German Politicians' Twitter Usage during the COVID-19 Crisis

Jonas Puchinger
jonas.puchinger@stud.uni-regensburg.de
University of Regensburg
Regensburg, Germany

Sabine Roth
sabine.roth@stud.uni-regensburg.de
University of Regensburg
Regensburg, Germany

Theresa Strohmeier
theresa.strohmeier@stud.uni-regensburg.de
University of Regensburg
Regensburg, Germany

ABSTRACT

As the COVID-19 pandemic affects nearly all aspects of daily life around the world, politicians are tasked with creating effective public health measures. Many politicians use Twitter as a tool for active and direct communication. We investigate the Twitter usage of German politicians throughout 2020 by analysing their tweets in regards to Twitter-inherent measures as well as NLP- and SNA-related methods. We find that politicians' tweets are generally characterised by negative sentiment, which gets slightly amplified when dealing with COVID-19. We also show that the networks created by the politicians Twitter accounts are strongly shaped by the political party-based landscape. These findings may help inform politicians on how to effectively use Twitter, and the general public on how to better understand government communication on social media, especially in combination with real world data.

CCS CONCEPTS

• **Networks** → **Social media networks**; • **Human-centered computing** → *Information visualization*; • **Computing methodologies** → **Natural language processing**; **Discourse, dialogue and pragmatics**; **Information extraction**.

KEYWORDS

COVID-19, social media, Twitter, politics

1 INTRODUCTION

The global pandemic COVID-19 presents an unprecedented crisis for the whole world. Because of the importance and frequency in communication, the largest microblogging platform Twitter already released a streaming endpoint[1] to enable researchers to collect tweets regarding this topic. The combination of Twitter's simplicity and popularity make it an effective tool for politicians to communicate their ideas to the general public, with the added possibility of direct feedback. Especially during a public crisis, politicians are the center of attention. Insights into the communication of German politicians on Twitter regarding COVID-19 can provide meaningful knowledge. Also differences between the political parties become clear. These findings can provide a glimpse for both, the politicians themselves and the general public. Therefore, we analyse the behaviour of German politicians on Twitter through quantitative and qualitative content-based methods such as sentiment analysis and topic detection, as well as social network analysis methods. All these tools are used with the distinct features and characteristics of Twitter in mind. This combination of approaches helps us to create a more holistic picture of the way politicians use Twitter. We show how to identify tweets regarding COVID-19, how to analyse their content, compare them to non-COVID-tweets, and how to contextualise these findings through real world data.

[1] https://developer.twitter.com/en/docs/labs/covid19-stream

2 RELATED WORK

2.1 COVID-19 on Twitter

The outbreak of COVID-19 has resulted in distinct communication across online environments [52]. Evaluating the emotions, concerns and criticisms enables an insight into the opinions of the population [33]. Particularly due to lockdown, quarantine and social distancing, more and more discussions were transferred to the internet [54]. Therefore, information about the pandemic is spreading even faster in social networks. This active exchange can cause the spread of myths and misinformation, but it also enables public health agencies to effectively disseminate information about the current situation [52].

Although the infectious disease only emerged about a year ago, there are already many scientific papers concerning the communication around the virus on Twitter. These studies often include sentiment analysis [44, 50], uncovering misinformation [20, 21] or dealing with the emergence and spread of COVID-related myths [56, 57]. The study by Gencoglu and Gruber proved that Twitter activity is more than six times higher and the sentiment is more negative on days with announcements of new deaths, infections or lockdowns [14].

Investigating public opinion in real time can help to predict the behaviour of a population in crisis situations [14]. Collected Twitter data have already proven to reflect reality as shown in the study of Culotta, in which incidence values of an influence were compared with the number of tweets concerning flu symptoms [11]. In addition, there are studies that show that Twitter data can provide important insights into public health crisis [35, 46]. Twitter studies were carried out in previous health crises such as Malaria or Ebola virus to analyse Twitter behaviour such as users' perspectives and reactions [5, 24]. In addition, the study of Signorini et al. concerning the H1N1 influenza demonstrate that Twitter data can estimate disease activity in real time [45].

2.2 Politics on Twitter

In recent years, Twitter has established itself as a political communication tool [6]. To date, there are various studies that deal with the influence of COVID-19 on the use of Twitter. The following studies provide an in-depth look at the start of the pandemic on the number and way political leaders have used Twitter and its impact on the public. By May 2020, 64.8% of 143 heads of state of the UN member states had already spoken out on the crisis on Twitter [16]. The fact that there is a demand for information from politicians is shown by the significant growth in followers of heads of state compared to before the pandemic [16]. In the study of Rufai and Bunce, the tweets of the G7 world leaders were analysed to find out how they are trying to draw attention to COVID-19 [39]. Out of a total of 203 viral tweets, 166 (82.8%) were used to provide information, including links to government sources. The remaining tweets were intended to raise morale or to politicise [39]. Yaqub examined the sentiment in tweets by Donald Trump during the early spread of the COVID-19 pandemic [58]. Two further studies looked at the political polarisation between parties and heads of state during the pandemic [19, 29]. The first shows that there is consensus among political elites and the general public in Canada on key actions and issues related to the pandemic [29]. For example, about social distancing or the severity of the measures taken. Whereas the second study shows that the tweets of the two leading parties in the USA are characterised by strong political polarisation [19].

2.3 Sentiment Analysis of Social Media Data

Sentiment analysis extracts emotional information from texts [60]. The content of social media platforms is characterised by, for example, shortness, contextual dependency and informativeness [60]. This presents a number of challenges for sentiment analysis. A tweet consists of a limited number of characters, which sometimes leads to the use of abbreviations that are not recognised by sentiment analysis algorithms [31, 60]. Other challenges include spelling mistakes, words with hashtags, emoticons but also content with images and videos [31, 60]. Research mainly deals with sentiment analysis of English-language content, meaning that fewer sentiment resources are available for other languages, which can lead to inaccurate results. Ways to compensate for this include translating the text into English or translating existing extensive corpora and lexicons from English into the respective focus language [32]. A disadvantage is the possible significant difference in sentiment assessment between the original and the translation [32]. Reasons for this include sarcasm, metaphorical expressions, and incorrect word rearrangements. Even correctly translated texts can result in incorrect sentiment assessment due to cultural context [32].

One of the approaches is lexicon-based sentiment analysis. This is a collection of predefined words, each of which is assigned a polarity value. With the help of this lexicon, word comparison is performed, which results in classification. The performance depends on the size and quality of the underlying lexicon. Moreover, the problem of synonyms can be reduced by using a domain-specific lexicon [43]. TextBlob DE[2] is a lexicon-based Python module for sentiment analysis that supports German as well as English. The study by Yaqub et al. used TextBlob to analyse English tweets [59], Schlör et al. for the analysis of German-language texts [41].

Given the challenges, we opted for an automated lexicon-based approach. This analysis can be performed straightforwardly, especially for large data sets. The option to translate tweets into English was eliminated due to error-proneness. TextBlob seems to be a suitable tool for analysing the sentiment of tweets.

2.4 Content Analysis of Social Media Data

Topic modelling enables the extraction of topics from a collection of words. It is based on the assumptions that frequently together appearing words in a text, belong to the same topic and every word collection can be described with a mixture of topics [8, 40]. A topic is meant to be "a group of significant words that share the same context and that are connected together because of their meaning" [40]. For the topic detecting many different models were developed, differing in the underlying assumptions but have a fast approximation to the inference using algorithms in common [8]. The Latent Dirichlet allocation (LDA) is one of the most common one [40]. It is based on a thematic modelling algorithm, which creates word groups from text documents based on a probability model without considering the word order and sentence context [30]. It also takes into account how the words in a document co-occur in the document [40].

Since the number of topics must be given as a parameter, one of the most difficult tasks is finding the best number of topics for the LDA model [40]. For this it is necessary to evaluate a created model and its quality. An LDA model can be assessed based on its values for perplexity and coherence [9]. The perplexity metric estimates how correctly a model built for most of the corpus predicts a smaller part of the word collection [26]. However, Chang et al. do not recommend the use of perplexity to validate the model as it is inconsistent with human comprehension [8]. More reliable results can be obtained with the coherence score, which measures how informative the topics are, to filter out topics that cannot be interpreted by humans [38, 51]. The most common coherence measurements (*c_v* and *u_mass*) calculate word similarity scores for the top words of each topic and the relative distance between the words [40, 51].

In the qualitative content analysis usually manual annotation or coding is used [22]. Annotation provides metadata for a text through manual assignment of labels and based on this, a classification into categories [34]. Since this procedure is very time-consuming, often a small set of data is labelled and then automatically applied to the entire corpus using algorithms [22]. However, this procedure is often flawed and imprecise, since the results depend on many parameters, such as human judgement, the attitude of the annotator, disagreements between annotators and lack of contextual knowledge [34, 49]. But manual annotation is also be used by researchers to evaluate or extend results obtained previously from automated procedures like the topic modelling [49].

Based on these findings and the fact that many studies working with data from social networks use LDA [2, 30, 40], we decided for this concept as well. LDA is proven to work well for large datasets like they are common in social networks [30]. To find the optimum value of topics we calculated the coherence score as presented before. For the qualitative content analysis we decided to use manual annotation with annotator-agreement for a subset of tweets with high engagement.

2.5 Social Network Analysis of Social Media

Social Network Analysis (SNA) aims to describe and explain interactions and connections in social clusters [42, 53]. The method relies on graph theory. Key components of a graph are nodes and

[2]https://textblob-de.readthedocs.io/en/latest/

vertices or edges, symbolising connections between nodes. Both nodes and vertices can have attributes and weights. A distinction is to be made between directed and undirected graphs. Whereas an undirected graph only depicts the existence of edges, a directed graph also shows the edge direction.

A prime subject for SNA is social media. Users of social media platforms are commonly simulated as nodes, and their interactions as vertices. Metrics for the evaluation of such networks can be classified as node-level, link-level and network-level measurements [17]. Node-level measurements include degree centrality (vertices per node), betweenness centrality (shortest path between two nodes), closeness centrality (average distance between a node and all other nodes), and eigenvector centrality (proximity to important nodes). Link-level measurements include directionality, link type or weight, and link reciprocity. Network-level measurements include density (possible connections over actual connections) and network reciprocity (cumulative link reciprocity) [17].

One use case for SNA is identifying influential users. Dubois and Gaffney differentiate between two types of influencers: the *opinion leader*, defined by Lazarsfeld et al. as a content-driven influencer [25], and the *influential*, important because of his position in a network [12]. Influential users can be identified by high node-level centrality measures, but content-based metrics are also applicable and can give different results [12]. Soares et al. use these theories to identify influential users in Twitter discussions regarding impeachment trials in Brazil. They found that influencers can indeed be identified via degree-centrality, but that their influence may be limited to strongly connected user groups [48].

Finding these groups is often called community detection or clustering. While community detection is achieved by analysing the structure of a network, clustering depends on node attributes to identify groups [55]. Both approaches can also be combined [55]. Most algorithms for community detection rely on modularity, a measure that identifies groups of strongly connected nodes with a large degree of separation from other groups in the same network. Popular algorithms include Louvian and Infomap [13]. Conflicting studies on which of these algorithms achieve higher performance exist [13, 23].

A big advantage of representing a social network as a graph is the ability to visualise the complex structure. Many tools for plotting network graphs exist [7, 27], with popular ones being Gephi [4] and PAJEK.

The connections between the accounts in our dataset are best represented as a directed network graph. This also allows us to determine the structure of the network via centrality measures, as well as the communities within the network via the Infomap algorithm. We plot the network and its clusters with Gephi.

3 DATA

3.1 Data Acquisition

To create a representative list of Twitter accounts of German politicians, we started by gathering all members of the German parliament (Bundestag) with an existing Twitter account. This list was expanded by adding all federal ministers with an existing Twitter account, that were not already part of the list. Additional politicians with highly active and popular, meaning high number of tweets and followers, Twitter accounts were also added to the list, which in total contains 551 politicians. Additional accounts are listed in appendix 5.

We scraped the tweets of these politicians from 1st January until 31st December 2020. The scraping took place on 21st January 2021, in order to give more recent tweets from the time frame a chance to gain engagement. We used TweetScraper[3] to gather the data from Twitter. This tool utilises the advanced search functionality of the Twitter web application to find the queried tweets. This approach has the drawback, that it can only scrape original tweets by the politicians and tweets quoted by them, native retweets are excluded. Further, the tool is able to provide the Twitter profile information for the queried accounts, which was scraped on the same day.

