Original Research Article

BIG DATA & SOCIETY

Big Data & Society
July–December: 1–11

DOI: 10.1177/20539517211041279
journals.sagepub.com/home/bds
SAGE

Different types of COVID-19 misinformation have different emotional valence on Twitter

Marina Charquero-Ballester*, Jessica G Walter*, Ida A Nissen and Anja Bechmann

Abstract

The spreading of COVID-19 misinformation on social media could have severe consequences on people's behavior. In this paper, we investigated the emotional expression of misinformation related to the COVID-19 crisis on Twitter and whether emotional valence differed depending on the type of misinformation. We collected 17,463,220 English tweets with 76 COVID-19-related hashtags for March 2020. Using Google Fact Check Explorer API we identified 226 unique COVID-19 false stories for March 2020. These were clustered into six types of misinformation (cures, virus, vaccine, politics, conspiracy theories, and other). Applying the 226 classifiers to the Twitter sample we identified 690,004 tweets. Instead of running the sentiment on all tweets we manually coded a random subset of 100 tweets for each classifier to increase the validity, reducing the dataset to 2,097 tweets. We found that only a minor part of the entire dataset was related to misinformation. Also, misinformation in general does not lean towards a certain emotional valence. However, looking at comparisons of emotional valence for different types of misinformation uncovered that misinformation related to "virus" and "conspiracy" had a more negative valence than "cures," "vaccine," "politics," and "other." Knowing from existing studies that negative misinformation spreads faster, this demonstrates that filtering for misinformation type is fruitful and indicates that a focus on "virus" and "conspiracy" could be one strategy in combating misinformation. As emotional contexts affect misinformation spreading, the knowledge about emotional valence for different types of misinformation will help to better understand the spreading and consequences of misinformation.

Keywords

Pandemic, infodemic, sentiment analysis, social media, disinformation, fact-checking

Introduction

"We're not just fighting an epidemic; we're fighting an infodemic"[1], the director general of the WHO declared in February 2020, highlighting for example, facts instead of fear, and rationality instead of rumors. Since then, the spread of false and misleading information about COVID-19 on social media has only intensified (Brennen et al., 2020). This potentially stokes public anxiety, which might further escalate into collective panic or other negative collective behaviors that health organizations want to avoid. Such concerns build on, for instance, survey studies that find a negative relationship between health-protective behaviors of citizens and exposure to COVID-19 misinformation (i.e. conspiracy beliefs) and that such exposure is higher when using social media as a source of information (e.g. Allington et al., 2020). Existing studies have also added understandings of how misinformation from debunked stories in general spread faster than non-debunked stories on Twitter (Vosoughi et al., 2018) and detailed accounts of the pattern and logics of spread have been made in particular on specific COVID-19 conspiracy theories and associated hashtags (e.g. Gruzd and Mai, 2020), hinting at the importance to study and detect misinformation. In addition, studies have examined emotions in social media at scale during the pandemic (Xue et al., 2020), however, without a focus on misinformation. Furthermore, studies show that negative sentiments related to communication about conspiracies and science, spread over time, where communication about conspiracy is more

DATALAB – Center for Digital Social Research, Aarhus University, Denmark
*These authors contributed equally to this paper.

Corresponding author:
Anja Bechmann, DATALAB – Center for Digital Social Research, Aarhus University, Helsingforsgade 14, Aarhus N, Midtjylland, Denmark.
Email: anjabechmann@cc.au.dk

negative (Zollo et al., 2015) and that negative information can spread for a longer time in comparison to neutral or positive information (Kušen and Strembeck, 2018). Emotions also play an important role in misleading the reader with false information and in detecting misinformation (Ghanem et al., 2020). Thus, veracity and emotional valence affect how information spreads, to what extent receivers are deceived and which sentiments are elicited. Knowledge about valence also helps to detect false information. Regarding COVID-19 misinformation, however, knowledge is missing about emotional valence associated with different types of misinformation. Our aim is to fill this gap and contribute to a better understanding of COVID-19 misinformation by looking at the extent of misinformation and associated emotional valence as one potential factor influencing spreading on social media. Our focus on COVID-19-related misinformation and its emotional valence is sparked by inconsistencies on the amount and character of misinformation and emotions in the context of COVID-19 (e.g. Brennen et al., 2020; Kouzy et al., 2020; Singh et al., 2020). We examine a similar period as these studies, but contribute to the field with a more content-sensitive perspective.

We supplement such accounts with a broader understanding of emotional valence of misinformation content related to COVID-19 on Twitter and investigate, if we empirically see any differences in this valence depending on the type of misinformation. Our study: (1) is based on all English fact-checked stories from Google Fact Check Explorer in March 2020, (2) coded these stories manually into types of misinformation to increase the sample size per type for a reliable sentiment analysis, (3) includes the creation of a classifier for each story and its application to tweets from March for related COVID-19 hashtags, and (4) is continued by the manual coding of a random subset of the tweets for misinformation to increase sensitivity (finding true positives) and specificity (avoiding false positives) before (5) reporting the measure of the emotional valence scores and differences between types.

The results of the study in this paper will be a first step to understand how the misinformation content itself might ignite different emotional valence when it spreads on social media. Even though tools to automatically detect misinformation are still improving, detecting misinformation in real time is likely to remain a significant and enduring challenge due to the high velocity, volume and variety of dis- and misinformation. Thus, a better understanding of any of its contexts and potential drivers of diffusion is essential for minimizing its potential impact on society. A better understanding of emotional valence associated with different types of misinformation can help addressing misinformation more specifically.

Theoretical background

Despite careful conceptual work done by scholars to identify differences between mis- and disinformation and the contested and political term "fake news" (Bechmann and O'Loughlin, 2020; Buning et al., 2018; Farkas and Schou, 2019; Kalsnes, 2018; Tandoc et al., 2018; Wardle and Derakhshan, 2017), we collapse them here and use these concepts as synonyms in this paper. Thus, we apply a general definition of misinformation and disinformation inspired by Buning et al. (2018) "false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit." However, misinformation is not necessarily created with the intention of causing harm or exerting influence (e.g. Carmi et al., 2020). In this paper, we emphasize what appears in the Google Fact Check Explorer API that was debunked as false, inaccurate or misleading from a fact-checkers point of view without taking intentions into consideration. Google Fact Check Explorer is an index site that includes verified fact check organizations' work on "re-reporting and researching the purported facts in published/recorded statements made by politicians and anyone whose words impact others' lives and livelihoods. Fact checkers investigate verifiable facts" (Graves, 2018: 615). Our conceptual understanding of misinformation, thus, is situational and expert/fact-checker driven.

