Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Online Social Networks and Media 26 (2021) 100164

Contents lists available at ScienceDirect

Online Social Networks and Media

journal homepage: www.journals.elsevier.com/online-social-networks-and-media

The Networked Context of COVID-19 Misinformation: Informational Homogeneity on YouTube at the Beginning of the Pandemic

Daniel Röchert [1,*], Gautam Kishore Shahi [1], German Neubaum [1], Björn Ross [2], Stefan Stieglitz [1]

[1] University of Duisburg-Essen, Duisburg, Germany
[2] The University of Edinburgh, Edinburgh, United Kingdom

ARTICLE INFO

Keywords:
COVID-19
Misinformation
Network Analysis
Deep Learning, Social Media
YouTube
Homogeneity
Infodemic

ABSTRACT

During the coronavirus disease 2019 (COVID-19) pandemic, the video-sharing platform YouTube has been serving as an essential instrument to widely distribute news related to the global public health crisis and to allow users to discuss the news with each other in the comment sections. Along with these enhanced opportunities of technology-based communication, there is an overabundance of information and, in many cases, misinformation about current events. In times of a pandemic, the spread of misinformation can have direct detrimental effects, potentially influencing citizens' behavioral decisions (e.g., to not socially distance) and putting collective health at risk. Misinformation could be especially harmful if it is distributed in isolated news cocoons that homogeneously provide misinformation in the absence of corrections or mere accurate information. The present study analyzes data gathered at the beginning of the pandemic (January–March 2020) and focuses on the network structure of YouTube videos and their comments to understand the level of informational homogeneity associated with misinformation on COVID-19 and its evolution over time. This study combined machine learning and network analytic approaches. Results indicate that nodes (either individual users or channels) that spread misinformation were usually integrated in heterogeneous discussion networks, predominantly involving content other than misinformation. This pattern remained stable over time. Findings are discussed in light of the COVID-19 "infodemic" and the fragmentation of information networks.

1. Introduction

Social media such as Facebook, Twitter, and YouTube play a paramount role in today's society for exchanging information, especially in times of a global pandemic that forces many to stay at home [1]. This information includes latest status reports on the disease and thus helps citizens to make informed decisions about their actions in daily life. In addition to these day-to-day communications, social media platforms also provide effective channels for authorities to disseminate risk messages [2] and for members of the public to ask for help [3]. However, the new and multiple communication channels offered by social media also allow misinformation to flourish [4], which poses a potential threat to our collective health and democracy [5,6]. According to a recent poll by the Pew Research Center, 30% of U.S. adults who were primarily seeking information through social media have received "a lot" of conspiracy theory news alleging that the pandemic was deliberately planned [7].

Ever since the beginning of the pandemic, there has been a flood of myths and false reports about the virus (e.g., eating garlic prevents infection with COVID-19[1], and COVID-19 spreads via 5G mobile networks)[2]. The World Health Organization (WHO) speaks of an "infodemic" and has warned of the threat of "an overabundance of information—some accurate and some not—that makes it hard for people to find trustworthy sources and reliable guidance when they need it" [8, p. 2]. Since content on networking platforms such as YouTube is in the public domain, it is particularly important that the medical information provided and widely consumed by citizens is accurate and of high quality [9]. This can sometimes be a challenge since scientific findings related to such a complex and multi-layered issue like a global pandemic are elusive and, given the accumulation of scientific knowledge at an accelerated pace, fast-changing [10]. Thus, the dynamic

* Corresponding Author. Daniel Röchert, University of Duisburg-Essen, Department of Computer Science and Applied Cognitive Science, Junior Research Group "Digital Citizenship in Network Technologies"; Forsthausweg 2, 47057 Duisburg
E-mail address: daniel.roechert@uni-due.de (D. Röchert).

[1] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters#garlic
[2] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters#5g

https://doi.org/10.1016/j.osnem.2021.100164
Received 25 November 2020; Received in revised form 11 March 2021; Accepted 8 August 2021
Available online 30 August 2021

nature of scientific knowledge and its recurrent effects on policymakers and citizens offers a breeding ground for the formation and spread of misinformation [6].

In relation to global public health emergencies such as the outbreak of epidemics and pandemics, previous studies addressed the presence of misleading content on the outbreaks of Ebola [11], Zika [12], and H1N1 [13]. A number of published studies have recently already addressed the spread of misinformation about the COVID-19 pandemic on Twitter [14–17], Instagram [18], and YouTube [19]. Since misinformation about COVID-19 appears to be a phenomenon across different social media platforms, the risk of users being exposed to such information appears to be continuously prevalent. Clearly, the effects of misinformation can be harmful: When reading or viewing falsehoods, for instance, about the origin of COVID-19 or the ultimate effectiveness of masks, individuals may decide to not protect themselves or others, ignoring recommendations by centers for disease control and contributing to the further spread of the infectious disease [20]. The impact of misinformation in relation to public health crises can become even more amplified when it is spread in homogeneous clusters in which false information is treated as "normal" and accurate information is absent [21]. In an era in which information and communication networks are assumed to be fragmented (i.e., divided into different groups) along ideological lines [22,23], it is conceivable that social media technologies unite individuals who believe in misinformation and, therefore, interact mainly within like-minded cocoons where information that contradicts falsehoods does not receive any attention. In light of the potential clustering of information networks within widely used platforms such as YouTube, it seems crucial to not only assess the prevalence of misinformation, but also to analyze the network in which misinformation is disseminated and discussed.

Drawing on the notion of fragmented information networks in social media, the present study introduces the concept of *informational homogeneity* to refer to the extent to which misinformation (vs. non-misinformation) is directly connected to other pieces of misinformation in a network. By relying on this concept and focusing on the increasingly popular video-sharing platform YouTube as a news source, the present study is intended to: (a) provide knowledge about the presence of misinformation related to COVID-19 on YouTube, (b) estimate the extent to which pieces of misinformation are connected among each other, and (c) analyze to what extent informational homogeneity as an indicator for fragmentation varies over time. To this end, this study analyzes a dataset of 2,585,367 comments and 10,724 videos related to COVID-19 gathered on YouTube in the period between January and March 2020 (representing the beginning of the pandemic). The analysis combines methods from deep learning and social network analysis, allowing insights into how different types of information are connected with each other in communication networks.

The paper is organized as follows: In Section 2, we explain the theoretical background of misinformation on social media in the age of the coronavirus pandemic and its relation to the fragmentation of informationally homogeneous/heterogeneous groups. We present our research approach, consisting of data collection, annotation of the data, the deep learning language model BERT, error analysis, and network analysis in Section 3. Section 4 summarizes the study results, while these are discussed in Section 5. Finally, in Section 6, we conclude with a summary of findings and future work.

2. Theoretical background

2.1. . Misinformation on social media

The pandemic outbreak of the severe acute respiratory syndrome coronavirus (SARS-CoV-2) that causes the disease COVID-19 has permanently changed the lives of millions and thus, society, in various respects. As of September 6, 2021, approximately 220 million cases of COVID-19 and 4.6 million deaths had been reported, thereby posing an enormous challenge for countries and their healthcare systems in their fight against the spread of the virus [24].

Social media have an essential function in the distribution of news during crises, since they are capable of reaching a large number of people in a short time [25,26]. In particular the information communicated by health authorities on the current status of the virus and its spread in the respective countries is an important component of prevention measures. According to previous studies, communication via social media can help to inform the public with risk messages, optimize decision-making processes [2], and ensure rapid dissemination of scientific information [27].

Making sense of the news in extreme events is a collective process; however, establishing a common consensus could also have serious consequences, especially if users are only indirectly involved in the events. If they are not well informed, this could cause rumors to arise and spread [28]. With regard to events such as the COVID-19 pandemic, Mirbabaie et al. [29] found that in particular "information-rich actors" (e.g., media organizations, emergency management agencies) are influential in social networks and that they therefore play a key role in reducing mistrust. The quest to disseminate fast-changing scientific knowledge about an urgent matter such as a global pandemic is directly linked to dealing with the emergence of misinformation, falsehoods, rumors, and misleading content [10,21].

Even at the beginning of the pandemic, the WHO recognized another problem besides the spread of the virus, i.e., the massive amount of information that could not be guaranteed to come from trustworthy and reliable sources, and defined this as an "infodemic" [8]. According to a survey, 48% of adult US Americans had already been exposed to misinformation about COVID-19 by mid-March 2020 [30].

In general, the content of political misinformation on social platforms represents a potential threat both to democratic systems and to global health. With regard to its effects on democracy [31,32], studies showed that misinformation about current events spreads faster and more widely than true information [17,33], which could lead to political misperceptions (i.e., false or inaccurate beliefs about politics [34]). In fact, the identification of misinformation is a challenge since messages mutate and are duplicated in different contexts as time goes by [35]. Misinformation related to global health issues, for instance, in the form of conspiracy theories about vaccines, has serious consequences such as reducing people's vaccination intentions and increasing distrust on this issue [36]. To counteract this, evidence-based corrections employed by algorithms can serve as preventive measures [37]. However, when misinformation is deeply rooted in people's beliefs, it is difficult to counteract [38], especially if this misinformation is embedded in communities that deal exclusively with misinformation and are more self-contained [39].