We used the Twitter API to scrape the follower relationships between the accounts of the politicians in our dataset, as well as their relations to the Twitter accounts of 13 German news portals and seven German virologists. These additional accounts were determined by selecting all mentions with an active Twitter account from surveys on popular news portals[4] and virologists[5] in Germany. These accounts are listed in appendix 7 and 6. We gathered data on COVID-19 case numbers in 2020 from ourworldindata.org [28].

3.2 Dataset Exploration

In total, the dataset contains 286,014 unique tweets. This number is made up of 254,861 (89.1%) original tweets by the politicians and 31,153 (10.9%) quoted tweets from other accounts. Going over all 551 politicians in the dataset, this comes to a mean of 462.5 original tweets per politician, with a median of 208 tweets. Politicians from eight parties are represented in the dataset. The distribution of original tweets per politician and party is displayed in Figure 1. The party *Bündnis 90/Die Grünen* gets shortened to *Die Grünen* in the following for the sake of simplicity.

On the tweet-level, the dataset notably contains, in addition to the content of the tweet, metadata about engagement metrics (likes, retweets, reply count), extracted tweet entities (hashtags, mentions, links), and, if present, connections to other tweets (quoted tweets, parent tweet, conversation). Additionally, the language of the tweet is annotated. While manual tests showed that these annotations may be false, especially for brief tweets, the overall impression is that the language is usually correctly annotated. A total of 91.4% of all original tweets by the politicians is annotated as German.

On the user-level, the dataset includes information such as the screen name, the account creation date, verification status, followers, following and friends counts, as well as counts of tweets and media posted.

3.3 Data Preprocessing

Data cleaning and preprocessing is required before certain methods of analysis can be performed. Cleaning steps include tokenization, lower-casing, emoji removal, URLs removal, mentions removal, punctuation removal, hashtags removal, whitespace removal, stop words removal and stemming or lemmatization [3, 18, 37]. Via

[3]https://github.com/jonbakerfish/TweetScraper
[4]https://de.statista.com/statistik/daten/studie/877238/umfrage/ranking-der-vertrauenswuerdigsten-nachrichtenquellen-in-deutschland/
[5]https://de.statista.com/statistik/daten/studie/1195709/umfrage/erwaehnung-von-virologen-epidemiologen-und-infektionsbiologen-in-der-presse/

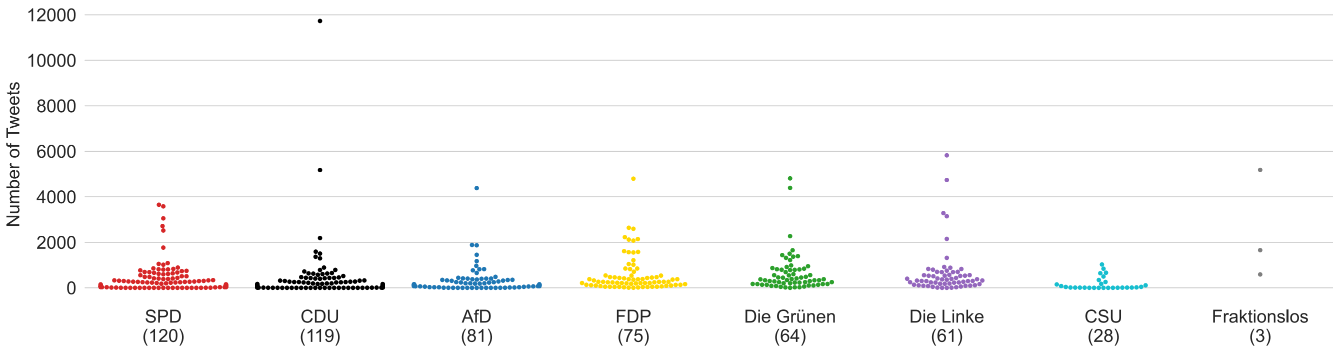

Figure 1: Distribution of politicians and their number of original tweets per party. Each dot represents a Twitter account.

tokenization a bag of words model was created [1]. For those, we used spacy[6], because of its better results in pretests compared to TweetTokenizer and the tokenizer of the nltk library. Next, with the corpus of the nltk[7] library we were able to remove the stopwords of the German tweets precisely. To remove the punctuation we used regex matching[8] and for getting the nouns in the bag of words the Textblob library with German extension. The remaining operations mentioned above were performed with spacy itself.

For later qualitative evaluations with manual annotations, we extracted the most viral tweets from the dataset. We selected 15 tweets per party with the most likes and the highest retweet count in combination to receive a manageable amount of data. Since we did this for both, the COVID-tweets (CT) and the non-COVID-tweets (NCT), we received 240 tweets. About this amount of data for qualitative research has proven itself in related work with similar approaches, e.g., in the study of Rufai and Bunce with 203 viral Tweets [39]. CT are the self-written tweets of the selected politicians, which either match at least one word from the word list in the matching process or the original tweet of a quoted tweet does (see section 4.1). Otherwise, the tweets are referred to as NCT. Most of the tweets in this dataset were from March (18.8%, 45 tweets) and December (12.1%, 29 tweets). Least tweets were from January and April (7 tweets each). Regarding the viral CT there were nearly a third from March (28.3%) and non from January. The most viral NCT were published in February (15%), but besides that the distribution is equal throughout the year.

4 TWEET-LEVEL ANALYSIS

4.1 Filtering COVID-19 related tweets

The DWDS topic glossary[9] for COVID-19 Pandemic served as basis for the word list. This is a collection of terms related to the COVID-19 pandemic and has been continuously updated since March 2020. For use in our study, the glossary was accessed on 12/08/2020. Words from the glossary were selected collectively based on the criterion that they were uniquely thematic to COVID-19. Through further Twitter research, analysis of word frequencies, and updating the listed words concerning the plural and female forms, the word list was expanded. During the final revision, it was determined that some words could not be clearly assigned to COVID-19. For these words to continue to be considered in the matching process, at least two of them must be included in a tweet to be considered one of the COVID tweets (CT), otherwise one word is sufficient. The extensive word list ensures that the largest possible number of COVID-related tweets is obtained. The final word list includes 624 words.

Before matching, the words in the word list and the tweets were cleaned in the same way (see section 3.3). To be considered as one of the CT there must be a unique match with a word from the word list in step 1. Step 2 is done by *Pattern Matching* with the remaining tweets, which allows matching words from the word list with flexible endings. In order to be able to evaluate tweets with possible spelling errors, *Fuzzy Matching* with ratio 90 is applied to the remaining tweets in the last step. The results were randomly checked in advance and the steps were adjusted if necessary. The matching is first done with the self-written tweets of the respective politician, but also quoted tweets are considered COVID-related if the original tweet was evaluated as such. The matching resulted in 44,312 self-written CT being found.

In a comparison between CT and the total number of self-written tweets by party, the CSU (20.6%) has the highest percentage, followed by the AfD (20.1%). Independent politicians have the lowest share with 9.1%, the rest of the parties are between 16% and 18%.

The highest proportion of CT was distributed among the parties, the SPD with 8,126 (18.3%) tweets, the lowest proportions CSU with 1,021 tweets and the independent politicians with 699 tweets. On average, the independent politicians have the highest number of CT with 233. The remaining parties range between 120 and 36 tweets.

Top tweeter from CT overall and within the CDU party is Ruprecht Polenz with 2,103 tweets, followed by Dr. Karl Lauterbach with 1,978 tweets as the top tweeter of the party SPD.

The course of the CT by months shows that most CT (8,002) were written in March. In the following months up to June the number dropped to 2,975 tweets. The number stayed relatively the same through October and then rose again to 4,000 to 5,000 tweets per month by the end of the year.

[6]https://spacy.io/api/doc
[7]https://www.nltk.org/api/nltk.html
[8]https://docs.python.org/3/library/re.html
[9]https://www.dwds.de/themenglossar/Corona

4.2 Verified Accounts

If a Twitter account is verified, other users can be certain that the account is operated by the person claiming to represent said account. The verification status is granted by Twitter after enquiry by the account owner. We examined the number of verified accounts for each party in our corpus. Notably, all parties have a percentage of over 50% when it comes to verified accounts, except for the far-right AfD, where only 27.2% of all members in our corpus are verified.

4.3 Hashtags

In order to obtain further information on the tweet behaviour of the politicians, we evaluated the hashtags used in their tweets. The most used hashtags in all tweets is *#Corona* (8,884 tweets), followed by *#AfD* (6,736 tweets) and the hashtag for the German parliament *#Bundestag* (5,417 tweets). The top hashtag *#Corona* was first used on 01/26/2020 by Renate Künast of the party Die Grünen. The temporal analysis states that this is the most frequently used hashtag throughout 2020 starting with march, with the exception of July (*#AfD*) and September (*#Moria*). While the former was mainly used by the politicians of the eponymous party (76.3%), the distribution within the parties of the latter is quite balanced.

The most common hashtags in the matched CT are *#Corona, #Coronakrise* (Corona crisis) and *#COVID19*. Nearly all popular hashtags of this dataset contain 'Corona' or 'Covid' and therefore are not specific and can be used for all kinds of CT. They were evenly distributed among the parties and the time frame. Notable hashtags with a time-limited popularity are *#CoronaWarnApp* (COVID Tracing App) in June, *#b2908* (demonstration against lockdown measures in Berlin) in August, *#Infektionsschutzgesetz* (infection protection law) in November and *#Lockdown* in December. Common hashtags in January were *#China*, *#Grippe* (influenza) and *#Panik* (panic).

The non-COVID-tweets (NCT) were led by the hashtags *#AfD*, *#Bundestag* (German parliament) and *#CDU*. Here too, the hashtag *#AfD* was largely used by politicians of the eponymous party. #Bundestag on second rank already pertain to less tweets, which indicates the wide range of the NCT. Additional frequently used hashtags were *#Wirecard* in July, *#Belarus* in August and *#Moria* in September. It is noticeable that *#Wirecard* and *#Moria* were used most frequently by the politicians of the party Die Linke, Belarus by the politicians of the parties FDP and SPD. More detailed information to the distribution of the hashtags within the parties can be seen in appendix B.

4.4 Domains

We also evaluated the tweets regarding the shared links. 75% of the 20 most frequent domains in the collected data throughout the year were news portals, the remaining were social media platforms and the video-sharing platform *YouTube*. The list is headed by the news portals *spiegel.de* and *welt.de*, *youtube.com* and the social media platform *facebook.com*. While the politicians of Die Grünen, Die Linke and SPD more often linked *SPIEGEL*; *WELT* was preferred by the parties FDP and AfD. *Facebook* was mainly shared from the politicians of the party AfD (72.5%). The highest circulation daily newspaper *BILD* was referenced most by the politicians of the parties CSU, FDP and AfD. The politicians of the party CSU referenced most frequently the Bavarian broadcasting corporation *br.de*. The members of the parties linked quite often the website of the own party except of the politicians of the AfD. Here the conservative newspaper *jungefreiheit.de* was linked often instead which is said to be the "ideological supply ship of right-wing populism" [10]. The independent politicians often referred to their own websites such as *mariomieruch.net* or *marcobuelow.de.*

Analysing the CT the news service *tagesschau.de* got more referenced in the CT (2.7%, rank 5 of all links in the tweets) in comparison to the NCT (1.9%, rank 9) as well as the *tagesspiegel* (2.8%, rank 4 in comparison to 2.3%, rank 6). They were most pronounced in July, October and November and within all parties. *Bild.de* was used more frequent in the pandemic context (2.3%, rank 7 in comparison to 1.5%, rank 12 of the NCT). *Twitter* itself was more common in NCT (2.8%, rank 5) than in CT (1.3%, rank 14). Despite that there were no meaningful differences between the parties and the used links in CT and the used links in NCT. The dataset is quite homogeneous between the groups mentioned. Further information to the frequently used domains can be looked up in appendix C.

4.5 N-Gram Analysis

In order to get a more detailed insight into the topics of tweets, the analysis was carried out on the basis of word frequencies. The CT were prepossessed as described in section 3.3.

The word *corona* occurs in the 20 most frequently used words in the CT of the individual parties. It is the most common in the CT of the CSU and it is among the Top 5 of the parties SPD, independent politicians, FDP and Die Linke.

Other words among the Top 20 in the CT are *lockdown* (CSU), *krise* (crises) in 4 out of 8 parties (Die Grünen, Die Linke, FDP and SPD). As well as *pandemie* (pandemic) in 5 out of 8 parties (CDU, SPD, FDP, Die Grünen and Die Linke).