The role of social media in the pandemic

We focus on misinformation spread on the social media platform Twitter due to its accessibility for systematic analysis at scale (Bruns et al., 2018; Freelon, 2018; Møller and Bechmann, 2019). Despite a lower penetration rate than Facebook and YouTube in most countries, Twitter is a widely used communication channel, which is used during the pandemic to share information or experiences and is one of the channels in which misinformation related to COVID-19 spreads (Singh et al., 2020). In general, Singh et al. (2020) describe communication about COVID-19 on Twitter (mid-January to mid-March 2020) and find that the two most common topics are either related to the virus, health itself or the global nature of the pandemic. Allington et al. (2020) show that people, who use social media to look for COVID-19-related information more extensively, are more prone to believe conspiracy theories and to a larger extent refuse health-protective behavior indicating a potential threat of social media use in the pandemic. We extend these studies with a focus on which COVID-19-related misinformation tweets contain positive, neutral, or negative emotions. This will help to better understand emotional valence as a potential factor in the spreading of misinformation on social media, thereby a small step to possibly help developing strategies for preventing negative behavior in the future.

Some studies already focus on misinformation related to COVID-19. In order to quantify misinformation related to COVID-19, Kouzy et al. (2020) select random samples of tweets within hashtags related to the pandemic and find

that ~25% of the analyzed tweets include misinformation, whereas they do not report the extent of misinformation of their original sample. Findings from Singh et al. (2020) suggest that tweets related to COVID-19 refer to misinformation only to a small extent. The discrepancies between these two pre-published studies are probably due to methodological differences in the identification of misinformation. This leaves open to what extent misinformation—especially regarding COVID-19—really spreads on Twitter. With our analyses, we add information about the extent to which COVID-19-related misinformation is present within Twitter. Regarding the creator or spreader of misinformation on Twitter, Brennen et al. (2020) show that most misinformation on social media related to COVID-19 come from ordinary people, but engagement is higher with misinformation coming from prominent public figures such as politicians and celebrities. Looking at debunked stories in March 2020, they also find that most misinformation is related to actions of public authorities or the spread within communities, and that misinformation deals less with the origin of the virus or public preparedness. Based on the findings of Brennen et al. (2020), our focus on misinformation identified on the basis of debunked stories of fact-checking organizations aims at picking the most influential misinformation as fact-checking organizations tend to debunk misinformation either stated by public figures, spreading on social media or a combination thereof.

The role of emotions for misinformation

A previous study show that misinformation spreads faster and wider than information (Vosoughi et al., 2018), with differences in emotional content contributing to an explanation of these differences in the spread. Looking at the role of emotions in misinformation, previous studies also suggest that sharing political misinformation is strongly related to a desire to show a clear political stance, and often originate in and/or elicit emotions such as anger and mistrust (Osmundsen et al., 2021; Weeks and Garrett, 2019). Apuke and Omar (2020) show based on survey data from Nigeria that a strong motivator for sharing COVID-19 disinformation is altruism. We therefore expect public health emergencies such as the current COVID-19 pandemic to lead to a larger variety of emotional expressions in tweets since related misinformation is shared out of altruistic reasons to some extent, for example regarding potential preventions. However, Han et al. (2020) also find that anger—as an emotion with negative valence—leads people to believe in and to spread misinformation related to COVID-19.

A study by Martel et al. (2020) shows that positive and negative emotions also make people more prone to believe in fake news and less able to distinguish fake from real news, thereby providing some evidence that emotional valence of tweets could affect the spread of misinformation. Further studies using social media data support this link between spread and sentiment, with one study showing that negative information spreads for a longer time than neutral or positive (Kušen and Strembeck, 2018). Another study shows that communication about misinformation and science also becomes more negative over time with misinformation being more negative in general (Zollo et al., 2015). Our aim is to investigate whether different types of misinformation tweets related to the COVID-19 crisis differ in the emotional valence. Emotional valence is one of the two dimensions—besides arousal—on which emotion can be located. It equals "a bipolar dimension of affective valence—from attraction and pleasure to aversion and displeasure" (Lang, 1995: 374).

In this article, we examine emotional valence in tweets related to COVID-19 misinformation by using sentiment analysis as it has been used in other contexts to detect or examine misinformation (cp. Ajao et al., 2019; Kušen and Strembeck, 2018; Zollo et al., 2015). Regarding health-related misinformation, Featherstone and Zhang (2020) show in an experiment, that misinformation about vaccines elicits anger. Kumar et al. (2020) analyze sentiments in tweets related to COVID-19 on three days in March 2020 and find that tweets with positive sentiments outnumber those with negative sentiments. Furthermore, regarding tweets with a positive sentiment, they find a considerable number of tweets expressing trust and less expressing joy. Regarding tweets with a negative sentiment, the most expressed sentiment was fear (also cp. findings from Medford et al., 2020) and surprise was the least common sentiment. Also Li et al. (2020) pre-published their findings about sentiments in Twitter and Weibo communication about COVID-19. They distinguish six main emotions sadness, anger, disgust, worry, happiness, and surprise, and find that in the period from end of January to mid-May 2020 the intensity of emotions decreased over time regarding communication on the Chinese platform Weibo. Here, the most intense sentiment was worry followed by sadness and anger. On Twitter, they find an increase in sentiment intensity after the beginning of March 2020. They also look at triggers of these sentiments and identify events and topics around the issues of testing positive and death triggering sadness, the issues of shutdown, quarantine and mandatory rules triggering anger and issues related to jobs, getting infected, payments and families triggering worry.

However, Kumar et al. (2020) as well as Li et al. (2020) do not focus on tweets related to misinformation but on communication about COVID-19 in general. Thus, our study contributes to this research gap, by looking at misinformation in particular. We expect that the emotional valence of COVID-19-related misinformation tweets depends on the type of misinformation as also Li et al. (2020) found a relation between topic and emotion.