Initial studies have already examined the emergence of misinformation during the COVID-19 pandemic. Fact-checking websites have analyzed the misinformation across multiple social media platforms, most notably YouTube, Twitter, Facebook, Instagram, etc., with the rise of the pandemic over time, and the misinformation also increases at the same rate across the world in multiple languages [40,41]. Further studies also report the rise of misinformation during the beginning of the pandemic and lockdown across numerous countries, followed by a sudden decrease in misinformation. After investigating misinformation on Facebook, Twitter, and YouTube regarding the current COVID-19 pandemic, Brennen et al. [42] were able to illustrate that while the greatest share of misinformation is disseminated by ordinary people in the social sphere, this share also seems to attract the least engagement. Kouzy et al. [16] analyzed a sample of tweets based on eleven COVID-19-related hashtags and three key terms ("Corona," "Coronavirus," and "COVID-19") on February 27, 2020, and found that Twitter accounts with a low number of followers or an unverified status were more likely to spread misinformation than verified accounts and those that had more followers. Recent studies also indicate that the dissemination of misinformation seems to be platform-dependent and that the

spread of misinformation is related to the respective users of those platforms [43]. They found that the highest level of interaction between comments and posts was on YouTube and Twitter, while the distribution of user activities (reaction dynamics and content consumption) was a commonality that was similar across all platforms. Another study examined a snapshot of the most-watched YouTube videos (N = 69) on COVID-19 and found that more than a quarter of these videos contained misleading information [19]. However, based on the small size of the sample and the limited time period it covered, it is difficult to generalize the prevalence of misinformation to all of the content that is available on YouTube. Initial evidence showed that videos on COVID-19 that contained misinformation were associated with a significantly higher number of comments that also featured misinformation [44]. The service YouTube has recognized the ongoing presence of misinformation and intends to remove content that does not adhere to its guidelines[3]. Nevertheless, due to its potential global health consequences, it seems urgent to investigate the prevalence of misinformation on YouTube related to a health issue such as COVID-19—not only on one specific day but based on a longer period of time. To comprehensively assess the presence of misinformation, it is important to not only analyze the videos but also the associated comments sections:

RQ1: What is the proportion of videos and comments that spread misinformation on YouTube in the context of the COVID-19 pandemic?

2.2. . The (informational) homogeneity in online networks

Since misinformation has become a pressing issue in the agenda of social media research [45,46], scholars proposed also taking into account the networked context in which misinformation is embedded [21,47]. These proposals address the notion that misinformation could have detrimental effects on individual actions and group dynamics if it spreads in homogeneous networks in which the misperception that the misinformation is accurate is reinforced and validated by many like-minded voices in the absence of any contradiction or correction. The juxtaposition of mass media content (e.g., news coverage) and interpersonal communication (e.g., exchanges in user-generated comments) in social media could lead to even accurate (health) information promoted by news coverage being misinterpreted or mistrusted by what readers/viewers read in the comments section [48]. Therefore, analyses of the informational homogeneity in online networks need to take into account both the main media content (e.g., journalistic videos) and corresponding comment threads.

In social media, users can choose their information sources and interaction partners in a self-determined way; selective and biased information gathering is possible because people share information without verifying it [49]. Drawing on the idea of homophily as "the principle that a contact between similar people occurs at a higher rate than among dissimilar people" [50], we propose the concept of *informational homogeneity*, which refers to the extent to which uniform types of information are connected to each other. In the context of misinformation, informational homogeneity would be high if actors who spread misinformation are closely connected to each other (forming an information cluster), while they are largely disconnected from non-misinformation (which could potentially contradict or correct the misinformation).

The level of homogeneity within online networks has already been addressed by a body of research focusing on ideologies or political opinions: While a series of studies showed that people are more likely to be connected to those who are ideologically alike [51–53], a more nuanced approach focusing on homogeneity at the topic level revealed that discussion networks are more heterogeneously structured than assumed by public concerns [54]. More specifically, on YouTube, dissimilar expressions of opinion in the form of user-generated comments were more likely to be connected to each other than comments that were similar in their stance towards a topic.

The level of homogeneity within networks is not only applicable to political views but also to the accuracy of information. Following this logic, it seems conceivable that pieces of inaccurate information are directly associated with further pieces of false information. There is reason to assume that this is prevalent in social media platforms. Recommendation algorithms, like those present on platforms such as YouTube, could lead users who initially followed a video recommendation with false or inaccurate information to further content that promotes misinformation, thereby catching those users in an information network (a "rabbit hole" [55]) predominantly comprising misinformation. Indeed, a study on YouTube found that users' individual search history is responsible for recommending them misinformation content [56]. Furthermore, it was found that videos about vaccinations that contained misinformation are promoted and thus lead the user to more misinformation. On Twitter, the findings of Shin et al. [57] suggest that the dynamic communication of political rumors (misinformation) spreads in virtual cocoons. More specifically, their network analysis revealed that polarized communities of users with the same political orientation have formed and selectively spread rumors about opposing candidates. Consequently, recommendation algorithms based on users' previous interests could even amplify the effects of misinformation by conveying users the impression that there is a whole legitimate network that promotes and discusses this kind of (mis)information [55].

Despite this initial evidence on the context of misinformation in social media, it remains unclear to what extent different pieces of misinformation are linked to each other in online networks: A recent analysis of YouTube content (videos and comments) featuring misinformation in the form of conspiracy theories suggested that there is a moderate level of opinion-based homogeneity among those nodes in the YouTube network that express a stance in support of the respective conspiracy theory [80]. While this evidence on conspiracy theories may suggest that misinformation is moderately connected to further misinformation in online networks, it is unclear whether this also applies to issues relying on fast-changing evidence such as the COVID-19 pandemic. It seems conceivable that the global uncertainty related to this pandemic has led to a stronger spread of misinformation, which also diffuses into networks with predominantly accurate information. Therefore, we ask:

RQ2: How high is the prevalence of informational homogeneity of misinformation in the context of the COVID-19 pandemic?

2.3. . The fragmentation of information networks over time

The idea that news or information could spread among certain groups of people but not among others has been best described by the term "fragmentation" of news media [58]. This has been associated with the risk that communication landscapes are segmented and divided into sub-groups that are homogeneous in terms of what kind of information they receive and discuss, but also disconnected from the other sub-groups, leading to an asymmetrical diffusion of news and information [59]. From a normative point of view, fragmentation of news channels on social media, on the one hand, can have a positive effect on the distribution of relevant information since more sources of information are available [60]. On the other hand, fragmentation also carries risks and dangers, especially when these fragmented groups polarize and spread extreme ideologies, misinformation, or hate speech [23].

In direct association with the concept of homophily, studies have examined to what extent a divergence of political ideologies is responsible for fewer interactions among individuals, resulting in a fragmentation of information and discussion networks [22,61]. Empirical evidence, however, showed that the actual division in communication only applies to the politically extreme—there are still cross-cutting interactions among those who have different political views [22]. Likewise, an analysis of audience segments across different media outlets

[3] https://www.youtube.com/howyoutubeworks/our-commitments/fighting-misinformation/

revealed a significant overlap of media consumers between all of these channels, refuting the idea of enclaves in communication networks [62]. With the diffusion of algorithms in people's communication practices, the idea of news audience segmentation has gained renewed relevance [63]: Indeed, a bounded confidence model revealed that algorithm bias in the flow of information can strengthen the fragmentation of information consumers and their opinion polarization [64].

In the context of misinformation, the fragmentation of subgroups marked by informational homogeneity would mean that certain segments of a network are disproportionately exposed to misinformation, while at the same time being disconnected from sources of accurate information. Such a network structure could lead those groups that are homogeneously exposed to misinformation to believe in the accuracy of that false information without encountering any contradiction or correction [21]. However, the informational homogeneity of a certain sub-network may not emerge instantly, but instead increase over time: One study that focused on network fragmentation in the context of the Syrian war over a period of 32 months showed that fragmentation and homogeneity were generally high in the network. However, the temporal evolution of these fragmented groups showed that only one group increased its ideological homophily over time [65].

While some research has investigated the fragmentation process in political issues, there is still very little scientific understanding of fragmentation in the context of misinformation. An investigation on the online consumption of fake news found fragmentation between a fake news audience (minority) and a real news audience (majority) [66]. The same study also determined that the rapid spread of misinformation has a massive impact on the media environment, making it difficult for users to determine which news is right and which is wrong.

In addition to the existence of misinformation, however, the temporal consideration of informational homogeneity is particularly relevant in order to examine whether the dissemination of misinformation leads to the formation of disconnected network segments over time. In line with suggestions made by Webster & Ksiazek [62], we argue that audience fragmentation is best addressed by a network analytic approach, assessing the links between nodes in a communication network. To assess the fragmentation of the information landscape related to the COVID-19 pandemic, we therefore rely on the concept of informational homogeneity and its manifestation over time and ask:

RQ3: Are there temporal (i.e., monthly) differences in informational homogeneity within YouTube information networks in the context of the COVID-19 pandemic?

3. Methodology

In order to assess the proportion of misinformation, we first needed to: (a) collect data, (b) annotate part of the collected data, and (c) train a model to predict all remaining data records. This data consists of information about YouTube videos related to the search term "coronavirus," along with the comments on these videos. A random sample was then annotated by determining, for each video or comment, whether it belonged to the "misinformation" or "non-misinformation" class. Finally, natural language processing (NLP) techniques were used to predict the class for the remaining data records that had not been annotated. In particular, our approach uses the deep learning technique BERT (Bidirectional Encoder Representations from Transformers) to detect misinformation based on the previously annotated comments and videos on YouTube. To ensure the quality of the classification model, we performed an error analysis to ensure error classes and validate the results.

Once this classification step had been completed, to examine the communication network of YouTube and compute its informational homogeneity, we: (a) transformed the YouTube videos and comments into a network structure and (b) computed the external–internal (E-I) index on the basis of the two classes. This combination of NLP and network analysis allowed us to identify the homogeneity of the network from the communication paths of users. To determine the fragmentation, i.e., the temporal aspect of homogeneity in our data, we examined subsequent months individually and compared them with each other.