Among the Top 20 covid-related bigrams are *(medizinisches, personal)* (medical, staff) (Die Linke), *(abstand, halten)* (distance, keep) (Die Grünen and CDU), *(schulen, kitas)* (schools, daycare centers) (Die Grünen), *(abstands-, hygieneregeln)* (distance, hygiene rules) (SPD).

Among the Top 20 covid-realated trigrams are *(gemeinsam, corona, bekämpfen)* (together, corona, fight) (Die Grünen), *(größte, wirtschaftskrise, seit)* (biggest, economic crisis, since) (FDP), *(risikogruppen, besser, schützen)* (risk groups, better, protect) (FDP), *(priorität, schule, kita)* (priority, school, daycare center) (CSU) and *(daheim, bleiben, kontakte)* (at home, stay, contacts) (CSU).

5 SENTIMENT ANALYSIS

5.1 Automatic Sentiment Detection

In order to extract the emotional information of a tweet, we carried out a sentiment analysis. The lexicon-based approach of Python TextBlob was used for this (see section 2.3). The words in the lexicon are assigned values in the range between 1 and -1 (1: positive, 0: neutral, -1: negative). Although TextBlob DE assigns subjectivity values to the words in addition to the polarity values, the underlying "German Polarity Lexicon" does not yet contain any subjectivity values.

[10]https://en.wikipedia.org/wiki/Junge_Freiheit

The analysis was carried out with the self-written COVID-tweets (CT) and non-COVID-tweets (NCT) of the politicians. Of all 44,312 CT, 17,761 (40.1%) were classified as positive and 8,688 were classified as negative (19.6%). 71,257 (33.8%) of the 210,511 NCT were classified as positive and 32,151 tweets as negative (15.3%). The highest proportion of tweets with negative sentiment was written by Die Linke (CT: 25.1%, NCT: 19.4%) and AfD (CT: 24.4%, NCT: 20.8%), the highest proportion with positive sentiment was written by CSU (CT: 48.5%, NCT: 37.3%) and SPD (CT: 47.4%, NCT: 36.9%).

The average sentiment of each party is almost identical when comparing CT and NCT. The highest sentiment on average is found for CSU and SPD (CT: both 0.13 (positive), NCT: both 0.12 (positive)) and the lowest AfD (CT: 0.04 (positive), NCT: 0.04 (positive)).

The course of the polarity over the entire year shows a minimum of -0.02 (negative) and a maximum of 0.34 (positive) of the CT. The average is 0.08 (positive). The large fluctuations of the polarity in the first quarter are probably due to the small amount of data. The minimum and maximum values of the NCT over the entire year are 0.05 (positive) and 0.12 (positive), the average is 0.09 (positive).

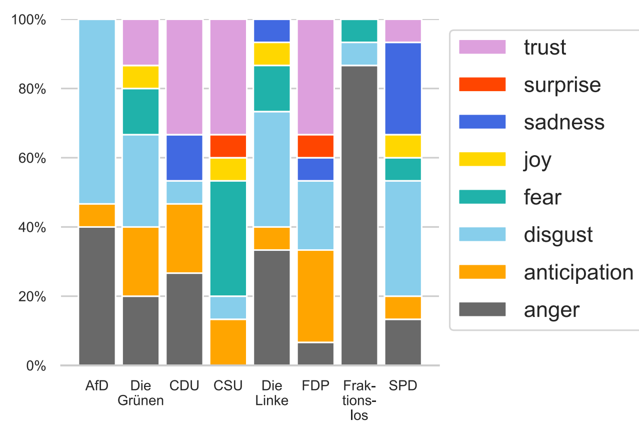

Figure 2: Emotion analysis of viral COVID-tweets per party.

5.2 Qualitative Sentiment and Emotion Analysis of Viral Tweets

In order to verify and extend the results of the automatic sentiment analysis, we carried out two more qualitative analyses, sentiment and emotion analysis. The dataset consisted of each 120 viral COVID-tweets (CT) and non-COVID-tweets (NCT) (see section 3.3). Each tweet was assigned a polarity (positive, neutral or negative) and a category of emotions. In detail, we focus on the emotions *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise* and *trust* according to Plutchik [36]. The full codebook definitions can be looked up in appendix D and E. The annotation was carried out without providing information about politicians and the party in order to enable an objective assessment. Furthermore, the authors commented separately from each other and then discussed for the annotator-agreement. The results of the sentiment analysis were: Of the total of 120 viral CT, 77 (64.1%) were classified as negative. The proportion of negative tweets is also highest with viral NCT at 88 (73.3%). The independent politicians (15 tweets), AfD (14 tweets) and Die Linke (13 tweets) largely contribute to these CT with negative polarity. For the NCT the AfD and Die Linke with 14 tweets each and the independent politicians with 13 tweets. A comparison of the results between the two sentiment analyses shows that the proportion of neutral tweets is higher in the automatic analysis. Both analyses show that Die Linke and AfD have the largest share of negative tweets.

Whereas the emotion analysis showed: Among the independent politicians, 13 of the 15 CT were assigned to the category *anger*. The second largest share of CT with the category *anger* is attributed to the AfD (40.0%), the remaining tweets are assigned to *disgust* (8 tweets) and *anticipation* (1 tweet).

Overall the three most common categories are: *anger* (28.3%, 34 tweets), *disgust* (23.3%, 28 tweets) and *trust* (15.0%, 18 tweets). The distribution of CT by party to the categories is shown in figure 2.

The three most common categories of all NCT are *anger* (41.6%, 50 tweets), *disgust* (21.6%, 26 tweets) and *trust* (15.0%, 18 tweets). The AfD (11 tweets) and independent politicians (10 tweets) have the highest proportion of NCT in the *anger* category. The party Die Linke has the largest share of NCT with the category *disgust* with 10 tweets.

The emotion analysis in comparison between CT and NCT shows that the biggest differences are in the categories *anger* (34 CT, 50 NCT), *fear* (11 CT, 3 NCT) and *anticipation* (15 CT, 6 NCT).

6 CONTENT ANALYSIS

6.1 Topic detection with LDA topic modelling

To detect the preferred topics in the tweets of the politicians we executed a content analysis. We decided to use the concept of topic modelling with Latent Dirichlet allocation (LDA) based on the findings of the research (see section 2.4). We examined the content of all tweets via the Python package *gensim*[11]. Therefore, the cleaned data were transformed into a document-term-matrix, and then into vectors with a series of tuples including the id and the word frequency [40, 47]. We decided to divide the dataset into the four quarters of the year because the course of the pandemic can be broken down into these. The first quarter was dominated by the emergence of the virus in Europe and Germany, followed by the very high infection numbers in the second period, the continuously low case numbers in summer and the rapidly increasing case numbers combined with the second lockdown in the last quarter.

In the first quarter all distinct topics were COVID related besides the Prime Minister election in Thuringia. In view of the high case numbers and the high proportion of COVID-tweets (CT) in March, this is not surprisingly. The second quarter was dominated by parliamentary COVID debates, school closures and the crisis in all. The party AfD in combination with the U.S. president Trump and racism were also dominant in this period. In the third quarter the topics were less dominated by the pandemic. The refugee disaster in Moria, the protests in Belarus and the debate regarding the German police were paramount. Nevertheless the German parliament and the school closures were key topics also in this quarter. The end of the year was dominated by the measures of the German parliament,

[11]https://radimrehurek.com/gensim/

such as exit restrictions and lockdowns. In this context democracy and the restriction of liberty rights were discussed extensively. The elections in the U.S. were the only non-COVID related main topic from October till December. All in all there are some topics standing out between pervasive COVID-related topics.

6.2 Qualitative content analysis of viral COVID-tweets

To extend the finding of the topic detection we performed a qualitative content analysis for a smaller dataset. Therefore we evaluated the 120 most viral CT of each party (see section 3.3). We labelled each tweet with a content category. The labelling of the tweets was done without knowing the politician or the party to prevent influence by subjective opinions. Each tweet was only assigned to one category describing the tweet best. The authors annotated the tweets separated and discussed the results afterwards for the annotator-agreement. The categories were took over from the scientific work of Chew and Eysenbach for classifying the CT [10]. The categories are *Humour or Sarcasm, Relief, Downplayed Risk, Concern, Frustration, Misinformation* and *Question*. According to Chew and Eysenbach the category *relief* contains any expression of happiness, joy or sense of peace and *concern* also includes fear, sadness and scepticism [10]. *Misinformation* is meant for every tweet contradicting the reference standard or containing unsubstantiated information, but also speculations, distrust and conspiracy. Also *Downplayed Risk* is broadly defined such as for tweets attempt to de-emphasise the potential risk as well as for tweets with a lack of interest in the risks [10]. The full codebook definitions of the 7 qualifiers with example tweets can be looked up in appendix F.

The findings of the qualitative content analysis of the viral CT confirm the negative mood resulted from the qualitative sentiment analysis. The category *frustration* was assign the most (42.5%), followed by *concern* (26.7%), while *relief* was only allocated 19 times (15.8%). As shown in figure 3 the politicians of the party CSU raised *concerns* in most of the tweets. The politicians of the party AfD expressed predominant *frustration*, as well as the independent politicians. Furthermore *misinformation* was only spread by these two groups. *Relief* was most shown by the politicians of the parties Die Grünen, CDU and FDP. The politicians of the party Die Linke attempted the most often with *humour or sarcasm*, however this category was often selected because of gloating or black humour.

The chronological sequence clarifies that in February tweets with *concerns* were dominant, while March and April were led by *frustration*. In July *downplayed risk* was the most widespread in terms of content combined with the most tweets containing *relief* for the year. Since October *frustration* dominated in the tweets again. In December also a lot of tweets containing *concerns* were published, especially from the politicians of the party CSU. *Misinformation* was spread in August and December. The most viral CT were tweeted in March (28.3%), followed by December (15%). There were no viral CT for January in our dataset.

7 NETWORK ANALYSIS

Social media interactions and structures can be effectively represented by networks. We generated such networks to examine the relationships between the politicians in our corpus, as well as their

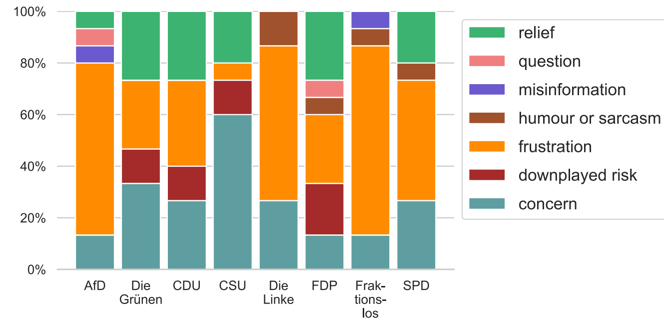

Figure 3: Qualitative content analysis of viral COVID-tweets. Percentage distribution of the categories per party.

connections to selected additional accounts (see section 3.1). Twitter accounts are represented by nodes in these networks. These nodes are identified by the Twitter-ID of the account, and are attributed with the screen name of the account, the name or title of the entity behind the account, as well as the association of the account respective to the structure of our corpus. Associations are therefore either the name of the corresponding party, if the account belongs to a politician, or *news portal* or *virologist*, depending of the type of additional account. We use a combination of directed edge types to model the structure of the networks. The first type is the follower relationship between two accounts. If given source account follows a given target account, an edge is created from source to target node. Additionally, we add edges for all mentions of an account in another accounts tweets, replies from one account to another accounts tweets, and quote tweets from one account of another accounts tweets. These last edge types are weighted by the number of occurrence of each given type of interaction. The network graphs are created via *networkx*[12], visualised via *Gephi*[13]. The resulting network based on all original tweets by the politicians can be seen in Figure 4.

The network shows a rather dense graph, without large distances separating single or clusters of nodes. This density is represented in the modularity value for this graph, which is rather low at 0.372. The colouring of nodes can be mapped almost perfectly to their position in the network, with the political parties forming clusters reminiscent of the rays of a star. Notable observations are the mixing of nodes belonging to CDU and CSU, which is to be expected with them being sister parties, the central positioning of the news portal nodes, and the relatively large separation of the far-right AfD from the other parties.