However, it is difficult to predict to what extent we will find emotional content of any valence, as existing studies do not report the amount of neutral information or do not rely on Twitter or COVID-19-related misinformation. Zollo et al. (2015), for example, report the amount of positive, negative, and neutral communication related to conspiracies and science and find in their Facebook dataset a large amount of negative and neutral information, with positive information being much less common. Thus, we also expect that a large part of the conversation about COVID-19-related misinformation is neutral. Regarding the direction of valence, the discussed topic might be crucial.

Method

We extracted misinformation debunked by fact-checking organizations via the Google Fact Check Explorer. Google Fact Check Explorer is one of the most widely used fact-check index sites by fact-checkers (Walter et al., 2020) and it is especially relevant to apply to research data as it contains an API solution that allows for research to legally extract data without violating IP rights. We also chose the Google Fact Check Explorer because it was central to English language misinformation. The tool is mainly used by fact-checkers to publish debunked disinformation stories. We use Google Fact Check Explorer via its API to extract a baseline dataset of all English language misinformation related to COVID-19 from March 2020 (that had a claim or review date between the 1st and 31st of March), where the pandemic started taking off in Europe and the United States. This dataset includes 247 debunked stories, and after removing duplicates 226 debunked stories remained. We included all stories irrespective of the rating result (e.g. misleading or false).

Our initial sample of tweets was obtained through a public coronavirus Twitter dataset containing over 123 million tweets, with over 60% of them in English collected through 76 hashtags related to COVID-19 (Chen et al., 2020).[2] Due to our focus on misinformation, we decided to narrow down the sample to March 2020, when the pandemic was globally widespread and the amount of misinformation was high due to still being in the early stages of the crisis and to make our analysis comparable to existing research. After selecting tweets in English language and removing retweets, we reached a total sample of 17,463,220 tweets.

We identify all tweets related to the misinformation stories using manual keywords and select tweets according to their overlap with the keywords. We manually extract the most central single word or bigram of each debunked story title by two annotators independently. For 89% of the stories, the two annotators extracted identical keywords. In case of disagreements, the final keyword was selected in discussions. The first keyword was supplemented by an exhaustive keyword list for each story, containing all relevant words of the story title (excluding e.g. stop words, repetitions) (similar to keywords selection by e.g. Allcott et al., 2019). Those secondary keywords were also selected by the two annotators. Examples for extracted keywords and keyword lists are presented in Table 1 and the first keywords of all stories are presented in the supplemental material (Tables A1a to A1f). The first keyword and the secondary keywords were then used as a classifier representing each story in the final Google Fact Explorer dataset.

We ran the classifiers for each story on the Twitter dataset in order to identify the tweets talking about the misinformation in two steps. In order to filter the data, the first keyword was used in a first step to identify relevant tweets through string matching. Only tweets that contained the first keyword were selected for the second step. In the second step, false positives (tweets that were captured but are not related to the debunked stories) were filtered out based on the second keywords. At least one word from the list of secondary keywords was required for selection. In the end, the selected tweets contained the first keyword and one or more of the secondary keywords. In comparison to a solely automated approach, this two-step procedure facilitates data processing by reducing processing time by filtering a large number of tweets in the first step. After having filtered the result through the classifiers to identify all tweets that discuss misinformation related content (comments have not been available in the streaming API of Twitter) we are reducing the dataset to 690,004 tweets.

Assigning type of misinformation

Each of the debunked stories within the final Google Fact Check Explorer dataset was assigned to one of six types of misinformation. The code book of the types of misinformation were made bottom up by two researchers from the stories at hand and were labeled "cure, prevention & treatment" (shortly "cures"), "conspiracy," "political measures" (shortly "politics"), "vaccine & test kits" (shortly "vaccine"), "virus characteristics & numbers" (shortly "virus") and "other," and represent different kinds of COVID-19-related misinformation. The clustering was inspired by a typology of narratives provided by the EU DisinfoLab (2020): health fears, conspiracy theories, lockdown fears, false cures, identity, societal and political polarization. Due to the sensitivity to the data at hand, we combined misinformation related to "lockdown fears" and "identity, societal, and political polarization" into "political measures." We extended the typology with the category "vaccine & test kits" as specific health-related misinformation. We also added the category "virus characteristics and numbers," which contains stories related to the spread and characteristics such as symptoms and transmission. "Health fears" were assigned to the "other" category due

Table 1. Types of misinformation.

Type of misinformation	Example in Google Fact Check Explorer database	Example debunked by	1st keyword	2nd keyword list
Cure, prevention, & treatment (18.6%)	"Eating bananas is a preventative against the COVID-19 coronavirus disease."	Snopes.com	bananas	eating; preventative; against; coronavirus
	"A Facebook user suggested that palm oil can stop the spread of Covid-19."	Dubawa	palm_oil	stop; spread
Conspiracy (14.6%)	"Russian Ministry of Health confirmed in a document that the novel coronavirus, COVID-19, is man-made."	BOOM	man-made	Russian; Ministry; Health; confirmed; document; corona-virus; COVID-19
	"Dean Koontz's novel predicted the COVID-19 pandemic."	The Logical Indian	Koontz	Dean; novel; predicted; COVID-19; pandemic
Political measures (23.5%)	"Zimbabwe's schools and colleges, which closed on March 24 over the COVID-19 crisis, will re-open in September 2020."	ZimFact	Zimbabwe	schools; colleges; closes; March; re-open; September
	"Residents in India to stay indoors during nighttime COVID-19 disinfectant spraying"	AFP Fact Check	disinfectant	residents; India; stay; indoors; nighttime; during; spraying
Vaccine & test kits (9.7%)	"Vaccine for COVID-19 is now available."	The Logical Indian	vaccine	now; available
	"A US biotech company has successfully created a vaccine for COVID-19"	BOOM	vaccine	US; biotech; company; successfully; created
Virus characteristics & numbers (19.0%)	"The virus causing COVID-19, SARS-CoV-2, stays in the throat for four days before going to the lungs."	VERA Files	throat	virus; stays; four; days; before; lungs
	"COVID19 could kill 45 million Nigerians."	AFP Fact Check	Nigerians	COVID-19; could: kill; 45; million
Other (14.6%)	"Dolphins and swans return to Italy waterways after Covid-19 lockdown."	India Today	dolphins	swans; return; Italy; waterways; after; lockdown
	"Cristiano Ronaldo will turn his hotels into hospitals for COVID-19 patients."	AFP Fact Check	Ronaldo	turn; hotels; into; hospitals; COVID-19; patients

Note: Percentage of debunked stories is given in parentheses in the first column.

to a low number of related stories. The "other" category contains, in addition, stories related to individual fates, economy or criminal behavior and stories that cannot be assigned to the remaining five categories. The coding scheme for the typology can be found in the supplemental material (Table A9). Two independent raters assigned the stories to the six categories and agreed upon a category, if the rating differed. The coding was validated by computing the inter-rater reliability (Krippendorff, 2018), which was acceptable above 0.7 ($\alpha = 0.74$). Table 1 presents the percentage of stories assigned to each category and examples.

