The Disinformation Dozen: An Exploratory Analysis of Covid-19 Disinformation Proliferation on Twitter

Gianluca Nogara
gianluca.nogara@supsi.ch
University of Applied Sciences and Arts of Southern Switzerland
Switzerland

Padinjaredath Suresh Vishnuprasad
psvp28@protonmail.ch
IISER-Tirupati
India

Felipe Cardoso
felipe.cardoso@supsi.ch
University of Applied Sciences and Arts of Southern Switzerland
Switzerland

Omran Ayoub
omran.ayoub@supsi.ch
University of Applied Sciences and Arts of Southern Switzerland
Switzerland

Silvia Giordano
silvia.giordano@supsi.ch
University of Applied Sciences and Arts of Southern Switzerland
Switzerland

Luca Luceri
luca.luceri@supsi.ch
University of Applied Sciences and Arts of Southern Switzerland
Switzerland

Abstract

Shortly after the outbreak of the novel coronavirus disease (Covid-19), the United Nations declared an infodemic due to an unprecedented amount of false information spreading about Covid-19. A study made by the center for countering digital hate found out that twelve individuals, referred to as *Disinformation Dozen* (*Disinfo12*), were responsible for 65% of Covid-19 misinformation circulating on social media. Given the *Disinfo12*'s detrimental impact in spreading misinformation, in this work, we perform an exploratory analysis on *Disinfo12*'s activity on Twitter aiming at identifying their sharing strategies, favorite sources of information, and potential secondary actors contributing to the proliferation of questionable narratives. In our study, we uncovered the distinctive facets that allowed *Disinfo12* to act as primary sources of information, and we recognized that YouTube represent one of the favorite information sources to spread questionable narratives and conspiracy theories. Finally, we recognized that right-leaning accounts are embedded in *Disinfo12*'s community and represent the main spreaders of content generated by the *Disinformation Dozen*.

Keywords

covid-19, disinformation, misinformation, social media

ACM Reference Format:
Gianluca Nogara, Padinjaredath Suresh Vishnuprasad, Felipe Cardoso, Omran Ayoub, Silvia Giordano, and Luca Luceri. 2022. The Disinformation Dozen: An Exploratory Analysis of Covid-19 Disinformation Proliferation on Twitter. In *14th ACM Web Science Conference 2022 (WebSci '22), June 26–29, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3501247.3531573

WebSci '22, June 26–29, 2022, Barcelona, Spain

ACM ISBN 978-1-4503-9191-7/22/06…$15.00
https://doi.org/10.1145/3501247.3531573

1 Introduction

Since the outbreak of the novel coronavirus disease (Covid-19), Online Social Networks (OSNs) have been playing a crucial role in disseminating information about the virus and spreading guidelines to protect against the disease. Global organizations such as the World Health Organization (WHO) and other major health-related centers such as the Centers of Disease Control and Prevention (CDC) have relied on social networks to disseminate guidance to help protect against the spread of Covid-19. Conversely, propagandists, conspiracy theorists, and speculators, exploiting the scarce knowledge about the disease and humans' susceptibility to false information [1, 17, 19, 26, 31], have massively contributed to the spread of misleading narratives relative to the Covid-19 virus [25], undermining the veracity of online information and the health of those who, unaware of the malicious nature of the information, believed it. The proliferation of false information about the virus has led the United Nations to declare a Covid-19 infodemic [16]. Researchers and news outlets attempted to mitigate the rising infodemic by showing examples of cases that restricted people from acting properly during the pandemic [10, 12, 21, 28, 32, 35, 40]. Conspiracy theories spread rumors regarding the cure for Covid-19, with some suggesting home-made remedies, preventing people from focusing properly on how to mitigate the spread and impact of the disease. Other conspiracies falsely claiming that 5G damages the immune system and that it is the cause of disease outbreak gained traction online and eventually resulted in vandalism of 5G cell towers [4].