When using only COVID-tweets (CT) or non-COVID-tweets (NCT) as the basis for the graph creation, the resulting graphs are highly similar to the one based on all tweets. The graphs for based on all tweets and NCT are almost identical, which is expressed in the numbers for edge count, average degree, edge density and modularity. Since there are fewer tweets identified as regarding COVID-19 in our corpus, the graph based on those tweets has a smaller number of edges, which in turn leads to smaller numbers for average degree and edge density. One slight difference is the

[12]https://networkx.org/
[13]https://gephi.org/

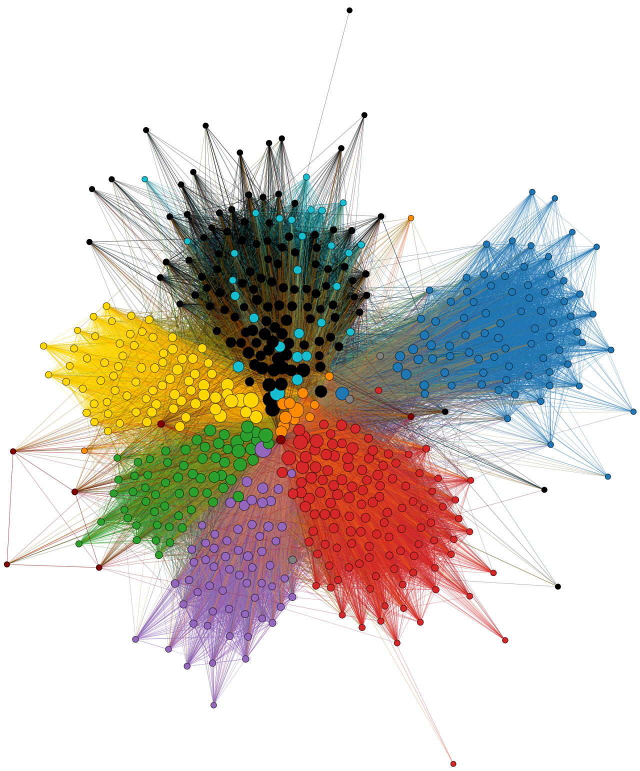

Figure 4: Graph of all original tweets by politicians. Nodes represent Twitter accounts, while edges represent following, replies, mentions, and quote tweets. Graph laid out using the Fruchterman Reingold layout algorithm.

higher value for modularity in this graph, indicating clearer division between the parties when dealing with COVID-19. Table 1 shows the relevant graph statistics for all three graphs.

Table 1: Comparison of the three graphs based on all tweets, non-COVID-tweets, and COVID-tweets.

Graph	All Tweets	Non-COVID Tweets	COVID Tweets
No. of Nodes	571	571	571
No. of Edges	74,279	71,101	56,502
Avg. Degree	130.1	126.3	98.9
Avg. Weighted Degree	239.9	215.4	108.9
Modularity	0.372	0.374	0.393

7.1 Community Detection

To examine clusters beyond what is discernible in the presented network, we used the Infomap algorithm[14] for community detection. This produces a list of eight communities for the graph based on all tweets, which we will therefore call *AT1 - AT8*.

[14]D. Edler, A. Eriksson and M. Rosvall, The MapEquation software package, available online at mapequation.org

This network is highly reminiscent of the original network, which becomes even clearer when checking which nodes belong to the identified communities. All larger communities can be mapped to the largest political parties in our corpus. Accounts belonging to the *news portal* category are distributed fairly evenly among different communities, while, apart from two exceptions, all *virologists* belong to *AT3*. A mix of nodes is grouped in *AT7*, where members of six different parties are present, alongside two news portals and one virologist. No inherent semantic relationship can be made out between those nodes. Table 2 lists the details of all communities present in this network.

Table 2: Communities detected by Infomap in the graph based on all tweets. The corresponding graph is shown in Figure 12.

Comm.	AT1	AT2	AT3	AT4	AT5	AT6	AT7	AT8
Nodes	149	122	67	76	61	81	14	1
%	26.1	21.4	11.7	13.3	10.7	14.2	2.5	0.2
Parties	CDU CSU	SPD Frk.los (1)	Die Grünen	FDP	Die Linke	AfD	Mixed	CDU
Add. Accs.	7 np. 1 vir.	2 np.	5 vir.	1np.	1 np.		2 np. 1 vir.	

In the graph based on the NCT, only six communities are detected by Infomap, henceforth called *NCT1 - NCT6*. Apart from *NCT6* the communities here are indicative of the current structure of German politics. Parties which currently form the opposition each have their own community, while the parties making up the government are grouped together in *NCT1*. Interestingly, all nodes belonging to the *news portal* category and six of seven nodes from the *virologist* category are also found in *NCT1*. A detailed rundown of these communities is shown in Table 3.

Table 3: Communities detected by Infomap in graph based on the non-COVID-tweets. The corresponding graph is shown in Figure 13.

Comm.	NCT1	NCT2	NCT3	NCT4	NCT5	NCT6
Nodes	290	63	75	61	81	1
%	50.8	11.0	13.1	10.7	14.2	0.2
Parties	CDU CSU SPD	Die Grünen	FDP	Die Linke	AfD	CDU
Add. Accs.	13 np. 6 vir.	1 vir.				

While the communities in the non-COVID-graph were tightly packed, the clusters found in the graph based on the CT are more differentiated. Nine communities are detected by Infomap, called *CT1 - CT9*. *CT1 - CT6* again are almost entirely made up of singular parties. *CT7* is similar to *AT7*, not only because it consists of a mix of nodes from different associations, but also because the nodes are for the most part identical. An interesting new cluster is *CT8*, which consist of five news portals, all virologists, one independent politician, and Dr. Karl Lauterbach, health policy spokesman of the SPD. Lauterbach is one of the most active and noted German politicians on Twitter during 2020, especially regarding COVID-19, so his connection to the virologists is logical. *CT9* only consists of RTL aktuell, a German news show, which is not very active on Twitter. All relevant statistics for these communities can be found in Table 4. The higher number of clusters in the COVID-graph compared to the tighter non-COVID-graph further indicates the notion that parties are more divided when it comes to COVID-19, while there are stronger connections when dealing with other political topics.

Table 4: Communities detected by Infomap in graph based on the COVID-tweets. The corresponding graph is shown in Figure 14.

Comm.	CT1	CT2	CT3	CT4	CT5	CT6	CT7	CT8	CT9
Nodes	144	119	61	76	61	82	13	14	1
%	25.2	20.8	10.7	13.3	10.7	14.4	2.3	2.5	0.2
Parties	CDU CSU	SPD Frk.-los (1)	Die Grünen	FDP	Die Linke CDU (1)	AfD Frk.-los (1)	Mixed	Mixed	
Add. Accs.	4 np.			1 np.			2 np.	7 vir.	1 np.

7.2 Influential Twitter users

To measure the relative influence of an actor in a network, different centrality measures can be employed. The main ones used for this analysis are indegree and outdegree.

Indegree measures how many incoming connections a single node has in a given network. In our case, incoming connections express follows, mentions, replies and quotes from other accounts. Users with high indegree measures are therefore influential because they receive a lot of attention for their content. Users like this are prime examples of opinion leaders [12, 25]. The nodes with the highest measures for indegree are generally popular politicians with positions such as party leaders or federal ministers. Christian Lindner, leader of the FDP, has the highest indegree at 657, with other notable mentions being German health minister Jens Spahn with an indegree of 567, and the aforementioned Karl Lauterbach with an indegree of 408. The only account in the top 50 that does not belong to a politician is n-tv, a German news TV station.

The outdegree of a node shows how many mentions, replies, quotes and follows one account generates in regards to others in the network. Accounts with a high outdegree therefore play a significant role in distributing content, and can take the role of a curator for sub-networks of users. They can be classified as superparticipants [15, 48]. When looking at the accounts with the highest outdegrees in our network, we mainly find very active Twitter users that produce large numbers of tweets and interactions. These are people like Ruprecht Polenz (CDU), who has the most original tweets in our corpus at 11,728, with an outdegree of 454, or Frank Pasemann (AfD) and Michael von Abercron (CDU), who have large outdegrees (554 and 459 respectively), but much smaller indegrees (109 and 87 respectively), cementing their role as content-spreading superparticipants. The number of news portals with large outdegrees is significantly higher than when looking at indegree measures, which is in line with their intention of distributing content rather than being the target of it.

When comparing these results to the graphs based on CT and NCT, it becomes apparent that much the same account have high scores for centrality measures in all three graphs. Accounts related to public health, like Jens Spahn or Karl Lauterbach place slightly higher in the COVID-graph, and loose some of their prominence when looking exclusively at NCT, but overall there are not many differences between the three networks.

Lastly, we analysed the network position of the additional account types *news portals* and *virologists*. We already mentioned that some news portals place highly when measuring centrality, especially with outdegree measures. Their high degrees in general, as well as their connections to many nodes from different communities position them at the center of the networks. They act as bridges between different groups, by creating own content and propagating topics concerning other members of the network. Mass media has inherently always been a target of attention, so their central positioning in the network is to be expected. Virologists on the other hand have only really begun to receive attention with the start of the COVID-19 pandemic. Since then, their popularity certainly has risen, but their positions and connections in the network are still minimal when compared to other groups. While the 13 news portals in our corpus of a total 571 nodes make up for 6.96% of all edges, all seven virologists can only account for 0.89% of all edges, with many of them connecting them among themselves.

8 COMPARISON OF TWITTER METRICS TO COVID-RELATED EVENTS

To contextualise the findings on the general Twitter activity of the accounts in our corpus, we provide a comparison between those Twitter metrics and case numbers from the COVID-19 pandemic. Figure 5 shows this comparison by plotting the daily new cases in Germany throughout 2020 against the number of COVID-tweets (CT) sent by the politicians on a given day.

This visualisation shows that an increase in infection rate corresponds to an increase in Twitter activity. A noticeable number of CT were posted before the pandemic really hit Germany. These tweets are a reaction to the course of the pandemic in other countries that were hit earlier than Germany, such as China, Spain, or Italy.

The peaks in the plot for the CT can be mapped to relevant events during the course of the pandemic. Especially political decisions like lockdowns and financial aid packages can easily be distinguished.

While the reaction on Twitter during the initial outbreak is very strong, the much larger wave of infections towards the end of the year elicits a somewhat weaker response.

9 DISCUSSION

The different types of analysis we performed give a holistic picture of the course of the pandemic in Germany throughout 2020. Our findings indicate that by tracking tweets of politicians, we can build an accurate timeline of important events, especially when they are related to COVID-19.

While many politicians have a Twitter account, only some are highly active and influential. Their example shows that Twitter can

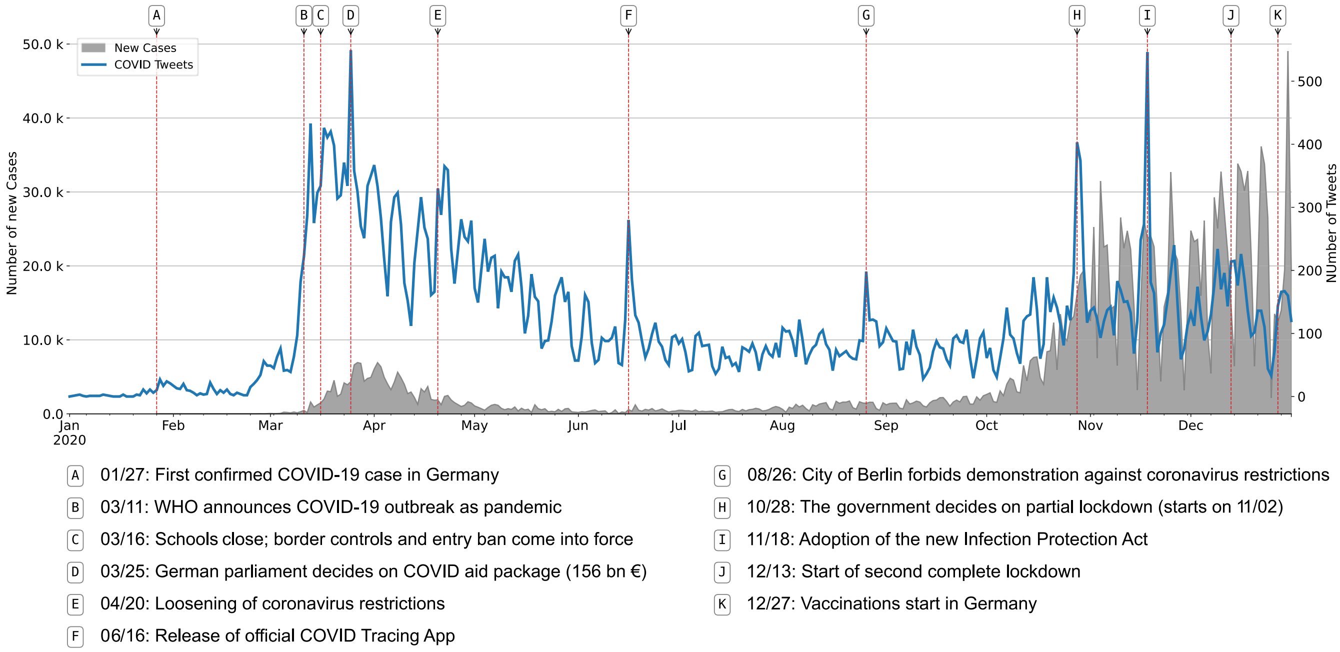

Figure 5: COVID-tweets and daily new COVID infections in Germany during 2020. Key events are marked and described below.

be an effective communication tool, and more politicians should embrace this direct way of conveying information.