Assigning tweets to types of misinformation

After having clustered all stories into the six types of misinformation, we assigned the tweets identified by the classifiers to the six categories. However, some of the classifiers were better than others due to the generic character of the stories and subsequent keyword classifiers (e.g. vaccine, virus) and therefore the ratio of misinformation/not misinformation varied to a high degree (Tables A1a to A1f in the supplemental material documents variance in all 226 classifiers when manually checking the subset of tweets, Table A10 provides an overview). To increase the validity of the study by reducing false positives, we needed to manually select a subset of tweets that was surely positive rather than running the analysis of emotional valence on tweets containing supposedly claims unrelated to the debunked stories. In order to do so, we randomly selected 100 tweets for each of the 226 debunked story or, if less tweets were assigned to a story, all related tweets. We then deleted all remaining unrelated tweets and recurring tweets manually. This filtered sample provided us with 2097 tweets for the sentiment analysis. Figure 1 illustrates the databases, the filtering process and also reports how many tweets were finally assigned to each type of misinformation.

Measuring emotional valence through sentiment analysis

After curating the dataset to contain only those tweets referring to the debunked stories, we measure the emotional valence of the tweets through sentiment analysis. The aim of sentiment analysis is the application of "automatic tools able to extract subjective information from texts in natural language, such as opinions and sentiments, so as

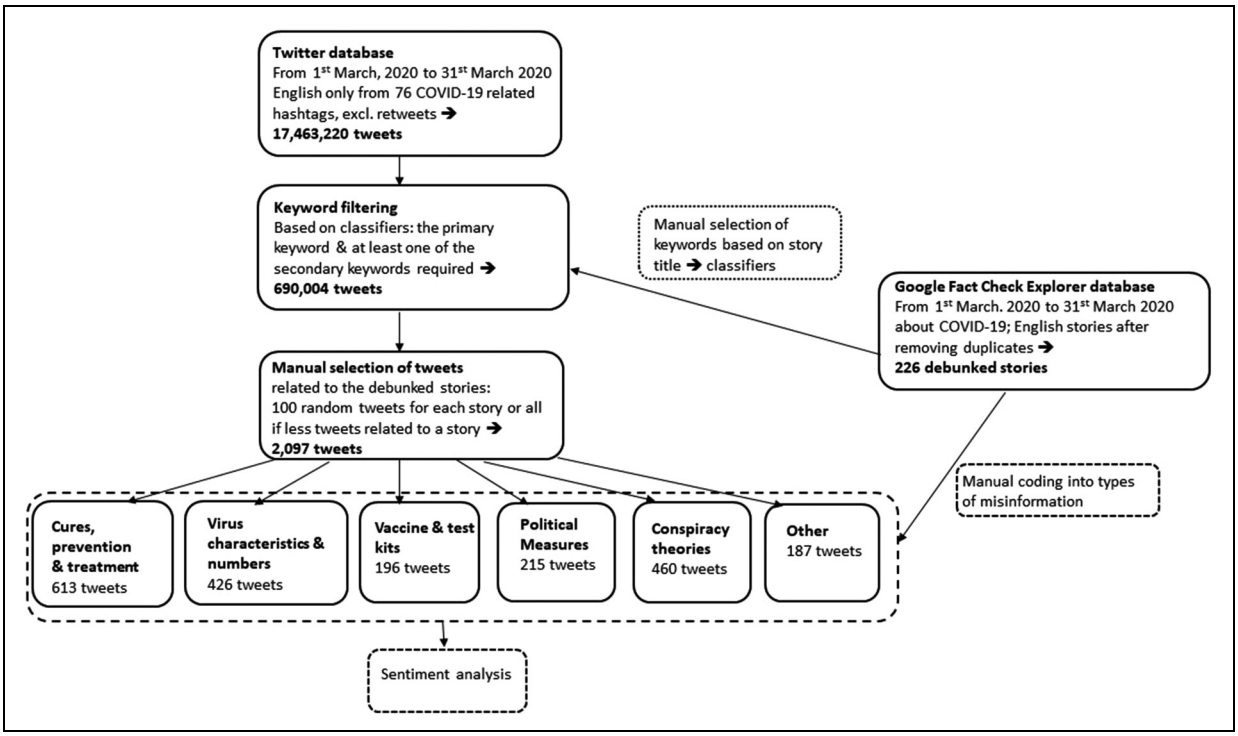

Figure 1. Illustration of filtering process, methods and distribution on types of misinformation.

to create structured and actionable knowledge" (Pozzi et al., 2017: 1). That is, "sentiment analysis attempts to conclude individual sentiments based on expressions provided with the help of any language to portray their emotions" (Pawar et al., 2016: 34). Sentiment analysis is used, for example, to get insights about sentiments about products and services or to understand voter behavior and has been applied to several domains such as health care, financial services, and politics (Pozzi et al., 2017). Sentiments are often analyzed using text messages, but analyses can be applied multimodal (Soleymani et al., 2017). We use sentiment analysis to detect underlying emotions that are expressed in the tweets of COVID-19 misinformation. We applied VADER (https://github.com/cjhutto/vaderSentiment), a rule-based model for general sentiment analysis (Hutto and Gilbert, 2014) to the entire dataset. This model is based on a list of lexical features specifically attuned to sentiment analysis, providing scores for sentiment valence. We chose this model because it is specifically made for microblog-like contexts, addressing and showing good performance in these types of contexts and their associated challenges (e.g. shortness of text and a use of abbreviated language conventions to express sentiments). The model produces four different scores: positive, negative, neutral, and a final compound score. Positive, negative, and neutral scores are ratios for proportions of text that fall in each category. The compound score is computed by summing the valence scores of each word that appears also in the lexicon, adjusted according to the rules, and normalized to be between −1 (most negative) and +1 (most positive). This last measure provides a single score of sentiment for any given sentence or tweet, making it the most useful measure to evaluate overall sentiment scores at the whole sample level as well as any possible differences between types of misinformation.