3.1. Data collection

For data collection, we ran a self-developed program from 1 January 2020 to 11 March 2020 that accesses the YouTube application programming interface (API) and retrieves metadata about the videos and content, as well as metadata of comments and replies. YouTube plays an increasingly important role in the consumption of news because it provides a platform where multifaceted information from different news channels comes together [67]. Based on a recent Pew Research Center survey, 26% of U.S. adults indicated that they used YouTube as a news source because it is a key source for staying up to date [68]. We used a similar method to that used by Röchert et al. [54] to obtain the data using the search, video, comments, and replies list. When passing the parameters responsible for the output of the search results, we sorted the videos with the parameter "order" by "date" in order to iteratively collect, for every single day, content related to the search term "coronavirus." By repeating this iterative procedure after short periods of time, it was possible to ensure that the number of collected videos could be heard. Furthermore, we carried out another collection in which we changed the parameter "order" to "relevance" in order to also collect the most relevant videos according to YouTube. For both procedures, we set the parameter "relevantLanguage" to "en" and "de" to get a wide range of videos. We searched for the word "coronavirus," which was used internationally at that time. Based on Google Trends and a worldwide comparison of the words "coronavirus" and "COVID-19," the term coronavirus received much higher attention during the investigation period[4]. Following data collection, we noticed that despite the filtering of the language, the term "coronavirus" was still used in multiple other languages. Focusing on the English language, we used the language classification API "detectlanguage"[5] to identify English videos based on the title and description. This step is necessary because, although we had specified a "relevance language" in our requests to YouTube's API, the API documentation warns that "results in other languages will still be returned if they are highly relevant to the search query term." In total, we collected 10,724 videos and 2,585,367 comments and replies. Figure 1 shows the crawling procedure of the dataset.

3.2. Annotation

We developed a coding scheme that serves as a guideline for the manual annotation of unlabeled videos and comments. For this purpose, we defined two mutually exclusive classes (misinformation and non-misinformation), which were used for annotating videos and comments. Misinformation is inaccurate information shared by the user without a clear intention to deceive. Often, the user is involved in circulating the misinformation without knowing the background truth, here in this study without knowing the truth about the YouTube videos. In contrast, disinformation is a piece of information that is deliberately misleading or biased. The user has the intention to mislead or deceive others. People alter the truth or repurpose the original story to spread propaganda, cheat people, etc. Without knowing the origin of YouTube videos, it is difficult to classify a video as misinformation or disinformation, so for this study, we classified videos as misinformation and non-misinformation. The misinformation category might include some videos that are disinformation, while in non-misinformation, we include YouTube videos that do not contain any false information.

In this study, the "misinformation" class contains all unintentionally

[4] https://trends.google.com/trends/explore?gprop=youtube&q=covid-19,%2Fm%2F01cpyy

[5] https://detectlanguage.com/

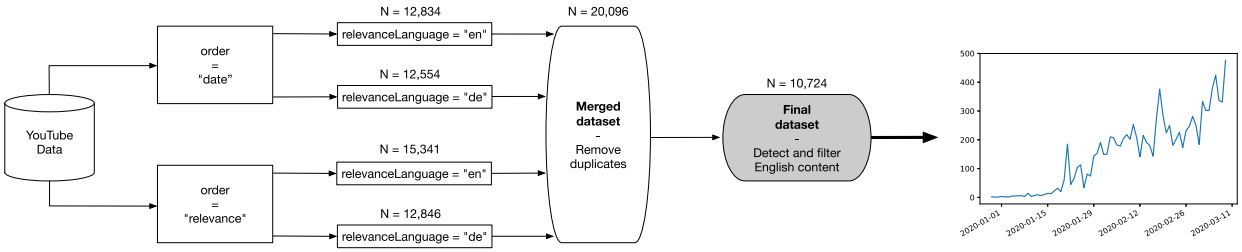

Figure 1. An illustration of the data collection process.

and intentionally false information about the origin, distribution, prevention, etc., of the COVID-19 virus and disease. This class also includes conspiracy theories and content that misleads the user with a wrong title, captions, misrepresented context, or statistics. In contrast, videos or comments that do not contain any information about the coronavirus or neutral news reporting, as well as satire or parodies, are annotated as "non-misinformation." Furthermore, this class includes videos or comments that do not contain any false information and therefore could be true or refer to a completely different topic. If the video or comment was not in English, it was also marked as non-misinformation. In line with ethical principles regarding misinformation that may lead to ostracism and profiling, we decided to consider only content-relevant information in the annotation; metadata such as the name of the channel was hidden or not considered in the video annotation.

To ensure the correct annotation, especially of the videos, the content was examined while watching the video and investigating the title and description, and the topic was additionally checked with the International Fact-Checking Network (IFCN) of the Poynter Institute. If the IFCN signatories had not fact-checked the information, then we searched for additional information from reliable sources such as government portals and reputable news websites.

Since annotating the entire dataset using this technique was not feasible, we annotated a portion that was sampled according to the number of videos and comments published in the respective months as follow: To ensure that all time periods were sufficiently represented in the sample, we used stratified sampling so that 20% of the sample consisted of data from January (when the overall number of videos about COVID-19 was still lower), 40% from February, and the remaining 40% from March. We only considered the videos that had public comments and found that some of the videos or comments had been deleted or removed from YouTube. The final sample consisted of 429 videos and 10,400 YouTube comments, which were annotated. An overview is presented in Table 1.

Each YouTube video and comment was annotated by three annotators, all undergraduate students. To measure inter-coder reliability, we used Fleiss' Kappa [65], which resulted in a value of 0.582 for the video dataset and a value of 0.473 for the comment dataset, indicating a moderate level of agreement. For the determination of the final class, we used a majority vote. If the class could not be determined, the annotators reviewed the videos and comments again in order to come to a decision.

Overall, the number of videos that contained misinformation was not sufficient to train a deep learning model. We pre-tested this in advance and found that the model overfitted due to the low training data and that too many errors occurred in the performance on the test data. This effect was not only observed with the undersampling procedure, but also with the distribution of the real dataset (unbalanced). As Zhang et al. [69] point out, a major challenge in developing a misinformation classification system is the lack of annotated data; therefore, we decided to add external data from the IFCN of the Poynter Institute, which stores known false information content about the coronavirus [41]. The database contains fact-checked articles on COVID-19 that have been identified from different signatories (fact-checking companies) from multiple countries. The IFCN provides basic information such as title, date, and country in English and points to the actual fact-checked article's webpage. A further advantage of these articles is that they cover a broad spectrum and report worldwide information regarding the COVID-19 pandemic, which is therefore ideal for the further course of our analysis. Since many of these statements are very short, they are similar to the YouTube video titles and are thus an ideal data source. Figure 2 demonstrates an example of the gathered information from Poynter.

For the collection of the data, we manually collected the headings from 14 January 2020 to 9 March 2020, which also corresponds to our investigation period and hence reflects comparable incidents related to the coronavirus. To obtain only clearly false information, we filtered the results to only include the category "false" (see Table 2).

Table 1
Overview of manually labeled YouTube videos and comments.

		Month			
Dataset	Class	January	February	March	**Total**
Videos	Misinformation	3	22	9	**34**
	Non-misinformation	61	152	182	**395**
Comments	Misinformation	119	379	298	**796**
	Non-misinformation	1283	3767	4554	**9604**

3.3. Pre-processing

As a first step in pre-processing the data, we merged the manually annotated video information with the fact-checked statements. In total, our video dataset contained 996 entries belonging to the misinformation class and 395 entries belonging to the non-misinformation class. For the comments, we used the 10,400 comments from the manual annotation. Since the class distributions were unbalanced in both datasets, we randomly undersampled the larger class so that both classes had the same size in the training process. As a result, we had 395 records for each class in the video dataset and 796 records in the comment dataset. Before training our classification model, we also performed common text pre-processing steps so that the text could be handled more efficiently by the algorithm. These processes were identical for both datasets. We removed the hyperlinks mentioned in the text and expanded contractions (e.g., "wasn't" to "was not", "we'll" to "we will"). We also removed punctuation marks from the text. Since we do not train our model on video files (video sequences), but only on the textual metadata given for the video, we decided to merge the title as well as the description of the YouTube videos to capture more meaning in the text and not lose essential information. Therefore, we concatenated the title and description together, while for the comments classification, we used only the textual information of the comments and replies from the videos.