Looking at the relations between the Twitter accounts of all politicians reveals, that political interaction is strongly shaped by party affiliations and the general political landscape. While the parties are strongly connected groups, their position to each other is far less divisive as in other countries such as the USA. We also find that parties are more divided regarding COVID-19 than 'everyday' political issues.

Content analysis of the politicians' tweets shows, that the general sentiment displayed in them is rather negative, only slightly higher when regarding COVID-19. While non-COVID-tweets have a tendency to filled by anger, COVID-tweets more often express fear and tense anticipation of the future. It can also be noticed that tweets by politicians are more concerned with decision making (e.g. deciding that a lockdown is happening), rather than the consequences of these decisions (e.g. living conditions during a lockdown).

The trend to get tired of the topic COVID-19 is generally felt throughout the population after dealing with the pandemic for over a year[15]. This trend is also noticeable in the politicians tweets, indicating a developed routine in dealing with the virus.

Connections of politicians' Twitter accounts to those of public health experts are far weaker than expected. This presents an opportunity for both sides to create stronger publicly visible relations, in order to better communicate health measures to the public.

10 LIMITATIONS & FUTURE WORK

As described before, there are some limitations regarding the NLP of non-English texts. Our study is also limited by the fact that we could not detect all COVID-tweets in the corpus, due to some tweets only containing images or videos, or not being matched by any keyword from our word list. Beyond that, we did not collect retweets, which would provide additional insights. Furthermore, we did not consider the development of follower numbers in the selected period, which would help to explain the increased popularity and importance of certain accounts. These considerations could be examined in future work, which should be made easier by Twitter enabling easier data access for researchers[16]. Along with the description of our methods, this opens up the opportunity to replicate and extend our study. Our results could also be the basis for analysing the communication of politicians on other online platforms.

11 CONCLUSION

Through our findings we explain how politicians use Twitter as a means of communication during a crisis. The results may help politicians to more effectively use Twitter, and the general public to better understand government communication on social media. We show that using a variety of methods helps to create a better understanding of behaviour on Twitter and social media in general. Additionally, our findings highlight the importance of contextualising results of data-driven social media analytics through real world qualitative data. By showing how these methods are used and how they are limited we provide a baseline for future work.

[15] https://projekte.uni-erfurt.de/cosmo2020/web/summary/34/

[16] https://blog.twitter.com/developer/en_us/topics/tools/2021/enabling-the-future-of-academic-research-with-the-twitter-api.html

REFERENCES

[1] Luca Maria Aiello, Georgios Petkos, Carlos Martin, David Corney, Symeon Papadopoulos, Ryan Skraba, Ayse Goker, Ioannis Kompatsiaris, and Alejandro Jaimes. 2013. Sensing Trending Topics in Twitter. *IEEE Transactions on Multimedia* 15, 6 (2013), 1268–1282. https://doi.org/10.1109/TMM.2013.2265080

[2] Tawfiq Ammari, Daniel Romero, and Sarita Schoenebeck. 2019. Self-declared Throwaway Accounts on Reddit: How Platform Affordances and Shared Norms enable Parenting Disclosure and Support. *Computer Supported Cooperative Work* 3 (2019), 135:1–30. https://doi.org/10.1145/3359237

[3] Arpita, Pardeep Kumar, and Kanwal Garg. 2020. Data Cleaning of Raw Tweets for Sentiment Analysis. In *2020 Indo – Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN)* (Rajpura, Punjab, India). IEEE, 273–276. https://doi.org/10.1109/Indo-TaiwanICAN48429.2020.9181326

[4] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media* 3, 1 (Mar. 2009).

[5] James Boit and Omar El-Gayar. 2020. Topical mining of malaria using social media. A text mining approach. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

[6] Gwen Bouvier and Judith E Rosenbaum. 2020. Communication in the Age of Twitter: The Nature of Online Deliberation. In *Twitter, the Public Sphere, and the Chaos of Online Deliberation*. Springer, 1–22.

[7] Enrico G. Caldarola and Antonio M. Rinaldi. 2017. Big Data Visualization Tools: A Survey. In *Proceedings of the 6th International Conference on Data Science, Technology and Applications* (Madrid, Spain) *(DATA 2017)*. SCITEPRESS - Science and Technology Publications, Lda, Setubal, PRT, 296–305. https://doi.org/10.5220/0006484102960305

[8] Jonathan Chang, Jordan Boyd-Graber, Chong Wang, Sean Gerrish, and David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Neural information processing systems*, Vol. 22. Citeseer, 288–296.

[9] Zhiyuan Chen and Bing Liu. 2014. *Topic Modeling using Topics from Many Domains, Lifelong Learning and Big Data*. Vol. 3. Journal Abbreviation: 31st International Conference on Machine Learning, ICML 2014 Publication Title: 31st International Conference on Machine Learning, ICML 2014.

[10] Cynthia Chew and Gunther Eysenbach. 2010. Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. *PLOS ONE* 5, 11 (2010), e14118. https://doi.org/10.1371/journal.pone.0014118 Publisher: Public Library of Science.

[11] Aron Culotta. 2010. Towards detecting influenza epidemics by analyzing Twitter messages. In *Proceedings of the first workshop on social media analytics*. 115–122.

[12] Elizabeth Dubois and Devin Gaffney. 2014. The Multiple Facets of Influence. *American Behavioral Scientist* 58 (08 2014), 1260–1277. https://doi.org/10.1177/0002764214527088

[13] Scott Emmons, Stephen Kobourov, Mike Gallant, and Katy Börner. 2016. Analysis of Network Clustering Algorithms and Cluster Quality Metrics at Scale. *PLOS ONE* 11, 7 (07 2016). https://doi.org/10.1371/journal.pone.0159161

[14] Oguzhan Gencoglu and Mathias Gruber. 2020. Causal modeling of twitter activity during covid-19. *Computation* 8, 4 (2020), 85.

[15] Todd Graham and Scott Wright. 2014. Discursive Equality and Everyday Talk Online: The Impact of "Superparticipants". *Journal of Computer-Mediated Communication* 19, 3 (2014), 625–642. https://doi.org/10.1111/jcc4.12016 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcc4.12016

[16] Michael Haman. 2020. The use of Twitter by state leaders and its impact on the public during the COVID-19 pandemic. *Heliyon* 6, 11 (2020), e05540.

[17] Itai Himelboim. 2017. Social Network Analysis (Social Media). In *The International Encyclopedia of Communication Research Methods*. American Cancer Society, 1–15. https://doi.org/10.1002/9781118901731.iecrm0236

[18] S. T. Indra, Liza Wikarsa, and Rinaldo Turang. 2016. Using logistic regression method to classify tweets into the selected topics. In *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (Malang, Indonesia). IEEE, 385–390. https://doi.org/10.1109/ICACSIS.2016.7872727

[19] Julie Jiang, Emily Chen, Shen Yan, Kristina Lerman, and Emilio Ferrara. 2020. Political polarization drives online conversations about COVID-19 in the United States. *Human Behavior and Emerging Technologies* 2, 3 (2020), 200–211.

[20] Ramez Kouzy, Joseph Abi Jaoude, Afif Kraitem, Molly B El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W Akl, and Khalil Baddour. 2020. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus* 12, 3 (2020). https://doi.org/10.7759/cureus.7255

[21] Chayakrit Krittanawong, Bharat Narasimhan, Hafeez Ul Hassan Virk, Harish Narasimhan, Joshua Hahn, Zhen Wang, and W. H. Wilson Tang. 2020. Misinformation Dissemination in Twitter in the COVID-19 Era. *The American Journal of Medicine* 0, 0 (Aug. 2020). https://doi.org/10.1016/j.amjmed.2020.07.012 Publisher: Elsevier.

[22] Katie Kuksenok, Michael Brooks, John J Robinson, Daniel Perry, Megan K Torkildson, and Cecilia Aragon. 2012. Automating large-scale annotation for analysis of social media content. In *IEEE Workshop on Interactive Visual Text Analytics for Analysis of Social Media*.

[23] Andrea Lancichinetti and Santo Fortunato. 2009. Community detection algorithms: A comparative analysis. *Phys. Rev. E* 80 (Nov 2009), 056117. Issue 5. https://doi.org/10.1103/PhysRevE.80.056117

[24] Allison J Lazard, Emily Scheinfeld, Jay M Bernhardt, Gary B Wilcox, and Melissa Suran. 2015. Detecting themes of public concern: a text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. *American journal of infection control* 43, 10 (2015), 1109–1111.

[25] Paul F. Lazarsfeld, Bernard Berelson, and Hazel Gaudet. 1968. *The People'S Choice*. Columbia University Press. https://doi.org/doi:10.7312/laza93930

[26] Daniel Maier, Annie Waldherr, Peter Miltner, Gregor Wiedemann, Andreas Niekler, Alexa Keinert, Barbara Pfetsch, Gerhard Heyer, Ueli Reber, Thomas Häussler, et al. 2018. Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures* 12, 2-3 (2018), 93–118.

[27] S. Majeed, M. Uzair, U. Qamar, and A. Farooq. 2020. Social Network Analysis Visualization Tools: A Comparative Review. In *2020 IEEE 23rd International Multitopic Conference (INMIC)*. 1–6. https://doi.org/10.1109/INMIC50486.2020.9318162

[28] Esteban Ortiz-Ospina Max Roser, Hannah Ritchie and Joe Hasell. 2020. Coronavirus Pandemic (COVID-19). *Our World in Data* (2020). https://ourworldindata.org/coronavirus.

[29] Eric Merkley, Aengus Bridgman, Peter John Loewen, Taylor Owen, Derek Ruths, and Oleg Zhilin. 2020. A rare moment of cross-partisan consensus: Elite and public response to the COVID-19 pandemic in Canada. *Canadian Journal of Political Science/Revue canadienne de science politique* 53, 2 (2020), 311–318.

[30] Sara Mifrah and EL Habib Benlahmar. 2020. Topic Modeling Coherence: A Comparative Study between LDA and NMF Models using COVID'19 Corpus. *International Journal of Advanced Trends in Computer Science and Engineering* 9, 4 (2020). https://doi.org/10.30534/ijatcse/2020/231942020

[31] Saif M Mohammad. 2016. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*. Elsevier, 201–237.

[32] Saif M Mohammad, Mohammad Salameh, and Svetlana Kiritchenko. 2016. How translation alters sentiment. *Journal of Artificial Intelligence Research* 55 (2016), 95–130.

[33] Akif Mustafa, Imaduddin Ansari, Subham Mohanta, and Shalem Balla. 2020. PUBLIC REACTION TO COVID-19 ON TWITTER: A THEMATIC ANALYSIS. *EPRA International Journal of Multidisciplinary Research (IJMR)* 6, 5 (May 2020), 349–358. https://doi.org/10.36713/epra4518

[34] Desmond Patton, Philipp Blandfort, William Frey, Michael Gaskell, and Svebor Karaman. 2019. Annotating social media data from vulnerable populations: Evaluating disagreement between domain experts and graduate student annotators. Proceedings of the 52nd Hawaii International Conference on System Sciences, 1–10.

[35] Michael J Paul and Mark Dredze. 2012. A model for mining public health topics from Twitter. *Health* 11, 16-16 (2012), 1.

[36] Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*. Elsevier, 3–33.

[37] Saurav Pradha, Malka N. Halgamuge, and Nguyen Tran Quoc Vinh. 2019. Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data. In *2019 11th International Conference on Knowledge and Systems Engineering (KSE)* (Da Nang, Vietnam). IEEE, 1–8. https://doi.org/10.1109/KSE.2019.8919368

[38] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (Shanghai, China) *(WSDM '15)*. Association for Computing Machinery, New York, NY, USA, 399–408. https://doi.org/10.1145/2684822.2685324

[39] Sohaib R. Rufai and Catey Bunce. 2020. World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. *Journal of Public Health* 42, 3 (Aug. 2020), 510–516. https://doi.org/10.1093/pubmed/fdaa049 Publisher: Oxford Academic.