Statistical analysis

We conducted statistical analyses using permutation testing to evaluate differences in valence between specific types of misinformation (e.g. disease prevention vs political measures). We used the compound scores for each type of misinformation and ran 10,000 permutations (two-sample *t*-tests) to test against the null hypothesis: the compound scores for the different types of misinformation do not differ between each other. Lastly, we corrected across multiple comparisons using the false discovery rate by Benjamini and Hochberg (1995).

Results

Overview of tweets related to misinformation

The overview in the supplemental material Table A1a to A1f displays the number of selected tweets per keyword. The overview in the supplemental material Table A10 provides numbers by type of misinformation. Filtering the Twitter dataset according to the overlap with keywords

resulted in an average of 3053 ± 20,027 tweets per debunked story (average ± standard deviation; range: 0–217,182; in total: 690,004). After the selection of tweets that related to the debunked stories based on the keywords, 195 of the 226 debunked stories could be associated with the tweets. On average, for the 195 debunked stories we had 3538 ± 21,527 tweets per story (range: 1–217,182). The percentage of tweets that were manually confirmed to be related to the debunked story varied widely across the different stories: the claim "Plasma from newly recovered patients from COVID-19 can treat others infected by COVID-19" (debunked by Full Fact) had 78% out of 100 analyzed tweets, whereas the claim "Tea can cure or alleviate novel coronavirus (COVID-19) infection" (debunked by AFP Fact Check) had no related tweets out of 100 analyzed. The average was 9.3 ± 16.6 related tweets (range: 0–89). In total, we collected 2097 tweets that related to the debunked stories.

A manual selection of tweets that related to misinformation was necessary, as some keywords collected many related stories whereas others collected mostly unrelated stories. Keywords that were specific for a well-defined concept, often captured tweets discussing the misinformed claim (such as "RFID" 7/9 (78%), "secret terrius" 62/100 (62%), and "Vitamin C" 89/100 (89%)). Unless this keyword appeared in a different much-discussed context that was unrelated to the claim, in those cases the number of selected tweets was very low (such as "Netherlands" 4/100 (4%), "Amazon" 4/100 (4%), and "sun" 2/100 (2%)).

How much misinformation do we observe?

The selection based on keywords yielded 690,004 tweets, out of 17,463,220 tweets. About half of the debunked stories had none or less than 100 tweets (125 of 226). The selected tweets together with copies (that were excluded for further analysis) made up 29% of the manually assessed tweets, and roughly extrapolating would mean that about 100,000 of the nearly 17.5 million tweets were related to misinformation that we captured. This means that generally, the amount of misinformation shared on Twitter is very small compared to other existing studies.

Is the discussion about misinformation related to emotion?

All tweets discussing the misinformation claims taken together had an average compound score of −0.0151. Most tweets were neutral (86.15%) and the proportion of positive (6.85%) or negative (6.99%) valenced tweets was similarly small. The neutral tone of the tweets could possibly be explained with the reasoning that tweets often stated facts without adding emotional words. Additionally, the positively and negatively loaded tweets balance each other out in the entire dataset. Following this line of reasoning, certain types of misinformation could differ from a neutral sentiment by evoking one predominant emotional valence. To test this hypothesis, we analyzed the sentiments of the six different types of misinformation. Thus, even if conversation about COVID-19 misinformation does not tend towards negative or positive valence and emotional content is less present in our dataset as in the Facebook dataset analyzed by Zollo et al. (2015), emotional valence can still matter if they differ between different types of COVID-19-related misinformation.

What are the sentiments in the different topics?

As shown in Figure 1, the distribution over the six types of COVID-19-related misinformation tweets was as follows: cures: $n = 613$, virus: $n = 426$, vaccine: $n = 196$, politics: $n = 215$, conspiracy: $n = 460$, and other: $n = 187$. While the sample sizes are not equal, the statistical tests used (i.e. permutation testing) do not suffer from a bias resulting from this fact, so the results preserve statistical validity.

Table 2 displays the compound score and permutation testing for the different types of misinformation. The compound scores had a difference in valence for the different types. Namely, the types "virus" and "conspiracy" had a negative compound score (−0.124 and −0.098, respectively), meaning that they had a negative valence. The other types of misinformation had a positive compound score and hence positive valence, with "cures" being most positive (0.073), "other" and "vaccine" slightly less positive (0.050 and 0.54, respectively) and "politics" almost neutral (0.007).

Several individual comparisons of the compound scores between the types of misinformation were significantly different. The negative valenced types "conspiracy" and "virus" both significantly differed from the positive valenced types "cures" (both: p-value = 0.0004), "other" (both: p-value = 0.0004), "politics" (p-value = 0.0007 and 0.0004, respectively), and "vaccine" (both: p-value = 0.0004).

Conspiracy-related misinformation might be more negative than misinformation related to other types of misinformation since, especially shortly after dramatic events, conspiracies elicit a negative emotional response, and a higher emotionality in dramatic situations potentially drive people towards conspiracies. Furthermore, in general emotions contribute to the spreading of conspiracies (Samory and Mitra, 2018; Sunstein and Vermeule, 2009). Misinformation related to virus characteristics and numbers are probably associated with a high uncertainty and especially in the beginning of the pandemic also fuel anxieties about a potentially deadly disease and are therefore especially negative.

Misinformed claims about false cures for COVID-19 are potentially damaging to society by fostering reckless behavior and thereby advancing the spread of the disease.

Table 2. Compound score results for comparisons of different types of COVID19-related misinformation.