3.4. Classification model

For the classification of misinformation in the comments and videos, we used the state-of-the-art neural network language model BERT, which has been pre-trained on a large corpus in order to solve language processing tasks [5]. An essential advantage of BERT is that it can be fine-tuned for task-specific datasets and allows high text classification accuracy even for smaller datasets. In the context of COVID-19, BERT

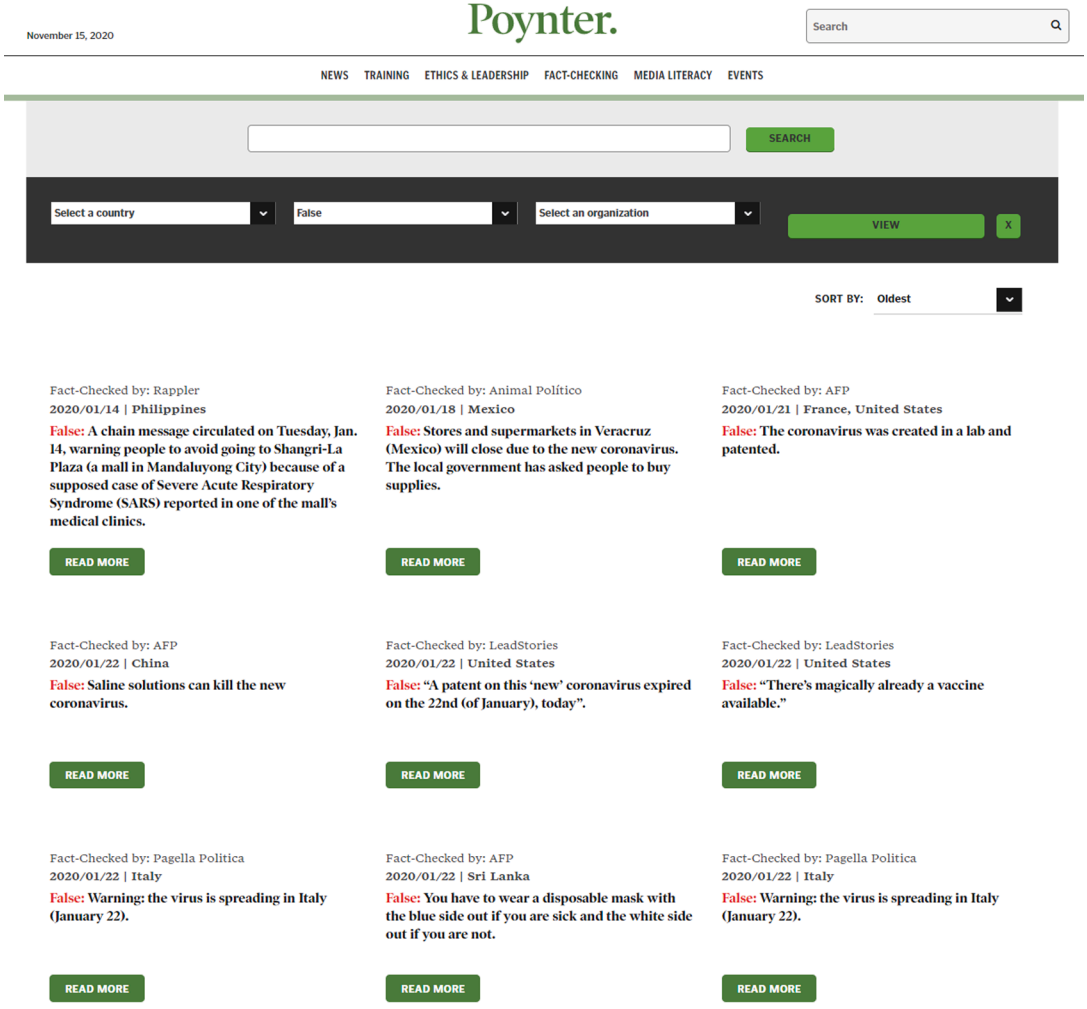

Figure 2. Screenshot of the Poynter database for COVID-19 in the category "false".

Table 2
Overview of additional fact-checked videos.

		Month			
Dataset	Class	January	February	March	**Total**
Fact-checked data	Misinformation	213	507	242	**962**

has already been applied for multiclass classification tasks, for example, on the Chinese social media platform Weibo, where it achieved considerable accuracy [70]. Furthermore, BERT was also used for other problems such as the detection of misinformation [71,72] or the identification of hate speech [73,74]. When using BERT, it should be noted that the texts must be formatted in a specific way in order to ensure that the training is carried out correctly. This pre-processing includes converting text to lowercase, tokenizing it, breaking words into word pieces, as well as attaching "CLS" and "SEP" tokens to represent the meaning of the entire sentence and to separate sentences for the next sentence prediction task. We split the video and comment data into 80% training data (videos: 632; comments: 1,273) and 20% test data (videos: 158; comments: 319). The randomization of the data prevents seasonal patterns from being learned by the model. For BERT fine-tuning, the model for the videos was trained for four epochs and the model for the comments was trained for three epochs with a learning rate of 2e-5. For the video dataset, we used a batch size of 8 with a sequence length of 128 because the dataset contains fewer records, and the titles of the fact-checking websites are also generally shorter. For the comments dataset, we chose a batch size of 32 with a sequence length of 128, since the average sequence length was 153 and the median sequence length was 88. After the individual prediction on the two test datasets, we evaluated the accuracy of the two models using the weighted F1 score.

3.4.1. *Baseline model*

We use two classical machine learning algorithms (Support Vector Machine [SVM] and Logistic Regression [LR]) and two deep learning techniques (Long Short-Term Memory [LSTM] and Convolutional Neural Network [CNN]) to compare the performance of BERT against those baseline models. For SVM and LR, we trained a term frequency–inverse document frequency (TF-IDF)-weighted character n-gram model with optimally selected hyperparameters based on grid search with five-fold cross-validation. The applied hyperparameters can be found in Appendix A.

In the deep learning techniques, we decided to keep the architecture the same for the comments and videos. For this reason, we will describe them globally, with individual parameters given in Appendix A.

In the LSTM model, our first layer is an embedded layer with an input length of 128. After this layer, an LSTM layer with 128 memory units is added. Following this layer, we set a dense layer with a unit of 128. The output layer is defined by one neuron with a sigmoid activation function. As an optimization function, we choose Adam [75] with the binary cross entropy loss function, suited for binary classification problems.

The CNN model is characterized by the first layer as an embedded layer with an input length of 128. After this layer, a Conv1D layer of 128 filters and a size of 3 with a ReLu activation function and max pooling of 3 is added. Following this layer, we set a flatten layer to reduce the dimension in our model and add a dense layer with a unit of 128. The output layer is the same as that described for the LSTM model with a single neuron and a sigmoid activation function.

After we had compared all the models, the results showed that BERT had the best performance in the video and in the comment dataset. The comparison of the different models and their performance can be seen in Table 3.

As can be seen in Table 3 above, the best F1 score for the commentary classifier was 0.81, and the score for the video classifier was 0.97. A detailed demonstration of the prediction within each class of the chosen BERT models can be found in Table 4. Since the values of the F1 score were acceptable for our further analysis, we proceeded to use the models to classify the entire dataset of videos and comments.

3.4.2. Error analysis

We performed an error analysis to evaluate the performance of the video and comment classification models. Therefore, we created an independent validation set that does not contain training and test sets and consists of 50 data records for each month of comment and video datasets. In total, we had 150 videos and 150 comments that we analyzed. Based on these sample datasets, we performed a manual analysis and checked the predicted content for their accuracy. In this manual analysis, the predicted values of the comments and videos were compared to the human annotation in which the comments were read and the videos were watched. The aim of the manual analysis is to identify specific classes of errors that may be responsible for the incorrect prediction and that have occurred most frequently. Since our models are binary classifiers, we can specifically address false negative and false positive errors.

Comments

Overall, we identified an 8% error rate of our 150 comments where these were predicted only as false positives. In diagnosing the predicted comments and their classes, we identified four reasons (off-topic, sarcasm/joke, lack of special knowledge, and lack of video context) that were responsible for the misclassification.

Off-topic: In this identified error class, which occurred most frequently, we could see that YouTube comments did not focus on the topic under investigation, "coronavirus," but rather dealt with different topics, which were kept very general.

Sarcasm/joke: This error class has already been found in other studies on hate speech and refers to comments that contain sarcastic or funny content. In particular, the topic of coronavirus was addressed here in conjunction with the eating habits and food that might have caused the disease (bats) and the treatment of the virus (handwashing).

Lack of special knowledge: We identified this error class because some misclassifications were related to healthcare information such as contagion, wearing masks, or information about the virus. This also includes information about specific locations that were not frequently included in the dataset.

Lack of video context: In this class of errors, we found errors that were directly related to the content of the YouTube video. For example, these comments contained spelling errors or declared the related video to be fake news.

Videos

As with the comments, we also manually checked a sample of the video dataset for errors. In general, we found an error rate of 7.33%, with false negative and false positive errors. Videos that were no longer available on the YouTube platform (N=20) were still coded based on the title and description to ensure comparability. In addition to the identified classes of errors, we noticed in particular that the descriptions had a major influence on the classification of the videos. While many official news channels add a description when publishing the videos, there are also some channels that do not have descriptions. Videos that do not have descriptions are more likely to be declared as misinformation by the algorithm. Overall, we were able to determine the following one class of error in the comments that were "false negative." The "conspiracy content" category was the most frequent with eight errors. In this category, as many as four videos had been deleted and were no longer available on YouTube due to violations of YouTube guidelines.

Conspiracy content: We defined this error class because it was most prevalent with conspiracy theory content about COVID. Here, the titles in particular consisted of rhetorical questions and were related to the outbreak of the virus. Furthermore, the length of the titles and the description of the videos were given with few characters.

For false positive errors, we were also able to identify one error class, in which the frequency of errors in the category "news channel content" occurred four times.

News channel content: The errors that were identified in this class were characterized by a short title in combination with a short description. More precisely, news channels used questions in the title (including rhetorical questions) and created a direct link to a specific scenario (e.g., disinfectant). Here, the description of the video may also be completely omitted.

3.5. Network analysis

After classifying the entire dataset of comments and videos using the trained models, we generated two different directed communication

Table 3
Model evaluation of deep learning and machine learning methods on the test dataset.