[40] Paola Sanfilippo. 2019. *Understanding user similarity from social media using topic analysis*. Ph.D. Dissertation. Dipartimento di Elettronica, Informazione e Bioingegneria Politecnico di Milano.

[41] Daniel Schlör, Albin Zehe, Konstantin Kobs, Blerta Veseli, Franziska Westermeier, Larissa Brübach, Daniel Roth, Marc Erich Latoschik, and Andreas Hotho. 2020. Improving Sentiment Analysis with Biofeedback Data. In *Proceedings of LREC2020 Workshop" People in language, vision and the mind"(ONION2020)*. 28–33.

[42] John Scott. 1988. Social network analysis. *Sociology* 22, 1 (1988), 109–127.

[43] Dipti Sharma, Munish Sabharwal, Vinay Goyal, and Mohit Vij. 2020. Sentiment Analysis Techniques for Social Media Data: A Review. In *First International Conference on Sustainable Technologies for Computational Intelligence*. Springer, 75–90.

[44] Wladimir Sidorenko. 2019. *Sentiment Analysis of German Twitter*. Ph.D. Dissertation. Universität Potsdam.

[45] Alessio Signorini, Alberto Maria Segre, and Philip M Polgreen. 2011. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PloS one* 6, 5 (2011), e19467.

[46] Lauren Sinnenberg, Alison M Buttenheim, Kevin Padrez, Christina Mancheno, Lyle Ungar, and Raina M Merchant. 2017. Twitter as a tool for health research: a

systematic review. *American journal of public health* 107, 1 (2017), e1–e8.
[47] D. A. Smith and C. McManis. 2015. Classification of text to subject using LDA. *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)* (2015), 131–135. https://doi.org/10.1109/ICOSC.2015.7050791
[48] Felipe Bonow Soares, Raquel Recuero, and Gabriela Zago. 2018. Influencers in Polarized Political Networks on Twitter. In *Proceedings of the 9th International Conference on Social Media and Society* (Copenhagen, Denmark) *(SMSociety '18)*. Association for Computing Machinery, New York, NY, USA, 168–177. https://doi.org/10.1145/3217804.3217909
[49] Hyunjin Song, Petro Tolochko, Jakob-Moritz Eberl, Olga Eisele, Esther Greussing, Tobias Heidenreich, Fabienne Lind, Sebastian Galyga, and Hajo G Boomgaarden. 2020. In validations we trust? The impact of imperfect human annotations as a gold standard on the quality of validation of automated content analysis. *Political Communication* 37, 4 (2020), 550–572.
[50] Himanshu Tanwar. 2020. Twitter Sentiment Analysis during Covid -19 Pandemic. Vol-06 (April 2020). https://doi.org/10.35291/2454-9150.2020.0300
[51] Poonam Tijare and P Jhansi Rani. 2020. Exploring popular topic models. *Journal of Physics: Conference Series* 1706 (2020), 012171. https://doi.org/10.1088/1742-6596/1706/1/012171
[52] Yan Wang, Haiyan Hao, and Lisa Platt. 2020. Examining Risk and Crisis Communications of Government Agencies and Stakeholders during Early-Stages of COVID-19 on Twitter. *Computers in Human Behavior* (Sept. 2020). https://doi.org/10.1016/j.chb.2020.106568
[53] Stanley Wasserman, Katherine Faust, et al. 1994. *Social network analysis: Methods and applications*. Cambridge university press.
[54] Philipp Wicke and Marianna Bolognesi. 2020. Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter. *PLoS ONE* 15 (Sept. 2020), e0240010. https://doi.org/10.1371/journal.pone.0240010
[55] Jaewon Yang, Julian McAuley, and Jure Leskovec. 2013. Community detection in networks with node attributes. In *2013 IEEE 13th international conference on data mining*. IEEE, 1151–1156.
[56] Shuiqiao Yang, Jiaojiao Jiang, Arindam Pal, Kun Yu, Fang Chen, and Shui Yu. 2020. Analysis and Insights for Myths Circulating on Twitter during the COVID-19 Pandemic. *IEEE Open Journal of the Communications Society* (Oct. 2020). https://doi.org/10.1109/OJCS.2020.3028573
[57] S. Yang, J. Jiang, A. Pal, K. Yu, F. Chen, and S. Yu. 2020. Analysis and Insights for Myths Circulating on Twitter During the COVID-19 Pandemic. *IEEE Open Journal of the Computer Society* 1 (2020), 209–219. https://doi.org/10.1109/OJCS.2020.3028573
[58] Ussama Yaqub. 2020. Tweeting During the Covid-19 Pandemic: Sentiment Analysis of Twitter Messages by President Trump. *Digital Government: Research and Practice* 2, 1 (Nov. 2020), 1–7. https://doi.org/10.1145/3428090
[59] Ussama Yaqub, Nitesh Sharma, Rachit Pabreja, Soon Ae Chun, Vijayalakshmi Atluri, and Jaideep Vaidya. 2020. Location-based Sentiment Analyses and Visualization of Twitter Election Data. *Digital Government: Research and Practice* 1, 2 (2020), 1–19.
[60] Wenping Zhang, Mengna Xu, and Qiqi Jiang. 2018. Opinion mining and sentiment analysis in social media: challenges and applications. In *International Conference on HCI in Business, Government, and Organizations*. Springer, 536–548.

Appendices

A ADDITIONAL ACCOUNTS

Table 5: The additional politicians and their Twitter accounts selected for the dataset. The accounts were chosen for their high number of tweets and/or their high number of followers and/or their political status (e.g. party leader).

Name	Party	Twitter Handle
Kramp-Karrenbauer, Annegret	CDU	@akk
Klöckner, Julia	CDU	@JuliaKloeckner
Schulze, Svenja	SPD	@SvenjaSchulze68
Söder, Markus	CSU	@Markus_Soeder
Walter-Borjans, Norbert	SPD	@NowaboFM
Meuthen, Prof. Dr. Jörg	AfD	@Joerg_Meuthen
Beck, Volker	Bündnis 90/Die Grünen	@Volker_Beck
Polenz, Ruprecht	CDU	@polenz_r
Kühnert, Kevin	SPD	@KuehniKev
Stegner, Ralf	SPD	@Ralf_Stegner

Table 6: The virologists and their Twitter accounts selected for the dataset.

Name	Twitter Handle
Drosten, Christian	@c_drosten
Streeck, Hendrik	@hendrikstreeck
Schmidt-Chanasit, Jonas	@ChanasitJonas
Kekulé, Alexander	@AlexanderKekule
Addo, Marylyn	@marylyn_addo
Brinkmann, Melanie	@BrinkmannLab
Ciesek, Sandra	@CiesekSandra

Table 7: The news portals and their Twitter accounts selected for the dataset.

Name	Twitter Handle
ARD Tageschau	@tagesschau
ZDF heute	@ZDFheute
Süddeutsche Zeitung	@SZ
Die ZEIT	@DIEZEIT
WELT	@welt
n-tv	@ntvde
FAZ	@faznet
Der Spiegel	@derspiegel
Focus	@focusonline
Stern	@sternde
RTL aktuell	@rtl_aktuell
t-online	@tonline
Bild	@BILD

B HASHTAGS

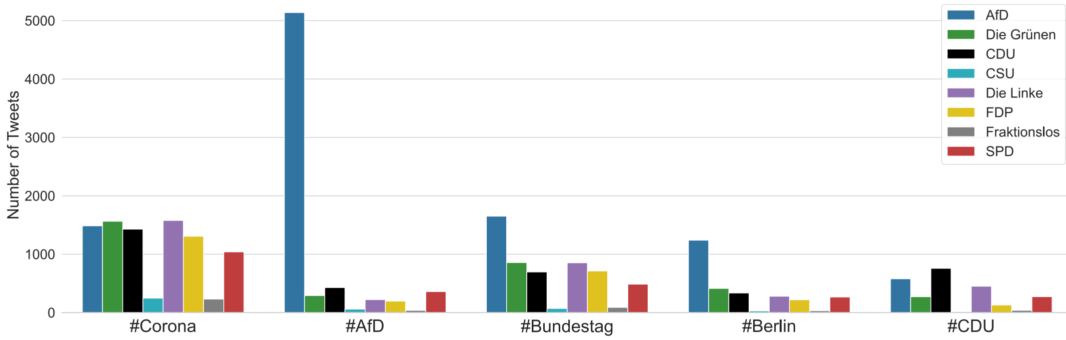

Figure 6: The 5 most frequently used hashtags in 2020 and their distribution within the parties. #Corona is the most popular hashtag of all tweets in our corpus. #AfD is mainly because of the eponymous party the second most used hashtag.

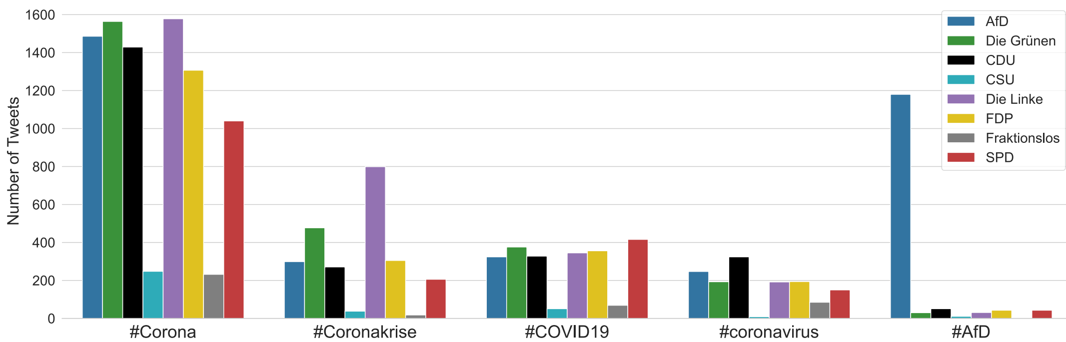

Figure 7: The 5 most frequently used hashtags of the COVID-tweets in our corpus in 2020 and their distribution within the parties. The popular hashtags contain 'Corona' or 'Covid', are not specific and can be used for all kinds of COVID-tweets.

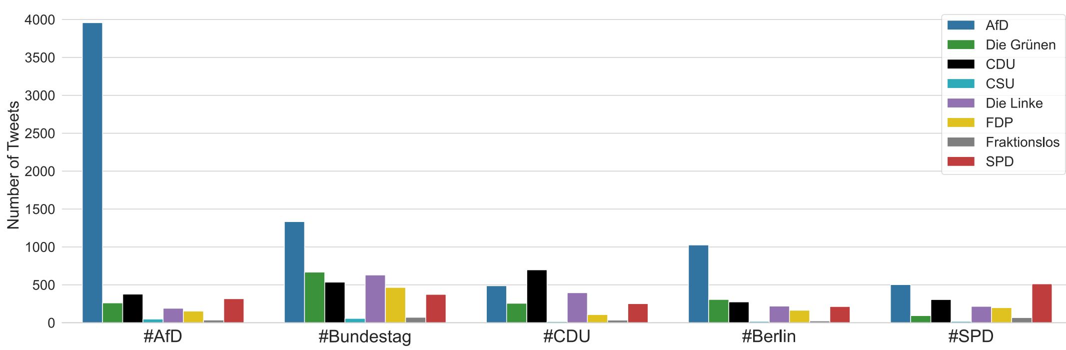

Figure 8: The 5 most frequently used hashtags of the non-COVID-tweets in our corpus in 2020 and their distribution within the parties. Again #AfD is mainly used from the eponymous party. #Bundestag on second rank already have less tweets, which indicates the wide range of the non-COVID tweets.