Compound scores	Conspiracy −0.098	Cures 0.073	Other 0.050	Politics 0.007	Vaccine 0.054	Virus −0.124
Conspiracy −0.098		1	1	1	1	0.3987
Cures 0.073	0.0004		0.5882	0.0617	0.5882	0.0004
Other 0.050	0.0004	1		0.3537	1	0.0004
Politics 0.007	0.0007	1	1		1	0.0004
Vaccine 0.054	0.0004	1	0.9177	0.2526		0.0004
Virus −0.124	1	1	1	1	1	

Note: *p*-values corrected for multiple comparisons using false-discovery rate correction according to Benjamini and Hochberg (1995)

Nonetheless, the associated sentiment with potential cures and vaccines for COVID-19 had an overall positive valence. The tweets with positive valence contained words related to hope (e.g. help, treat, progress). Even though this type of misinformation expressed positive emotions, the effect on society is not necessarily positive but probably rather negative as it can enhance transmission of the disease.

The tweets related to politics were about government regulations mainly for curbing the spreading of COVID-19. Here the limiting consequences in people's daily lives in combination with a feeling of hope for things to get better as a result, or the fact that most regulations stated were just reported as facts, might explain the neutral valence. Tables A2 to A7 in the supplemental material list the most negative and positive tweets (anonymized) for each type of misinformation to illustrate the expressed sentiments.

Continuing the line of the previous reasoning, positive sentiments relating to hope for cures would naturally produce stronger positive sentiments than the rather neutral fact-relating sentiments for political measures.

In order to have a clearer overview of the factors driving the results in the compound scores, we computed the mean for positive, negative, and neutral scores and each of the types of misinformation. Figure 2 presents the average for the sentiment valence for each type of misinformation. We observed that in every type of misinformation, the neutral score was above 0.82, highlighting the predominance of neutral language independently of misinformation type. However, if we consider the most common threshold values used to classify sentences as neutral as stated by Hutto and Gilbert (2014) in the VADER sentiment analysis repository (https://github.com/cjhutto/vaderSentiment) only politics would qualify as a neutral type of misinformation based on its mean (i.e. compound score >−0.05 and <0.05). This indicates that, despite the high neutral scores, all the other types of misinformation contain some positive and/or negative words driving the differences between types of misinformation. With regard to our results regarding misinformation around conspiracy theories and the virus characteristics and numbers, we observe that in both cases the language is still mainly neutral, but the ratios for positive and negative words show a more marked imbalance than in the other types of misinformation. Furthermore, we observe that misinformation around cures contains a relatively higher percentage of both positive and negative words than other types, but leans more towards positive words.

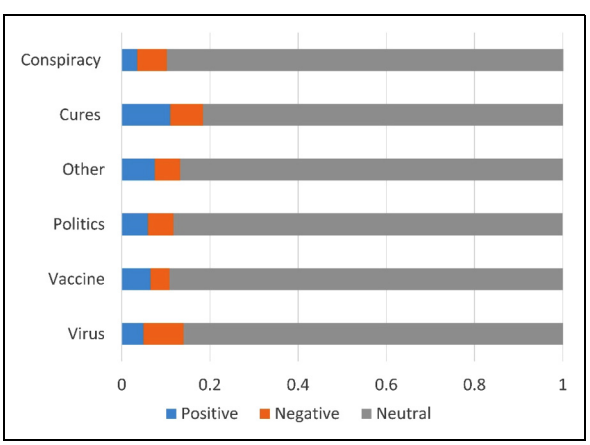

Figure 2. Average for the sentiment valence within each type of misinformation.

Discussion and conclusion

Existing studies examine relations between health-unrelated misinformation and emotions (e.g. Vosoughi et al., 2018), and misinformation related to COVID-19 without a focus on emotions (e.g. Singh et al., 2020) or examine emotions but without a focus on misinformation (e.g. Kumar et al., 2020). In this article, we add to those studies by showing how the emotional valence of COVID-19-related misinformation within Twitter differs by type of misinformation, thus looking at health-related misinformation, emotions, and adding the aspect of types of misinformation. We add knowledge about (a) the extent of COVID-19 misinformation, (b) the extent of emotional content regarding COVID-19 misinformation, and (c) emotional valence associated with different types of COVID-19 misinformation. We identify 2097 tweets related to misinformation debunked by fact-checkers and being accessible via the Google Fact Check Explorer. Within the tweets we analyzed, we found ~29% are related to misinformation, however, extrapolating this finding to our starting sample (nearly 17.5 million

tweets) leads to the estimation that only a very small number of the tweets relate to misinformation. Overall, the tweets included in the final dataset, do not show a clear positive or negative emotional valence, but only a slight tendency towards a negative emotional valence. However, looking at different types of misinformation, we found significant differences. Communication around misinformation related to "conspiracy" and "virus characteristics and numbers" is characterized by stronger negative emotional valence than misinformation related to "cures, prevention, and treatment," "vaccine and test kits," "political measures," or "other."

Our findings have several implications, namely that they support the argument for a more differentiated analysis of COVID-19-related misinformation. We suggest that strategies for fighting COVID-19 misinformation should focus on a fast response to misinformation regarding conspiracies and virus characteristics, as well as on reported numbers, given that previous research has shown that emotional valence potentially affects the spreading of (mis)information (cp. Kušen and Strembeck, 2018; Vosoughi et al., 2018). As emotional valence can also help to detect misinformation (Ghanem et al., 2020), the knowledge about emotional valence associated with different types of misinformation might also help to adjust detection approaches depending on the type of misinformation. Furthermore, our findings showed that communication about cures, prevention, and treatments has a more positive emotional valence. This finding raises new questions and it would be interesting for future studies to investigate further the link between misinformation with positive emotional valence and its consequences for society—that is, is misinformation with a positive language, which potentially elicits positive emotions, particularly dangerous because it leads to health risking behavior?

Furthermore, it would be interesting in future research to address some of the limitations of our study. Using our keyword classifier, only 29% of the tweets were actually related to misinformation, which led us to complement our approach with manual coding. The reliance on manual detection hinders to some extent misinformation detection in real time. In the future, we plan to improve our classifier to create a more comprehensive sample of misinformation in a shorter period of time, which could be used in future studies to detect misinformation even more efficiently by reducing the amount of required manual coding.