			Weighted average			Macro average		
Dataset	**Models**	**Epoch**	**Precision**	**Recall**	**F1 score**	**Precision**	**Recall**	**F1 score**
Comments		1	0.78	0.77	0.77	0.78	0.78	0.77
	BERT	2	0.80	0.80	0.80	0.80	0.80	0.80
		3	**0.81**	**0.81**	**0.81**	**0.81**	**0.81**	**0.81**
		4	0.81	0.81	0.81	0.81	0.81	0.81
	LSTM	10	0.71	0.70	0.70	0.71	0.71	0.70
	CNN	10	0.71	0.70	0.70	0.70	0.70	0.70
	LR	-	0.76	0.74	0.74	0.75	0.75	0.74
	SVM	-	0.75	0.74	0.74	0.74	0.74	0.74
Videos		1	0.97	0.97	0.97	0.97	0.97	0.97
	BERT	2	0.97	0.97	0.97	0.97	0.97	0.97
		3	0.97	0.97	0.97	0.97	0.97	0.97
		4	**0.97**	**0.97**	**0.97**	**0.97**	**0.97**	**0.97**
	LSTM	10	0.92	0.91	0.91	0.91	0.92	0.91
	CNN	10	0.93	0.93	0.93	0.93	0.93	0.93
	LR	-	0.90	0.90	0.90	0.90	0.90	0.90
	SVM	-	0.91	0.91	0.91	0.90	0.91	0.90

Table 4
Summary of the precision, recall, and F1 for each class based on the final BERT models.

Dataset	Class	Metrics Precision	Recall	F1 score	Support	Prediction
Comments	Misinformation	0.79	0.81	0.80	149	154
	Non-misinformation	0.83	0.81	0.82	170	165
	Weighted avg.	0.81	0.81	**0.81**	319	319
Videos	Misinformation	0.97	0.96	0.97	72	71
	Non-misinformation	0.97	0.98	0.97	86	87
	Weighted avg.	0.97	0.97	**0.97**	158	158

networks (1. video comment, 2. comment network) from the entire predicted YouTube data. The distinction between the two networks is intended to clarify the analysis in terms of network homogeneity between video and comment misinformation.

The first network reflects the entire YouTube network with links to videos and comments. Here, the nodes represent uploaded videos and users who have written at least one comment. Interactions are represented by the directed edges. Nodes A and B are linked by a directed edge from A to B if: (a) user A has commented on video B or (b) user A has replied to a comment made by user B. The second network, on the other hand, was generated only from comments and their replies, in order to determine the communication within the comments. Videos that were represented previously as hubs were removed in this network.

Based on the output of the classification results for the videos and comments, we computed for each node whether the particular user has spread misinformation or not. In the case that users had written numerous comments, we computed the aggregated value of the classification outputs for each comment by applying the arithmetic mean (compare [54,80]). In addition, we also eliminated self-loops (comments regarding one's own video and replies to one's own comments) and disconnected nodes (videos without comments) because they have no further impact on the final outcome. To measure the informational homogeneity, we used the global E-I Index of Krackhardt & Stern [76]. The E-I index is defined as follows:

$$EI = \frac{E - I}{E + I}$$

where E represents the number of external ties and I the number of internal ties.

Furthermore, we computed the directed per-group E-I indices, considering the direction of the edges by counting only outgoing links as external links. In this context, the main purpose of the computed per-group E-I index is to focus on the interaction of the members of a specific group, i.e., which users they have communicated with. Compared to the undirected groupwise E-I index, this gives a much more accurate representation of users' interactions.

We performed a permutation test to determine whether the given E-I index is significantly smaller or larger than the expected E-I index when the connections in the network are randomly generated. This involves creating multiple iterations of graphs based on the sampling distribution, where each edge is randomly rewired. In this way, we can test the null hypothesis that the edges are randomly distributed among the nodes and ensure that the number of nodes in each group and the ties is constant.

Table 5 below provides an overview of the evaluated networks based on their network properties.

For a summary of our methodological approach, see Figure 3. First, the videos and their comments were collected using the YouTube API, and then a subset was manually annotated. We then trained the classifier on most of the annotated videos and comments using two independent BERT models and evaluated them using the remaining annotated data as test datasets. We then used the two trained models to classify the entire dataset and transformed the data into a network structure. Using this network, we were able to compute the informational homogeneity and determine how the discussion of misinformation developed over a period of three months.

4. Results

Regarding RQ1, we found that 26.37% (N = 681,811) of comments were classified as containing misinformation, while the proportion of non-misinformation content was 73.63% (N = 1,903,556). Of the videos, 3.5% (N = 376) contained misinformation and 96.5% (N = 10.348) non-misinformation. After aggregating the classifications across all of the content posted by each user, we found that in January, 16% of users primarily posted misinformation (compared with 84% who did not). In February, this number rose to 20%, and in March it dropped again to 16.4%. The proportion of misinformation from the interaction of videos and comments, which we could observe on the basis of our network perspective (after pre-processing), was 21.8% in January, 19.9% in February, and 16.3% in March. In order to validate the error of the classification and thus ensure the quality of the results, we decided to perform an error analysis. Based on this error analysis, we were able to identify four different error classes of the comments and two error classes of the videos, making a correct prediction of the comments difficult. The errors of the comments refer to thematic points of view, with a lack of additional information such as the content of the video watched or specific medical knowledge. Based on the content of the

Table 5
Network properties.

Network parameter	Video and comments network			Comments networks		
	January	February	March	January	February	March
Nodes	222,204	460,816	308,367	131,991	244,017	166,841
Edges	394,008	1,102,352	603,704	203,790	484,651	280,503
Avg. degree	1.77	2.39	1.96	1.54	1.99	1.68
Diameter	31	24	27	31	24	27
Max. out-degree	159	411	252	152	400	241
Max. in-degree	1,712	1,625	1,663	1,712	884	304
Density	0.000008	0.000005	0.000006	0.000012	0.000008	0.00001
Assortativity	0.004	0.005	0.019	0.062	0.060	0.067
Clustering coefficient	0.001	0.002	0.002	0.001	0.003	0.003

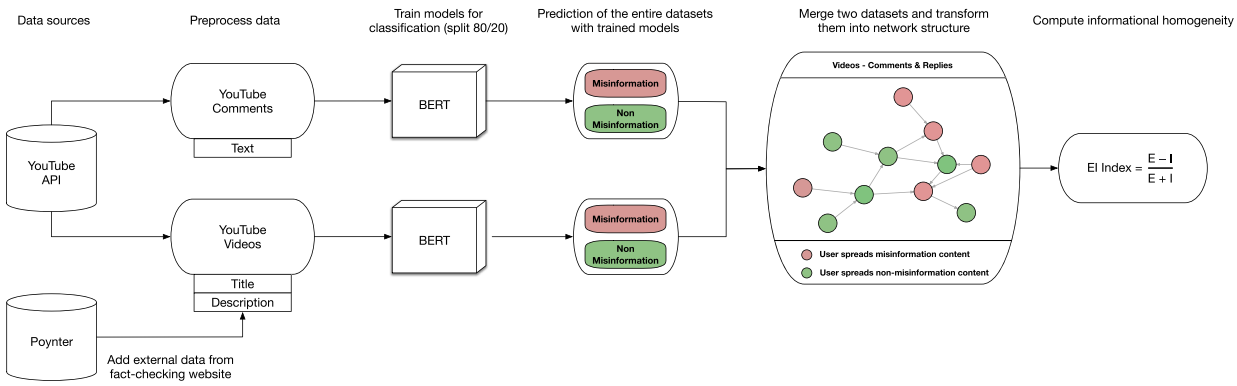

Figure 3. Description of the classification process and informational homogeneity analysis.

comments, we also found that comments that did not address the topic of coronavirus were misclassified and, thus, had a unusual number of words in the learned corpus, as well as containing sarcastic/funny content. On the other hand, when diagnosing the errors of the videos, we found that videos that have already been deleted, lack information, or contain conspiracy beliefs are also incorrectly predicted. Nevertheless, we have to mention that the percentage of errors in the analysis of errors is very low.

To address RQ2, the extent of informational homogeneity in YouTube networks among user-generated comments and videos on COVID-19, the results show that there is a significant difference in the class E-I indices of misinformation and non-misinformation. In our analyses of the two networks (video-comment network, comment network) over three months, the results indicate that people who disseminated misinformation find themselves in a heterogeneous discussion environment. Table 6 demonstrates the results of the homogeneity analysis of the three different months based on the video and comment network. While the per-class E-I indices for non-misinformation are all negative (January: –0.795, February: –0.850, March: –0.869) and thus show a homogeneous communication pattern, they are all positive for the class of misinformation (January: 0.788, February: 0.842, March: 0.839), which indicates heterogeneous communication. Similar results in Table 7 were also found in the communication-only network, where we considered only the links of comments and removed the links to the video. Compared to the whole network, the per-class E-I index of the commentary network has slightly lower values over the three months for all classes and therefore also a lower global E-I index. This results from the fact that videos, which are seen as a central hub in the network, are dropped and thus no longer have significant influence. Here, the per-class E-I indices for non-misinformation also show a homogeneity trend (January: –0.649, February: –0.698, March: –0.760), whereas the misinformation class indicates a more heterogeneous communication pattern (January: 0.508, February: 0.558, March: 0.605).

Taking into account the permutation test, the results in Table 6 and 7 indicate that the expected E-I index is negative for the "non-misinformation" class and positive for the "misinformation" class. With respect to the results on the null hypothesis test in Table 6, one can see that in all months the observed E-I index of the "non-misinformation" class is significantly closer to –1 than the expected E-I index, while the observed E-I index of the "misinformation" class is significantly closer to +1 than the expected E-I index. Concerning the null hypothesis test in the comment-only network, Table 7 shows that in all months the observed E-I index of the "non-misinformation" class is significantly closer to –1 than the expected E-I index. For the "misinformation" class, in contrast, one can see that in all months the expected E-I index of the "misinformation" class is significantly closer to +1 than the observed E-I index.