Consequently, governments and health organizations pressured social media to fight the infodemic and limit the spread of false news to ensure people were not misled by medically inaccurate information. However, social media platforms were not prepared to face such a global and rapidly spreading infodemic [22]. As a result, propagandists and conspiracy theorists were capable of massively spreading false claims about Covid-19 [12]. This spread of misinformation further continued with the introduction of Covid-19 vaccines, with conspiracy theorists claiming that Covid-19 vaccines will implant microchips into people or that Covid-19 vaccine will alter people's DNA to make them into genetically modified organisms [36]. These and other similar false theories penetrated through society leading to vaccination hesitancy and contributing to the rise of the what is known as *No Vax* or *Anti-Vax* movements [3, 24, 41]. A report of the center for countering digital hate highlights that 65% of misinformation circulating on social media originated from 12 individuals only [11, 20]. The report goes further and identifies the accounts of these 12 individuals, which are referred to as *Disinformation Dozen* (in short *Disinfo12*). We will describe in more detail the *Disinfo12* in Section 2. The fact that only twelve accounts were capable of playing such a significant role in spreading misinformation globally raises several questions regarding the means they adopted to spread their theories and involve other accounts in the diffusion of questionable narratives.

In this work, we aim at identifying the strategies adopted by the *Disinfo12* to spread misinformation on Twitter, the topics they primarily concentrated on, and their sources of information. Our purpose is also to uncover secondary actors contributing to the spread of misinformation originated by the *Disinfo12* by analyzing the interaction and endorsement that these twelve accounts receive from other users. To reach these aims, we first analyze the hashtags and keywords that were primarily used by the *Disinfo12*, their sharing activity performed on Twitter (e.g., original tweets, retweets, shared links and videos, etc.), and their main topics of discussion, comparing them to a set of users that contributed positively to the spread of accurate information during the Covid-19 pandemic (we provide more details about these users in Section 2). By contrasting this set of quality information spreaders with the *Disinfo12*, we then analyze the sources of information both group of accounts rely on, and their position in the social network while observing possible secondary actors that contributed to the spread of (mis-)information. To systematically discuss our findings, we pose the following research questions (RQs) and then address them in Sec. 5:

- RQ1: What are the primary online activities of *Disinfo12* and how do they differ from those of other users?
- RQ2: Which are the sources of information the *Disinfo12* relied on?
- RQ3: How is *Disinfo12*'s online community composed? Are there secondary actors contributing to the spread of misinformation?

To the best of our knowledge, our work is the first to examine in detail the online activity of the *Disinfo12* and to investigate their strategies in spreading misinformation on Twitter. The contributions of this work can be summarized as follows.

- We gathered a novel dataset that encompasses the messages published on Twitter by *Disinfo12*, including the tweets of the accounts that endorsed their sharing activity. The collected dataset, made available to the research community, also includes tweets from quality information spreaders and their followers.
- We investigated the online activities of *Disinfo12* recognizing their attempt to take the role of primary sources of information. We also uncover their distinctive traits in sharing original content and elicit distrust in the health system within their community.
- We examined the information sources leveraged by *Disinfo12* to share their content, and we identified proprietary websites and YouTube as their favorite resources to propagate questionable information and conspiracy theories.
- We explored the online communities around the *Disinfo12*, recognizing low-credibility media outlets, conservatives political figures, and journalists as secondary actors contributing to the proliferation of misleading narratives.

The paper is organized as follows. Section 2 provides an overview of the *Disinfo12*, while Section 3 discusses related work. In section 4, we discuss the data collection process and the resulting dataset. In Section 5, we present our exploratory analysis to address the aforementioned research questions. Section 6 discusses the main findings and concludes the paper.

2 The Disinformation Dozen

The Center for Countering Digital Hate published a report entitled *The Disinformation Dozen* where twelve users, or *anti-vaxxers*, as referred to in the report, were identified as being responsible for waves of misinformation relative to Covid-19 and anti-vaccine claims circulating on social media [11, 20]. The report highlights that the motives driving these 12 accounts were primarily economical.