C DOMAINS

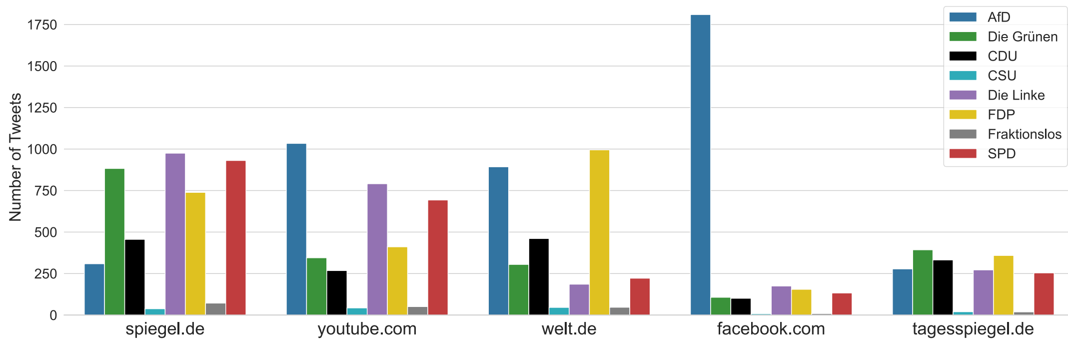

Figure 9: The 5 most frequently used domains in 2020 and their distribution within the parties. Spiegel.de is the most popular, but also YouTube get referenced quite often. Facebook is linked mostly by the AfD and quite less from the other parties.

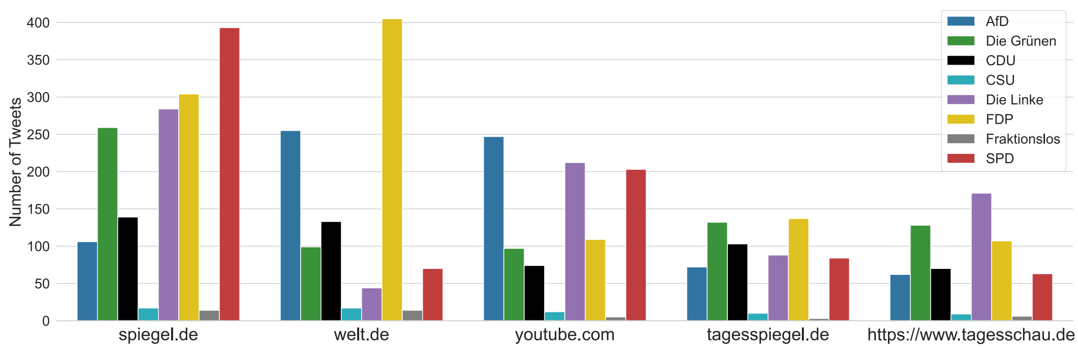

Figure 10: The 5 most frequently used domains of the COVID-tweets in 2020 and their distribution within the parties. In the pandamic context welt.de was mentioned more often than in all tweets. Also, tagesspiegel.de and tagesschau.de gain in importance.

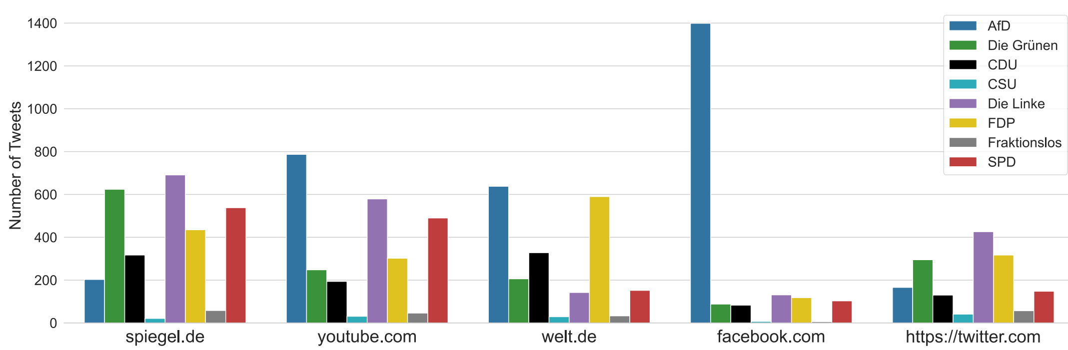

Figure 11: The 5 most frequently used domains of the non-COVID-tweets in 2020 and their distribution within the parties. Just two news portals were most popular when excluding the tweets about the pandemic.

D CODING GUIDELINE FOR QUALITATIVE SENTIMENT ANALYSIS

Table 8: Codebook for qualitative sentiment analysis of viral COVID-tweets and non-COVID-tweets.

Qualifier	Description	Example Tweet
positive	The dominant sentiment of the tweet is positive. This is supported by words and/or emoticons that can be clearly assigned to this sentiment.	"*Yes he can! Herzlichen Glückwunsch an @JoeBiden Bin sehr erleichtert, dass der Wahlkrimi ein gutes Ende genommen hat. Mein Vertrauen in die amerikanische Demokratie ist wieder gestärkt. Jetzt sollte das bizarre Schauspiel der letzten Tage ein Ende finden.*"
neutral	The dominant sentiment of the tweet is neutral. The words and/or emoticons cannot be assigned to positive or negative sentiment.	"*Die FDP-Bundestagsfraktion hat einstimmig beschlossen, den Gesetzentwurf für ein Infektionsschutzgesetz morgen im Deutschen Bundestag abzulehnen.*"
negative	The dominant sentiment of the tweet is negative. This is supported by words and/or emoticons that can be clearly assigned to this sentiment.	"*Atombomber sind nicht systemrelevant. 18,5 Mrd. € für 138 neue Kampfflugzeuge fehlen für Investitionen und bessere Bezahlung in Gesundheit und Pflege, Bildung und Erziehung, Handel und Logistik und die Bewältigung der Coronakrise. Katastrophale Entscheidung stoppen! https://t.co/iURFBMkMkn*"

E CODING GUIDELINE FOR QUALITATIVE EMOTION ANALYSIS

Table 9: Codebook for qualitative emotion analysis of viral COVID-tweets and non-COVID-tweets.

Qualifier	Description	Example Tweet
anger	The dominant emotion expressed in the tweet is anger. This is supported by words and/or emoticons that can be clearly assigned to this emotion. Also includes emotions such as fury and rage.	"*Tausende #Covidioten feiern sich in #Berlin als „die zweite Welle", ohne Abstand, ohne Maske. Sie gefährden damit nicht nur unsere Gesundheit, sie gefährden unsere Erfolge gegen die Pandemie und für die Belebung von Wirtschaft, Bildung und Gesellschaft. Unverantwortlich!*"
anticipation	The dominant emotion expressed in the tweet is anticipation. This is supported by words and/or emoticons that can be clearly assigned to this emotion.	"*Starker Start: Die #CoronaWarnApp wurde bereits 6,5 Millionen Mal heruntergeladen. Das sollte noch mehr Bürger motivieren, mitzumachen. Denn Corona einzudämmen, ist ein Teamspiel. Jeder, der die App nutzt, macht einen Unterschied. #IchAppMit*"
disgust	The dominant emotion expressed in the tweet is disgust. This is supported by words and/or emoticons that can be clearly assigned to this emotion. Also includes emotions such as aversion and loathing.	"*Es sind keine #Covidioten. Der Begriff verharmlost die Ziele der Drahtzieher. Es ist eine Querfront, die unsere Demokratie verachtet, mit Fake News eine geschlossene Gegenwelt aufbauen will (Stichwort:Lügenpresse) um sie gegen den freiheitlichen Rechtsstaat in Stellung zu bringen*"
fear	The dominant emotion expressed in the tweet is fear. This is supported by words and/or emoticons that can be clearly assigned to this emotion. Also includes emotions such as anxiety and fright.	"*Wir müssen damit rechnen, dass Corona mit voller Wucht wieder auf uns zukommt. Mir machen die steigenden Fallzahlen in Deutschland große Sorgen. Es ist absolute Wachsamkeit gefragt und deshalb ist jetzt nicht die Zeit für neue Lockerungen oder naive Unvorsichtigkeit. 1/2*"
joy	The dominant emotion expressed in the tweet is joy. This is supported by words and/or emoticons that can be clearly assigned to this emotion.	"*Freie Verpflegung für alle Mitarbeiter in Krankenhäusern, Kliniken, Alten-, Pflege- und Behinderteneinrichtungen: Bayern übernimmt ab 1. April die Kosten für Essen und Getränke während der Corona-Krise. Herzlichen Dank für die Arbeit, die rund um die Uhr geleistet wird.*"
sadness	The dominant emotion expressed in the tweet is sadness. This is supported by words and/or emoticons that can be clearly assigned to this emotion.	"*Diese Todesanzeige in meiner Lokalzeitung hat mich heute sehr bewegt. Wer sagt, es treffe nur die Alten: Wollen Sie mit 70 sterben? Wollen Sie mit 47 den Vater verlieren? Jedes Leben ist gleich viel wert. #aha #zusammengegencorona https://t.co/JugMeEt46p*"
surprise	The dominant emotion expressed in the tweet is surprise. This is supported by words and/or emoticons that can be clearly assigned to this emotion.	"*Was hat die Corona-Politik mit @Markus_Soeder gemacht, dass er die @fdp in die Nähe der #AfD rückt, weil wir an der Beteiligung der Parlamente bei Eingriffen in Grundrechte festhalten und die Wirksamkeit von Maßnahmen begründet sehen wollen? CL*"
trust	The dominant emotion expressed in the tweet is trust. This is supported by words and/or emoticons that can be clearly assigned to this emotion.	"*Ich finde der Söder macht das gerade gut. Krisenzeiten sind in Deutschland Zeiten, wo Demokraten über Parteigrenzen hinweg zusammenrücken. Es braucht jetzt Entschlossenheit und mehr Tempo. #coronavirusdeutschland #Corona #COVID19 #Soeder*"

F CODING GUIDELINE FOR THE QUALITATIVE CONTENT ANALYSIS

Table 10: Adapted codebook of the seven categories for COVID-19 related tweets for the qualitative content analysis by Chew and Eysenbach [10].

Qualifier	Description	Example Tweet
Humour or Sarcasm	Tweet is comedic or sarcastic.	"*Wenn drei FreundInnen unter freiem Himmel zusammentreffen, ist es verboten. Wenn die selben drei allerdings in der Werkshalle für Rheinmetall Waffen bauen oder für Amazon Pakete füllen, ist es kein Problem. Fazit: Kapitalismus schützt vor Viren! #Kontaktverbot*"
Relief	Tweet expresses joy, happiness, or sense of peace.	"*Freie Verpflegung für alle Mitarbeiter in Krankenhäusern, Kliniken, Alten-, Pflege- und Behinderteneinrichtungen: Bayern übernimmt ab 1. April die Kosten für Essen und Getränke während der Corona-Krise. Herzlichen Dank für die Arbeit, die rund um die Uhr geleistet wird.*"
Downplayed Risk	Tweet attempts to de-emphasize the potential risk of [COVID-19] or bring it into perspective. May also express a lack of conern or disinterest.	"*Corona ist die größte Herausforderung für Bayern seit dem 2.Weltkrieg. Die grundlegenden Ausgangsbeschränkungen werden eingehalten. Danke! Es war notwendig zu handeln. Wir sind vorangegangen, viele Länder sind gefolgt. Alle sind einig, dass wir soziale Kontakte verringern müssen. https://t.co/dKTmKcZ2AM"*"
Concern	Tweet expresses [COVID-19]-related fear, anxiety, worry, or sadness for self or others. May also express scepticism.	"*Die steigende Zahl der Todesfälle zeigt den Ernst der Lage. Wir sind nicht bereit, das so hinzunehmen. Das ständige Kleinreden und Leugnen von Corona ist ein Problem. Die Grundphilosophie heißt jetzt: Daheim bleiben.*"
Frustration	Tweet expresses anger, annoyance, scorn, or volatile contempt. May include coarse language.	"*Das Verbot der Berliner Corona-Demo am 29.8. ist ein Schlag ins Gesicht der Freiheit und des Rechtsstaats. Diese totalitäre Maßnahme darf und wird keinen Bestand haben!#Freiheit #COVID19 #Corona*"
Misinformation	Tweet contradicts the reference standard or contains unsubstantiated information. May make speculations or express distrust of authority or the media. May include conspiracy or doomsday theories.	"*Wie zuverlässig #Corona-Tests sind, beweist ein Abgeordneter der FPÖ in diesem kurzen Mitschnitt. Er nutzte seine Redezeit im Parlament, um einen Schnelltest zu machen - an einem Glas #Cola. Das ""überraschende"" Ergebnis sehen Sie im Video! #Lockdown #LockdownJetzt #AfD https://t.co/dNqvnA4Rxk*"
Question	Tweet asks a question or contains a question mark.	"*Was hat die Corona-Politik mit @Markus_Soeder gemacht, dass er die @fdp in die Nähe der #AfD rückt, weil wir an der Beteiligung der Parlamente bei Eingriffen in Grundrechte festhalten und die Wirksamkeit von Maßnahmen begründet sehen wollen? CL*"