Addressing the RQ3, there is a trend in both networks (videos/comments, comments only) for communication to become more informationally homogeneous over time. A consideration of the global E-I index for both networks indicates a clear trend towards a more homogeneous information network over time. With respect to the individual classes, however, there are minor differences. While in the video comment network the values for the misinformation class become more heterogeneous from January to February, the E-I index stagnates at a similar value of 0.839 in March. In the pure commentary network, it can be seen that for the misinformation class the communication within the comments becomes continuously more heterogeneous from January to March. For the non-misinformation class, the findings show that the communication between the videos and the comments or only within the comments becomes more homogeneous from January to March.

5. Discussion

In the COVID-19 pandemic, the world has not only seen a virus spread all over the world—an overabundance of information, including misinformation and conspiracy theories, has also been disseminated through online social networks [77]. The fight against misinformation on social media platforms poses many challenges: One step towards addressing those challenges is to understand whether the diffusion of misinformation divides users into segments in online networks, leading some users—in the long run—to be caught in clusters that are predominantly filled with misinformation and disconnected from the clusters that provide corrections or contradictions. To examine this question, we used a combination of deep learning and network analysis methods to compute the informational homogeneity among videos and comments on COVID-19 on the video-sharing platform YouTube.

Results showed that, over the period from January to mid-March, approximately 3.5% of videos and 26.37% of comments contained misinformation. These findings of misleading videos are lower than the proportion found by Li et al. [19], who revealed that about 23%–26% of YouTube videos were misleading, generating attention from millions of viewers worldwide. A possible explanation for this might be that Li et al. analyzed data crawled on one day at the end of March 2020. This was undoubtedly a "hot" stage in which information needs might have been remarkably higher, but also the potential publication of misinformation in the form of videos may have likewise been higher. In our view, our results do not challenge the findings presented by Li et al., but indicate that the amount of misinformation may vary depending on the stage of a crisis. When comparing our results with those of Li et al., stages might be more ephemeral in the sense that the amount of information might not increase month by month (as we show in our results), but significantly from day to day. Future analyses need to investigate the emergence of misinformation in much smaller units to do justice to the information needs created in the face of (health) crises. Considering our results, we can see that the spread of videos containing misinformation is low and that some videos have already been deleted, but the number of comments containing misinformation and thus having an influence on users' information processing is relatively high at 26.37%. It seems even more

essential for social media service providers to take action against misinformation comments given that the spread of such comments could most certainly have severe consequences on individual and collective health [6,20].

Using error analysis with the validation dataset, we were able to examine the quality of the classification model and identify specific sources of error related to comments and videos. In the case of comments, it was noticeable that they were more often incorrectly predicted if, for example, they were not related to the context of COVID, and thus were off-topic, or if they required specific medical knowledge to correctly identify the context. In addition to these findings, however, there are also parallels with other research that has looked at text classification of hate speech on online social media, which also found sources of error from texts such as sarcasm [78,79]. Text classification seems to work better using state-of-the art techniques such as BERT, but errors still occur when there is ambiguity or too little context. Reviewing the videos, it was apparent that many videos had already ceased to exist, potentially having been deleted due to the current YouTube guidelines, as YouTube is increasingly taking action against misinformation.

Misinformation in the domain of public health can pose a significant risk if people believe in the accuracy of this information and act accordingly. The mistaken belief in the accuracy of misinformation could be reinforced if that misinformation is embedded in a network in which misinformation is predominantly present without any correction or contradiction [21]. To analyze the networks in which misinformation is spread, we transformed our YouTube dataset into a network and computed the extent to which this discussion network may contain homogeneous clusters. Our results indicate that the communication paths of users who disseminate misinformation in the network are quite heterogeneous, since they are predominantly connected with nodes that disseminate non-misinformation. The E-I index indicated a relatively high level of informational heterogeneity associated with misinformation and this pattern slightly increased over time, suggesting that the spread of misinformation does not lead to an increase in *misinformational* homogeneity in networks in the long run. This result would speak against the notion of network fragmentation consisting of enclaves with certain types of information that are not available to others [59,62].

In this context, it seems worthwhile to compare the level of informational homogeneity between networks containing videos and comments versus networks containing only comments (see Tables 6 and 7): In fact, results showed that the misinformation was connected to non-misinformation to a larger extent when networks included both types of content, i.e., videos and user-generated comments. Therefore, it seems that the blending of mass and interpersonal communication that characterizes many social media platforms [48] is responsible for higher levels of informational heterogeneity. While this appears to be a desirable result, it also raises questions: Given that the prevalence of COVID-19-related misinformation was higher in user-generated comments than in videos, future (experimental) research needs to test under

Table 6
Results of the permutation test with the observed and expected class E-I index (videos and comments network).

Month	Sentiment	Observed E-I index	Expected E-I index	P (obs ≥exp)	P (obs ≤ exp)
January	Global	–0.508	–0.318	<0.01*	1.00
	Misinformation	0.788	0.564	1.00	<0.01*
	Non-misinformation	–0.795	–0.564	<0.01*	1.00
February	Global	–0.587	–0.363	<0.01*	1.00
	Misinformation	0.842	0.603	1.00	<0.01*
	Non-misinformation	–0.850	–0.603	<0.01*	1.00
March	Global	–0.653	–0.455	<0.01*	1.00
	Misinformation	0.839	0.674	1.00	<0.01*
	Non-misinformation	–0.869	–0.674	<0.01*	1.00

Table 7
Results of the permutation test with the observed and expected class E-I index (comment-only network).

Month	Sentiment	Observed E-I index	Expected E-I index	P (obs ≥ exp)	P (obs ≤ exp)
January	Global	–0.488	–0.383	<0.01*	1.00
	Misinformation	0.508	0.619	<0.01*	1.00
	Non-misinformation	–0.649	–0.619	<0.01*	1.00
February	Global	–0.553	–0.405	<0.01*	1.00
	Misinformation	0.558	0.637	<0.01*	1.00
	Non-misinformation	–0.698	–0.637	<0.01*	1.00
March	Global	–0.632	–0.491	<0.01*	1.00
	Misinformation	0.605	0.701	<0.01*	1.00
	Non-misinformation	–0.760	–0.701	<0.01*	1.00

which circumstances user-generated comments challenging or contradicting health-related information featured in journalistic videos or articles can exert an impact on their viewers'/readers' ultimate health-related knowledge and attitudes (e.g., on the acceptance of a COVID-19 vaccine).

There are two possible interpretations of our results, one optimistic, one pessimistic, yet both equally valid. The fact that misinformation is not concentrated in closed networks consisting of nodes that are predominantly associated with false information may prevent the formation of cohesive groups in which individuals mutually reinforce misperceptions and attitudes [64]. At the same time, it seems that misinformation successfully diffused in mainstream networks that were otherwise filled with non-misinformation. While this certainly does not lead to a segregation of certain information consumers, it may make the detection of misinformation more difficult for users who encounter false information in juxtaposition with accurate information [37]. At this point, it remains unclear whether misinformation is spread deliberately in those networks.

6. Limitations

As with most research, this research also has a number of limitations. First, we would like to emphasize that our results are based only on an English language dataset and on one specific search keyword, "coronavirus." Thus, we cannot state whether the results are transferable to other languages. Due to this random factor of sampling, we were faced with the challenge that there were too few datasets in the video dataset for the training of the BERT model, and we overcame this by increasing the amount of under-represented data by using fact-checking. In addition, time passed during the data collection and annotation process, which led to some videos being removed from YouTube due to violations of the guidelines and, thus, also excluded from our data analysis. Another limitation is the fact that we analyzed content published at the beginning of the pandemic; more precisely, we analyzed the videos and comments from 1 January 2020 to 11 March 2020. For this reason, it should be pointed out that after this period of time, further videos as well as comments may have been produced, thereby potentially providing more misinformation. For this reason, we cannot make any statements about the further course of the pandemic. A more comprehensive analysis could include later months of the pandemic and cover the full information landscape related to COVID-19 on YouTube. Moreover, it is worthy of note that our conclusions are based on predictions by a deep learning model (BERT), which has shown good results in previous research in different areas. The results should nevertheless be considered with some circumspection, since our results show that despite the high F1 score, there are still a few incorrect classifications in the test dataset. The final limitation is that YouTube Data API developer policy does not allow publication or distribution of the data used in this study, which does not ensure reproduction of the same results.

7. Conclusion and future work

This study investigated the informational homogeneity of misinformation on YouTube in the context of the current COVID-19 pandemic. We annotated random comments and videos from YouTube between January and March that were relevant to the search keyword "coronavirus" and applied a combination of NLP and network analysis to compute the informational homogeneity. The results showed that, despite small variations regarding the proportion of misinformation on YouTube between the three months analyzed, approximately one third of the content contained certain forms of misinformation. One of the more significant findings of this study is that although misinformation exists on YouTube, it is not concentrated in homogeneous networks filled with predominantly false information—instead, misinformation is moderately associated with non-misinformation. This finding indicates that the YouTube network is not fragmented in the sense that some groups are largely confronted with misinformation while others are not. Since our analysis is limited to the keyword "coronavirus," it would also be interesting for future research to include keywords that are explicitly related to misinformation or conspiracy theories. Thus, network structures based on single conspiracy theories could be investigated to get an even more precise understanding of (mis)informational homogeneity in online networks. Future work may also involve using additional metadata from videos (i.e., visual, audio and subtitles) to improve the automatic classification of misinformation. Also, it would be worthwhile to investigate the spread of misinformation and the identification of relevant actors in the network with their intentions. Our findings could be complemented by analyses of regional differences in the spread of misinformation, to examine whether users in some parts of the world are more likely to receive misinformation on a public health crisis. Addressing these questions could help to assess the actual role of social media platforms in shaping information diffusion processes and fostering the spread of misinformation that could put global health at risk.