Even though our analyses relied on a broad collection of tweets across many relevant COVID-19-related hashtags, we do not assume that we capture all misinformation published on Twitter (e.g. on even more specific or unrelated hashtags or without hashtags). As our aim was to understand the sentiment around different types of misinformation, the analysis is not too sensitive to not having all misinformation tweets available, but a future aim could be to compare our results with results from approaches, which collect data unrelated to any hashtags. As previous studies indicate a link between emotions and the spread of misinformation without a focus on COVID-19 misinformation (e.g. Vosoughi et al., 2018), future studies could also clarify the relation between emotional valence and the spread of COVID-19-related misinformation.

Our study is also limited in the sense that we cannot measure readers' immediate emotional response to COVID-19 misinformation, but only observe emotions expressed in the tweets. Future studies aiming to gain a better understanding of the relationship between the emotional responses to misinformation and their impact on society should consider analyzing Twitter comments in response to misinformation stories. Our approach to detect misinformation based on fact-checked stories probably also contributes to the in general neutral tone of the analyzed communication. Debunked stories are usually phrased neutrally without rhetoric (Graves, 2018). Since emotional framed information is more likely to elicit emotions (e.g. Ferrara and Yang, 2015; Zollo et al., 2015), we might underestimate the differences in emotional valence and the extent of emotions in general by using this neutral baseline. Thus, the differences we do find are probably even more reliable. It should also be mentioned that our classifier does not require a strict match with the debunked stories and therefore captures more emotional valence than the original debunked story.

Furthermore, another implication of identifying misinformation based on debunked stories from fact-checkers, which focuses on widely spread misinformation, is that we potentially examine emotional valence only for the most influential misinformation. Future studies could examine how representative fact-checking based misinformation databases are for misinformation in general.

The research community would also benefit from future studies directly comparing the emotional valence of misinformation to the emotional valence of information, in order to account for the inherent valence of specific topics. Besides addressing the limitations of this study, future work could also analyze emotions related to different types of misinformation in more detail by looking at more specific emotions such as anger, anxiety, or joy as studies have already done regarding communication about COVID-19 in general. Furthermore, different sentiment analysis methods could be tested on Twitter data also in order to replicate our findings, such as those incorporating machine learning (Al-Natour and Turetken, 2020). The use of different sentiment analysis methods could, for example, help to assess, whether the method affects the findings regarding the in general neutral tone of conversation about misinformation. As supplementation of our quantitative approach, qualitative analysis as it is applied in the field of argumentation theory, rhetoric or persuasion research, could shed light on mechanisms and reasoning behind the emotional expressions related to COVID-19 misinformation.

Overall, our study provides evidence for differences in emotional valence of the different types of COVID-19-related misinformation, highlighting the importance of investigating sentiment within meaningful clusters of misinformation. Our results, together with insights from previous research showing that negative misinformation spreads faster (Vosoughi et al., 2018), suggest that one strategy for combating misinformation could be focusing on misinformation around virus characteristics and conspiracies. This includes a focus of policies, information campaigns, or projects on these topics. Furthermore, in-depth analyses on the emotional characteristics of misinformation, and how these affect society, could further help to optimize strategies aiming to counteract the spread of misinformation.

Acknowledgments

The authors would like to thank developer Peter B. Vahlstrup for retrieving data through the Google fact checker API, student assistants Marie Damsgaard Mortensen and Petra de Place Bak for helping out with the manual coding. Furthermore, we thank the editors, and anonymous reviewers for their insightful suggestions.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the EU CEF (grant agreement number: INEA/CEF/ICT/A2020/2394203, 825469, 825652).

ORCID iDs

Jessica G Walter https://orcid.org/0000-0003-1593-3260
Ida A Nissen https://orcid.org/0000-0001-6910-1967
Anja Bechmann https://orcid.org/0000-0002-5588-5155

Supplemental material

Supplemental material for this article is available online.

Notes

1. https://www.who.int/director-general/speeches/detail/munich-security-conference.
2. The list of keywords: https://github.com/echen102/COVID-19-TweetIDs/blob/master/keywords.txt and the tweet IDs: https://github.com/echen102/COVID-19-TweetIDs.

References

Ajao O, Bhowmik D and Zargari S (2019) Sentiment Aware Fake News Detection on Online Social Networks. In: ICASSP 2019–2019 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2507–2511. https://doi.org/10.1109/ICASSP.2019.8683170

Allcott H, Gentzkow M and Yu C (2019) Trends in the diffusion of misinformation on social media. *Research & Politics* 6(2): 1–8. 2053168019848554.

Allington D, Duffy B, Wessely S, et al. (2020) Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine* 51(10): 1763–1769. https://doi.org/10.1017/S003329172000224X.

Al-Natour S and Turetken O (2020) A comparative assessment of sentiment analysis and star ratings for consumer reviews. *International Journal of Information Management* 54: 102132.

Apuke OD and Omar B (2020) Fake news and COVID-19: Modelling the predictors of fake news sharing among social media users. *Telematics and Informatics* 56: 101475. https://doi.org/10.1016/j.tele.2020.101475.

Bechmann A and O'Loughlin B (2020) *Democracy & Disinformation: A Turn in the Debate*. Thinkers' Programme, 1–30. Brussels: Royal Flemish Academy of Belgium for Sciences and the Arts.

Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1): 289–300.

Brennen JS, Simon FM, Howard PN, et al. (2020) *Types, Sources, and Claims of COVID-19 Misinformation*. Oxford; Reuters Institute for the Study of Journalism, 13. https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation

Bruns A, Bechmann A, Burgess J, et al. (2018) Facebook shuts the gate after the horse has bolted, and hurts real research in the process. *Internet Policy Review*. https://policyreview.info/articles/news/facebook-shuts-gate-after-horse-has-bolted-and-hurts-real-research-process/786.

Buning MdC, Allan R, Bargaoanu A, et al. (2018) *A Multidimensional Approach to Disinformation: Report of the independent high level group on fake news and online disinformation* (pp. 1–40). EU Commission. https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation.

Carmi E, Yates SJ, Lockley E, et al. (2020) Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review* 9: 2.

Chen E, Lerman K and Ferrara E (2020) Tracking social Media discourse about the COVID-19 pandemic: Development of a public coronavirus twitter data Set. *JMIR Public Health and Surveillance* 6(2): e19273.