G NETWORK ANALYSIS

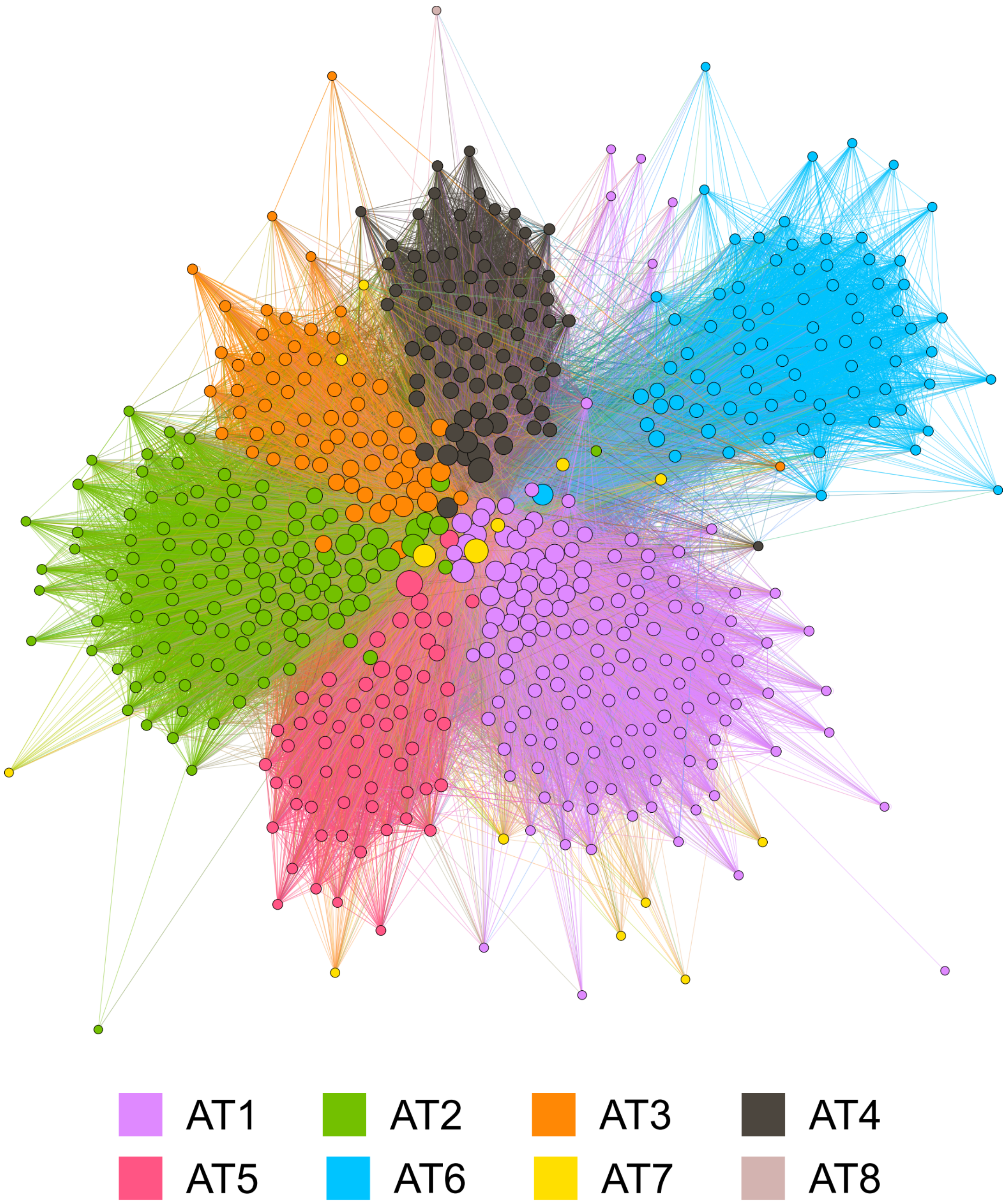

Figure 12: Communities detected by Infomap in the graph based on all original tweets by the politicians in the corpus. Colors represent individual communities. Nodes represent individual Twitter accounts, while edges represent follower relationships, replies, mentions,and quotes tweets. Graph laid out using the Fruchterman Reingold layout algorithm.

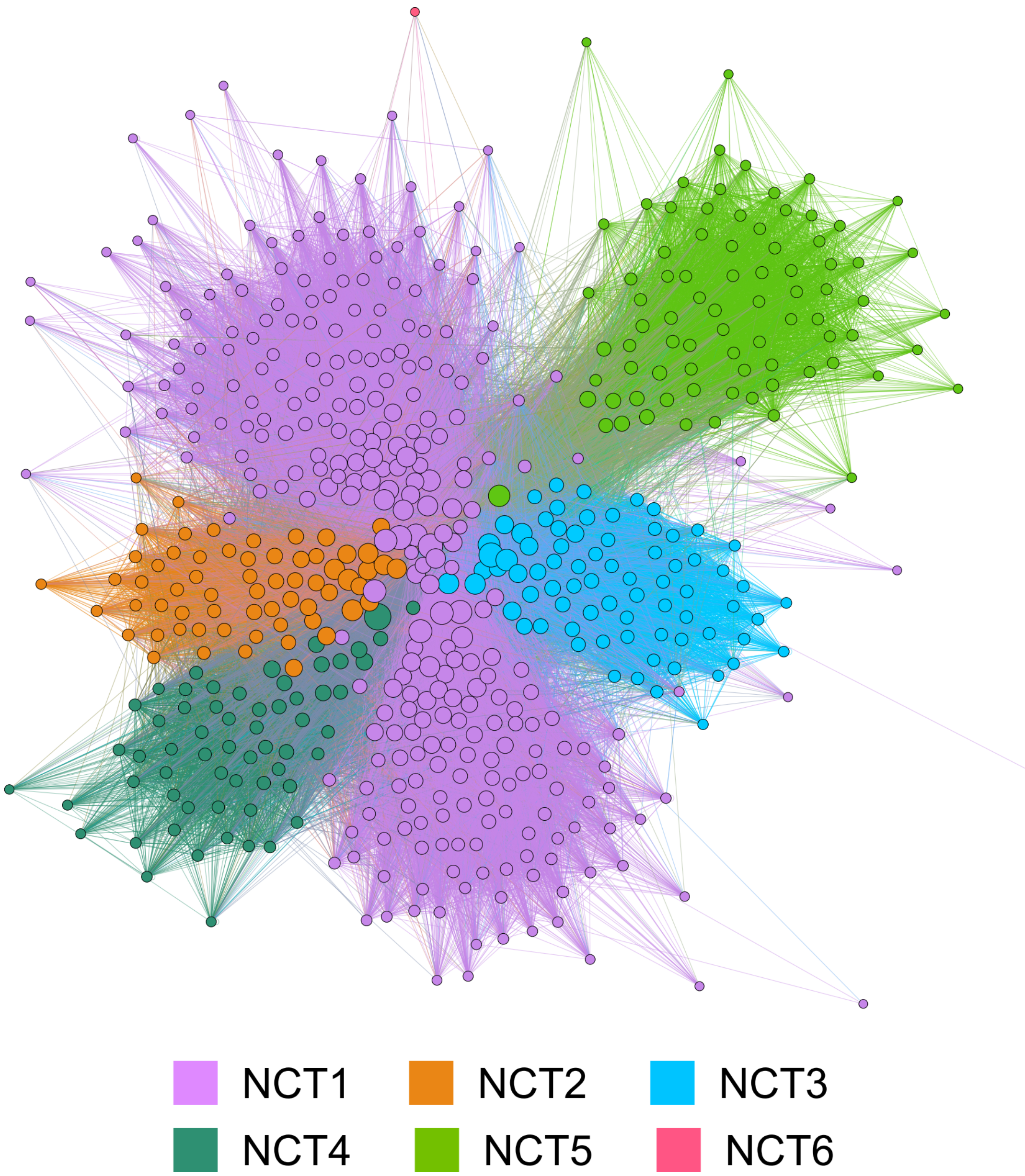

Figure 13: Communities detected by Infomap in the graph based on the non-COVID-tweets by the politicians in the corpus. Colors represent individual communities. Nodes represent individual Twitter accounts, while edges represent follower relationships, replies, mentions,and quotes tweets. Graph laid out using the Fruchterman Reingold layout algorithm.

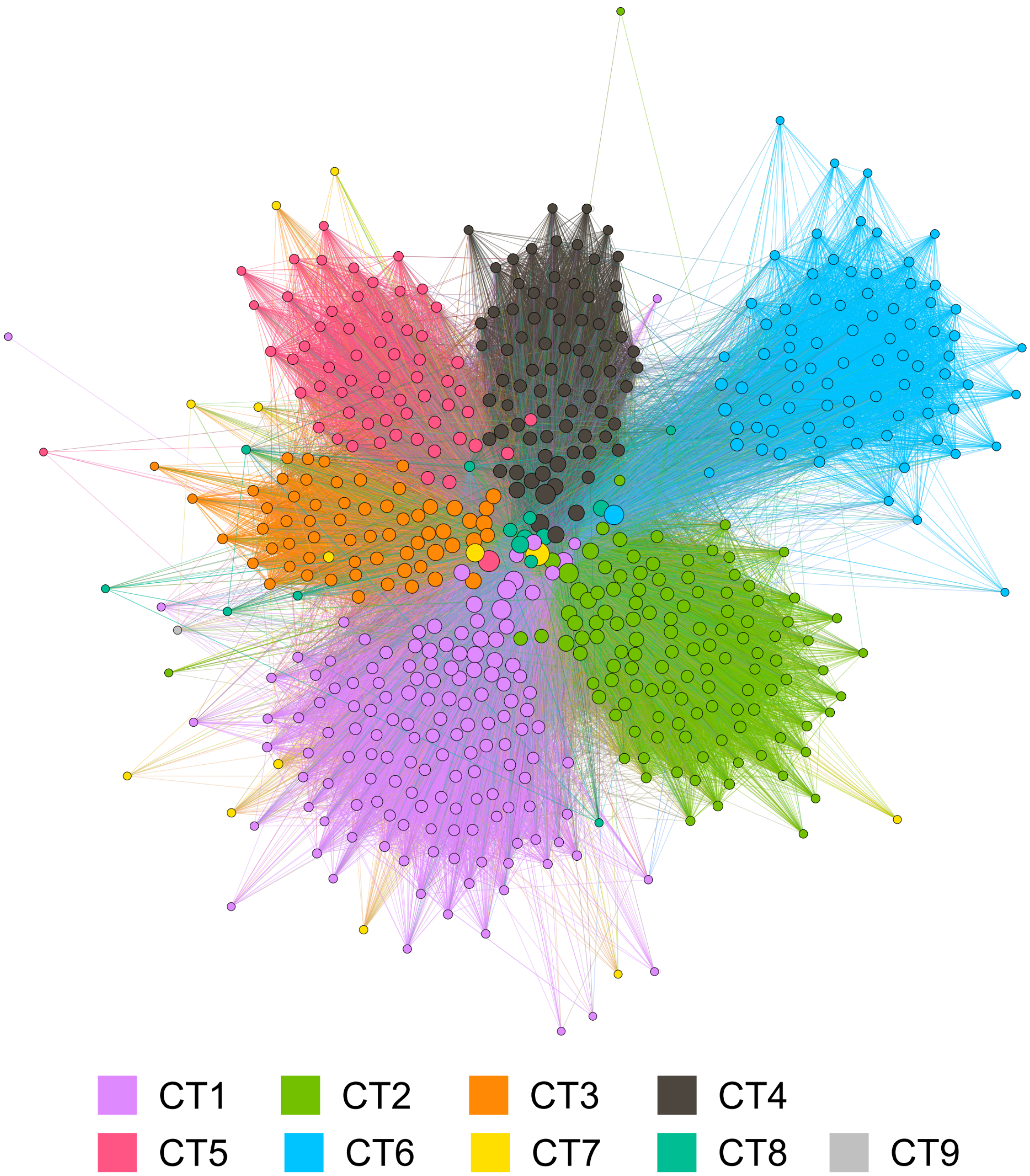

Figure 14: Communities detected by Infomap in the graph based on the COVID-tweets by the politicians in the corpus. Colors represent individual communities. Nodes represent individual Twitter accounts, while edges represent follower relationships, replies, mentions,and quotes tweets. Graph laid out using the Fruchterman Reingold layout algorithm.