CRediT authorship contribution statement

Daniel Röchert: Data curation, Investigation, Conceptualization, Methodology, Software, Validation, Formal analysis, Project administration, Writing – original draft, Writing – review & editing. **Gautam Kishore Shahi:** Data curation, Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **German Neubaum:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing, Funding acquisition. **Björn Ross:** Methodology, Writing – review & editing. **Stefan Stieglitz:** Conceptualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia (Grant Number: 005-1709-0004), Junior Research Group "Digital Citizenship in Network Technologies" (Project Number: 1706dgn009).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.osnem.2021.100164.

References

[1] N. Newman, R. Fletcher, A. Schulz, S. Andi, R. Nielsen, Reuter Institute for the Study of Journalism. Digital news report 2020, 2020 https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf.
[2] H. Ding, J. Zhang, Social media and participatory risk communication during the H1N1 flu epidemic: a comparative study of the United States and China, China Media Research 6 (2010) 80–91. https://scholar.google.com/scholar_lookup?hl=en&volume=6&publication_year=2010&pages=80-90&journal=China+Media+Res&author=Ding+H.&author=Zhang+J.&title=Social+media+and+participatory+risk+communication+during+the+H1N1+flu+epidemic%3A+a+comparative+study+of+the+United+States+and+China.
[3] C. Huang, X. Xu, Y. Cai, Q. Ge, G. Zeng, X. Li, W. Zhang, C. Ji, L. Yang, Mining the characteristics of COVID-19 patients in China: analysis of social media posts, J Med Internet Res 22 (2020) e19087, https://doi.org/10.2196/19087.
[4] Y. Wang, M. McKee, A. Torbica, D. Stuckler, Systematic literature review on the spread of health-related misinformation on social media, Social Science & Medicine 240 (2019), 112552, https://doi.org/10.1016/j.socscimed.2019.112552.
[5] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, ArXiv Preprint ArXiv: 1810.04805. (2018).
[6] N.M. Krause, I. Freiling, B. Beets, D. Brossard, Fact-checking as risk communication: the multi-layered risk of misinformation in times of COVID-19, Journal of Risk Research 0 (2020) 1–8, https://doi.org/10.1080/13669877.2020.1756385.
[7] A. Mitchell, M. Jurkowitz, J. Oliphant, E. Shearer, Three months in, many Americans see exaggeration, conspiracy theories, and partisanship in COVID-19 news, Pew Research Center (2020).
[8] W.H. Organization, Novel Coronavirus (2019-nCoV): situation report, 13, World Health Organization, 2020.
[9] R.S. D'Souza, S. D'Souza, N. Strand, A. Anderson, M.N.P. Vogt, O. Olatoye, YouTube as a source of medical information on the novel coronavirus 2019 disease (COVID-19) pandemic, Global Public Health 15 (2020) 935–942, https://doi.org/10.1080/17441692.2020.1761426.
[10] D.A. Scheufele, N.M. Krause, I. Freiling, D. Brossard, How not to lose the COVID-19 communication war, Issues in Science and Technology 17 (2020). https://issues.org/covid-19-communication-war/.
[11] R. Pathak, D.R. Poudel, P. Karmacharya, A. Pathak, M.R. Aryal, M. Mahmood, A. A. Donato, YouTube as a source of information on Ebola virus disease, North American Journal of Medical Sciences 7 (2015) 306.
[12] K. Bora, D. Das, B. Barman, P. Borah, Are internet videos useful sources of information during global public health emergencies? a case study of YouTube videos during the 2015–16 Zika virus pandemic, Pathogens and Global Health 112 (2018) 320–328.
[13] A. Pandey, N. Patni, M. Singh, A. Sood, G. Singh, YouTube as a source of information on the H1N1 influenza pandemic, American Journal of Preventive Medicine 38 (2010) e1–e3.
[14] A. Gruzd, P. Mai, Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter, Big Data & Society 7 (2020), 205395172093840, https://doi.org/10.1177/2053951720938405.
[15] G. Kawchuk, J. Hartvigsen, S. Harsted, C.G. Nim, L. Nyirö, Misinformation about spinal manipulation and boosting immunity: an analysis of Twitter activity during the COVID-19 crisis, Chiropr Man Therap 28 (2020) 34, https://doi.org/10.1186/s12998-020-00319-4.
[16] R. Kouzy, J. Abi Jaoude, A. Kraitem, M.B. El Alam, B. Karam, E. Adib, J. Zarka, C. Traboulsi, E.W. Akl, K. Baddour, Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter, Cureus (2020) 12, https://doi.org/10.7759/cureus.7255.
[17] G.K. Shahi, A. Dirkson, T.A. Majchrzak, An exploratory study of COVID-19 misinformation on Twitter, Online Social Networks and Media 22 (2021) 100104, https://doi.org/10.1016/j.osnem.2020.100104.
[18] P. Mena, D. Barbe, S. Chan-Olmsted, Misinformation on Instagram: the impact of trusted endorsements on message credibility, Social Media+ Society 6 (2020), https://doi.org/10.1177/2056305120935102, 2056305120935102.
[19] H.O.-Y. Li, A. Bailey, D. Huynh, J. Chan, YouTube as a source of information on COVID-19: a pandemic of misinformation? BMJ Glob Health 5 (2020), e002604 https://doi.org/10.1136/bmjgh-2020-002604.
[20] S. Tasnim, M.M. Hossain, H. Mazumder, Impact of rumors and misinformation on COVID-19 in social media, J Prev Med Public Health 53 (2020) 171–174, https://doi.org/10.3961/jpmph.20.094.
[21] D.A. Scheufele, N.M. Krause, Science audiences, misinformation, and fake news, Proc Natl Acad Sci USA. 116 (2019) 7662–7669, https://doi.org/10.1073/pnas.1805871115.
[22] J. Bright, Explaining the emergence of political fragmentation on social media: the role of ideology and extremism, Journal of Computer-Mediated Communication 23 (2018) 17–33, https://doi.org/10.1093/jcmc/zmx002.
[23] C.R. Sunstein, #Republic: divided democracy in the age of social media, Princeton University Press, Princeton ; Oxford, 2017.
[24] W.H. Organization, Weekly Operational Update on COVID-19 - 6 September 2021. https://www.who.int/docs/default-source/coronaviruse/weekly-updates/wou_20-nov_cleared.pdf.
[25] Y. Kryvasheyeu, H. Chen, N. Obradovich, E. Moro, P. Van Hentenryck, J. Fowler, M. Cebrian, Rapid assessment of disaster damage using social media activity, Science Advances 2 (2016), e1500779.