EU DisinfoLab (2020) *The Few Faces of Disinformation* (p. 7). https://www.disinfo.eu/wp-content/uploads/2020/05/20200512_The-Few-Faces-of-Disinformation.pdf.

Farkas J and Schou J (2019) *Post-Truth, Fake News and Democracy: Mapping the Politics of Falsehood*. New York (NY): Routledge.

Featherstone JD and Zhang J (2020) Feeling angry: The effects of vaccine misinformation and refutational messages on negative emotions and vaccination attitude. *Journal of Health Communication* 25(9): 1–11. https://doi.org/10.1080/10810730.2020.1838671.

Ferrara E and Yang Z (2015) Measuring emotional contagion in social Media. *PLoS One* 10(11): e0142390.

Freelon D (2018) Computational research in the post-API Age. *Political Communication* 35(4): 665–668.

Ghanem B, Rosso P and Rangel F (2020) An emotional analysis of false information in social Media and news articles. *ACM Transactions on Internet Technology* 20(2): 1–18.

Graves L (2018) Boundaries Not drawn: Mapping the institutional roots of the global fact-checking movement. *Journalism Studies* 19(5): 613–631.

Gruzd A and Mai P (2020) Going viral: How a single tweet spawned a COVID-19 conspiracy theory on twitter. *Big Data & Society* 7(2): 2053951720938405.

Han J, Cha M and Lee W (2020) Anger contributes to the spread of COVID-19 misinformation 16: 1–14.. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-39.

Hutto CJ and Gilbert E (2014) VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In: Proceedings of the 8th AAAI conference on weblogs and social media, pp.216–225. http://eegilbert.org/papers/icwsm14.vader.hutto.pdf

Kalsnes B (2018) Deciding what's True: The rise of political fact-checking in American journalism. *Digital Journalism* 6(5): 670–672.

Kouzy R, Abi Jaoude J, Kraitem A, et al. (2020) Coronavirus goes viral: Quantifying the COVID-19 misinformation epidemic on twitter. *Cureus* 12(3): e7255; 1–9. https://doi.org/10.7759/cureus.7255.

Krippendorff K (2018) *Content Analysis: An Introduction to Its Methodology*. Los Angeles: Sage. https://uk.sagepub.com/en-gb/eur/content-analysis/book258450

Kumar A, Khan SU and Kalra A (2020) COVID-19 pandemic: A sentiment analysis. *European Heart Journal* 41(39): 3782–3783.

Kušen E and Strembeck M (2018) Politics, sentiments, and misinformation: An analysis of the twitter discussion on the 2016 Austrian presidential elections. *Online Social Networks and Media* 5: 37–50.

Lang PJ (1995) The emotion probe: Studies of motivation and attention. *American Psychologist* 50(5): 372–385.

Li X, Zhou M, Wu J, et al. (2020) Analyzing COVID-19 on Online Social Media: Trends, Sentiments and Emotions. *ArXiv:2005.14464 [Cs]*. http://arxiv.org/abs/2005.14464.

Martel C, Pennycook G and Rand DG (2020) Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications* 5(1): 47.

Medford RJ, Saleh SN, Sumarsono A, et al. (2020) An 'infodemic': Leveraging high-volume twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. *Open Forum Infectious Diseases* 7(7): ofaa258.

Møller LA and Bechmann A (2019) *Research Data exchange (and transparency) solution with platforms*. D.2.2 SOMA EU report.

Osmundsen M, Bor A, Vahlstrup PB, et al. (2021) Partisan polarization is the primary psychological motivation behind 'fake news' sharing on twitter. *American Political Science Review* 115(3): 999–1015. https://doi.org/10.31234/osf.io/v45bk.

Pawar AB, Jawale MA and Kyatanavar DN (2016) Fundamentals of sentiment analysis: Concepts and methodology. In: Pedrycz W and Chen S-M (eds) *Sentiment Analysis and Ontology Engineering. Studies in Computational Intelligence*. Cham: Springer International Publishing, 25–48. DOI: 10.1007/978-3-319-30319-2_2.

Pozzi FA, Fersini E, Messina E, et al. (2017) Challenges of sentiment analysis in social networks. In: *Sentiment Analysis in Social Networks*. Cambridge, MA: Elsevier, 1–11. DOI: 10.1016/B978-0-12-804412-4.00001-2.

Samory M and Mitra T (2018) Conspiracies Online: User Discussions in a Conspiracy Community Following Dramatic Events. In: *Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)*, pp. 340–349. https://par.nsf.gov/servlets/purl/10082962

Singh L, Bansal S, Bode L, et al. (2020) A first look at COVID-19 information and misinformation sharing on Twitter. *ArXiv:2003.13907 [Cs]*. http://arxiv.org/abs/2003.13907.

Soleymani M, Garcia D, Jou B, et al. (2017) A survey of multimodal sentiment analysis. *Image and Vision Computing* 65: 3–14.

Sunstein CR and Vermeule A (2009) Conspiracy theories: Causes and cures. *Journal of Political Philosophy* 17(2): 202–227.

Tandoc EC, Lim ZW and Ling R (2018) Defining 'fake news'. *Digital Journalism* 6(2): 137–153.

Vosoughi S, Roy D and Aral S (2018) The spread of true and false news online. *Science* 359(6380): 1146–1151.

Walter J, Sørensen MH and Bechmann A (2020) *Detecting Mis- and Disinformation in Digital Media*, 1–48. Aarhus (DK): SOMA H2020-ICT-2018-825469.

Wardle C and Derakhshan H (2017) *INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making*. Council of Europe. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

Weeks BE and Garrett RK (2019) Emotional characteristics of social Media and political misperceptions. In: Weeks BE and Garrett RK (eds) *Journalism and Truth in an Age of Social Media*. New York (NY): Oxford University Press, 236–250. https://doi.org/10.1093/oso/9780190900250.003.0016

Xue J, Chen J, Chen C, et al. (2020) Public discourse and sentiment during the COVID 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter. *PLoS One* 15(9): e0239441.

Zollo F, Novak PK, Del Vicario M, et al. (2015) Emotional dynamics in the Age of misinformation. *PLoS One* 10(9): e0138740.