[26] J. Li, H.R. Rao, Twitter as a rapid response news service: an exploration in the context of the 2008 China earthquake, The Electronic Journal of Information Systems in Developing Countries 42 (2010) 1–22.
[27] A. Goel, L. Gupta, Social Media in the Times of COVID-19, Journal of Clinical Rheumatology : Practical Reports on Rheumatic & Musculoskeletal Diseases. 26 (2020) 220–223. https://doi.org/10.1097/RHU.0000000000001508.
[28] S. Stieglitz, D. Bunker, M. Mirbabaie, C. Ehnis, Sense-making in social media during extreme events, J Contingencies Crisis Man 26 (2018) 4–15, https://doi.org/10.1111/1468-5973.12193.
[29] M. Mirbabaie, D. Bunker, S. Stieglitz, J. Marx, C. Ehnis, Social media in times of crisis: learning from Hurricane Harvey for the coronavirus disease 2019 pandemic response, Journal of Information Technology 35 (3) (2020), 026839622092925, https://doi.org/10.1177/0268396220929258.
[30] A. Mitchell, J. Oliphant, Americans immersed in COVID-19 news; most think media are doing fairly well covering it, Pew Research Center 18 (2020).
[31] H. Allcott, M. Gentzkow, C. Yu, Trends in the diffusion of misinformation on social media, Research & Politics 6 (2019), 205316801984855, https://doi.org/10.1177/2053168019848554.
[32] R. Ehrenberg, Social media sway: Worries over political misinformation on Twitter attract scientists' attention, Science News 182 (2012) 22–25, https://doi.org/10.1002/scin.5591820826.
[33] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (2018) 1146–1151, https://doi.org/10.1126/science.aap9559.
[34] B. Nyhan, J. Reifler, When corrections fail: the persistence of political misperceptions, Polit Behav 32 (2010) 303–330, https://doi.org/10.1007/s11109-010-9112-2.
[35] J. Shin, L. Jian, K. Driscoll, F. Bar, The diffusion of misinformation on social media: Temporal pattern, message, and source, Computers in Human Behavior 83 (2018) 278–287, https://doi.org/10.1016/j.chb.2018.02.008.
[36] D. Jolley, K.M. Douglas, The effects of anti-vaccine conspiracy theories on vaccination intentions, PloS One 9 (2014) e89177.
[37] L. Bode, E.K. Vraga, See something, say something: correction of global health misinformation on social media, Health Communication 33 (2018) 1131–1140, https://doi.org/10.1080/10410236.2017.1331312.
[38] M.J. Wood, K.M. Douglas, R.M. Sutton, Dead and alive: beliefs in contradictory conspiracy theories, Social Psychological and Personality Science 3 (2012) 767–773, https://doi.org/10.1177/1948550611434786.
[39] A. Bessi, M. Coletto, G.A. Davidescu, A. Scala, G. Caldarelli, W. Quattrociocchi, Science vs conspiracy: collective narratives in the age of misinformation, PloS One 10 (2015), e0118093.
[40] G.K. Shahi, D. Nandini, FakeCovid-A multilingual cross-domain fact check news dataset for COVID-19. Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media (2020), in: http://workshop-proceedings.icwsm.org/pdf/2020_14.pdf.
[41] G.K. Shahi , T.A Majchrzak , AMUSED: An Annotation Framework of Multi-modal Social Media Data, ArXiv:2010.00502 [Cs]. (2020). http://arxiv.org/abs/2010.00502 (accessed March 7, 2021).
[42] J.S. Brennen, F. Simon, P.N. Howard, R.K. Nielsen, Types, sources, and claims of Covid-19 misinformation, Reuters Institute 7 (2020).
[43] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C.M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, A. Scala, The COVID-19 social media infodemic, Sci Rep 10 (2020) 16598, https://doi.org/10.1038/s41598-020-73510-5.
[44] J.C.M. Serrano, O. Papakyriakopoulos, S. Hegelich, NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube, in: Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020, 2020.
[45] S. Taylor, B. Pickering, P. Grace, M. Boniface, V. Bakir, danah boyd, S. Engesser, R. Epstein, N. Fawzi, P. Fernbach, D. Fisher, B.G. Gardner, K. Jacobs, S. Jacobson, B. Krämer, A. Kucharski, A. McStay, H. Mercier, M. Metzger, F. Polletta, W. Quattrociocchi, S. Sloman, D. Sperber, C.H.B.M. Spierings, C. Wardle, F. Zollo, A. Zubiaga, Opinion forming in the digital age, Zenodo (2018), https://doi.org/10.5281/ZENODO.1468575.
[46] J. Tucker, A. Guess, P. Barbera, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, B. Nyhan, Social media, political polarization, and political disinformation: a review of the scientific literature, SSRN Journal (2018), https://doi.org/10.2139/ssrn.3144139.
[47] B.E. Weeks, H. Gil de Zúñiga, Six observations for the future of political misinformation research, American Behavioral Scientist (2019), 0002764219878236, https://doi.org/10.1177/0002764219878236.
[48] G. Neubaum, N.C. Krämer, Opinion climates in social media: blending mass and interpersonal communication: opinion climates in social media, Hum Commun Res 43 (2017) 464–476, https://doi.org/10.1111/hcre.12118.
[49] J. Shin, K. Thorson, Partisan selective sharing: the biased diffusion of fact-checking messages on social media: sharing fact-checking messages on social media, J Commun 67 (2017) 233–255, https://doi.org/10.1111/jcom.12284.
[50] M. McPherson, L. Smith-Lovin, J.M. Cook, Birds of a feather: homophily in social networks, Annual Review of Sociology 27 (2001) 415–444, https://doi.org/10.1146/annurev.soc.27.1.415.
[51] E. Bakshy, S. Messing, L.A. Adamic, Exposure to ideologically diverse news and opinion on Facebook, Science 348 (2015) 1130–1132, https://doi.org/10.1126/science.aaa1160.
[52] A. Boutyline, R. Willer, The social structure of political echo chambers: variation in ideological homophily in online networks, Political Psychology 38 (2017) 551–569.
[53] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H.E. Stanley, W. Quattrociocchi, The spreading of misinformation online, Proceedings of the National Academy of Sciences 113 (2016) 554–559, https://doi.org/10.1073/pnas.1517441113.
[54] D. Röchert, G. Neubaum, B. Ross, F. Brachten, S. Stieglitz, Opinion-based homogeneity on YouTube: combining sentiment and social network analysis, Computational Communication Research 2 (2020) 81–108, https://doi.org/10.5117/CCR2020.1.004.ROCH.
[55] L. Tang, K. Fujimoto, M. (Tuan) Amith, R. Cunningham, R.A. Costantini, F. York, G. Xiong, J.A. Boom, C. Tao, Down the rabbit hole" of vaccine misinformation on YouTube: network exposure study, J Med Internet Res 23 (2021) e23262, https://doi.org/10.2196/23262.
[56] E. Hussein, P. Juneja, T. Mitra, Measuring misinformation in video search platforms: an audit study on YouTube, Proc. ACM Hum.-Comput. Interact. 4 (2020) 1–27, https://doi.org/10.1145/3392854.
[57] J. Shin, L. Jian, K. Driscoll, F. Bar, Political rumoring on Twitter during the 2012 US presidential election: rumor diffusion and correction, New Media & Society 19 (2017) 1214–1235, https://doi.org/10.1177/1461444816634054.
[58] R. Fletcher, R.K. Nielsen, Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication, Journal of Communication 67 (4) (2017) 476–498, https://doi.org/10.1111/jcom.12315.
[59] S. Flaxman, S. Goel, J.M. Rao, Filter Bubbles, Echo Chambers, and Online News Consumption, PUBOPQ. 80 (2016) 298–320. https://doi.org/10.1093/poq/nfw006.
[60] P. Mancini, Media Fragmentation, Party System, and Democracy, The International Journal of Press/Politics. 18 (2013) 43–60. https://doi.org/10.1177/1940161212458200.
[61] B.E. Weeks, T.B. Ksiazek, R.L. Holbert, Partisan enclaves or shared media experiences? a network approach to understanding citizens' political news environments, Journal of Broadcasting & Electronic Media 60 (2016) 248–268, https://doi.org/10.1080/08838151.2016.1164170.
[62] J.G. Webster, T.B. Ksiazek, The dynamics of audience fragmentation: public attention in an age of digital media, Journal of Communication 62 (2012) 39–56, https://doi.org/10.1111/j.1460-2466.2011.01616.x.
[63] T. Harper, The big data public and its problems: Big data and the structural transformation of the public sphere, New Media & Society 19 (2017) 1424–1439, https://doi.org/10.1177/1461444816642167.
[64] A. Sîrbu, D. Pedreschi, F. Giannotti, J. Kertész, Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model, PLoS ONE 14 (2019), e0213246, https://doi.org/10.1371/journal.pone.0213246.
[65] D. Freelon, M. Lynch, S. Aday, Online fragmentation in wartime: a longitudinal analysis of tweets about Syria, 2011–2013, The ANNALS of the American Academy of Political and Social Science 659 (2015) 166–179.
[66] J.L. Nelson, H. Taneja, The small, disloyal fake news audience: the role of audience availability in fake news consumption, New Media & Society 20 (2018) 3720–3737, https://doi.org/10.1177/1461444818758715.
[67] J.M. Sumiala, M. Tikka, Broadcast yourself-global news! a netnography of the "Flotilla" news on YouTube: broadcast yourself-global news!, communication, Culture & Critique 6 (2013) 318–335, https://doi.org/10.1111/cccr.12008.
[68] G. Stocking, P. van Kessel, M. Barthel, K.E. Matsa, M. Khuzam, Many Americans get news on YouTube, where news organizations and independent producers thrive side by side, Pew Research Centre (2020).
[69] D. Zhang, L. Zhou, J. Lim, From networking to mitigation: the role of social media and analytics in combating the COVID-19 pandemic, Information Systems Management (2020) 1–9, https://doi.org/10.1080/10580530.2020.1820635.
[70] T. Wang, K. Lu, K.P. Chow, Q. Zhu, COVID-19 sensing: negative sentiment analysis on social media in China via BERT model, IEEE Access 8 (2020) 138162–138169, https://doi.org/10.1109/ACCESS.2020.3012595.
[71] Y. Geng, Z. Lin, P. Fu, W. Wang, Rumor detection on social media: a multi-view model using self-attention mechanism, in: J.M.F. Rodrigues, P.J.S. Cardoso, J. Monteiro, R. Lam, V.V. Krzhizhanovskaya, M.H. Lees, J.J. Dongarra, P.M. A. Sloot (Eds.), Computational Science – ICCS 2019, Springer International Publishing, Cham, 2019, pp. 339–352, https://doi.org/10.1007/978-3-030-22734-0_25.
[72] E. Masciari, V. Moscato, A. Picariello, G. Sperli, A deep learning approach to fake news detection, in: D. Helic, G. Leitner, M. Stettinger, A. Felfernig, Z.W. Raś (Eds.), Foundations of Intelligent Systems, Springer International Publishing, Cham, 2020, pp. 113–122.
[73] K. Florio, V. Basile, M. Polignano, P. Basile, V. Patti, Time of Your Hate: The Challenge of Time in Hate Speech Detection on Social Media, Applied Sciences. 10 (2020) 4180. https://doi.org/10.3390/app10124180.
[74] J. Pavlopoulos, N. Thain, L. Dixon, I. Androutsopoulos, Convai at semeval-2019 task 6: Offensive language identification and categorization with perspective and bert, in: Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 571–576.
[75] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, ArXiv Preprint ArXiv:1412.6980. (2014).

[76] D. Krackhardt, R.N. Stern, Informal Networks and Organizational Crises: An Experimental Simulation, Social Psychology Quarterly 51 (1988) 123–140, https://doi.org/10.2307/2786835.
[77] W.H. Organization, Naming the coronavirus disease (COVID-19) and the virus that causes it, (2020). https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it.
[78] C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, Y. Chang, Abusive language detection in online user content, in: Proceedings of the 25th International Conference on World Wide Web, 2016, pp. 145–153.
[79] B. van Aken, J. Risch, R. Krestel, A. Löser, Challenges for toxic comment classification: An in-depth error analysis, ArXiv Preprint ArXiv:1809.07572. (2018).
[80] D. Röchert, G. Neubaum, B. Ross, S. Stieglitz, Caught in a networked collusion? homogeneity in conspiracy-related discussion networks on YouTube, Information Systems (2021), 101866, https://doi.org/10.1016/j.is.2021.101866.