In Table 1, we report the names of these accounts, their occupation, status on Twitter, and number of followers and retweets. Table 1 shows that four users are alternative medicine activists and in total nine have occupations in the health sector. In Figure 1, we show two exemplary tweets of two members of *Disinfo12* [11], with Rizza Islam claiming to have beaten Covid-19 following a homemade diet, and Sherri Tenpenny claiming that masks are not used to limit the spread of the disease but instead to control and suppress the immune system. These are just some examples of many of *Disinfo12*'s online messages either claiming to have ground-breaking cures, demoting vaccines, or spreading novel conspiracy theories[1].

In our study, and to contrast the strategies and means of the *Disinfo12*, we consider another set of users, referred to as *GoodInfo*, that contributed in spreading of quality and accurate information on social media to help raise awareness for combating Covid-19 diffusion. The set of *GoodInfo* considered in our analysis consists of 260 Twitter accounts. These accounts were selected from among 350 accounts[2] (260 user and 90 media outlet accounts) that were listed in a report of [27] as *to follow on Covid-19* due to their proliferation of accurate information relative to Covid-19 on OSNs, to their involvement in the health sector, and for having scientific competences[3].

3 Misinformation in Covid-19

The unprecedented spread of misinformation on online social networks during the Covid-19 era attracted the utmost interest of the research community. In this section, we discuss some of the studies investigating misinformation on Covid-19 [2, 7, 8, 30, 38, 39] and

[1] We refer the reader to [11] for more examples.
[2] We discarded media outlet accounts from our analysis and considered only accounts relating to users (260 in total).
[3] At the time of the writing of this paper, a new report shortlisting 100 accounts of the 350 accounts was published [27]. To validate our findings, we repeated our analysis considering only those 100 accounts with no significant changes on the results.

Table 1: Name, occupation, account status, number of followers, and number of retweets of the *Disinfo12* (up to August 2021)

Name	Occupation	Status	Followers	Retweet
Joseph Mercola	osteopathic physician	Active	314,302	158,962
Robert Kennedy, Jr.	environment supporter	Active	311,481	512,840
Christiane Northrup	obstetrics and gynecology	Active	115,215	16,764
Rashid Buttar	osteopathic physician	Active	86,368	596
Erin Elizabeth	alternative medicine activist	Active	39,384	81,479
Kelly Brogan	alternative medicine activist	Active	18,618	-
Sayer Ji	alternative medicine activist	Active	11,285	716
Kevin Jenkins	head of a novax group	Active	897	389
Sherri Tenpenny	osteopathic physician	Suspended	-	-
Ben Tapper	chiropractor	Suspended	-	-
Ty and Charlene Bollinger	alternative medicine activist	Suspended	-	-
Rizza Islam	novax activist	Suspended	-	-

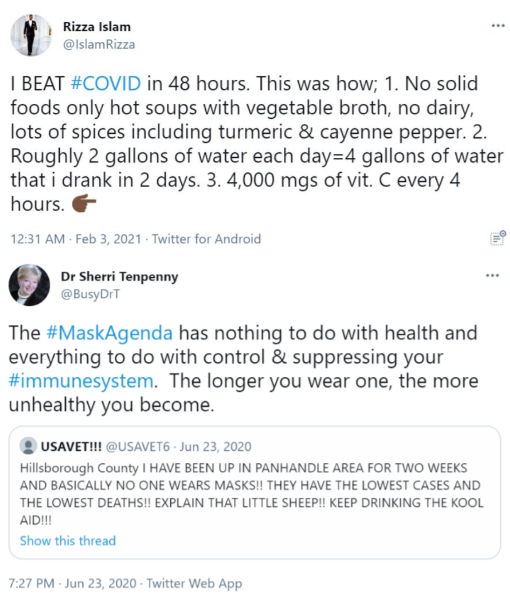

Figure 1: Exemplary tweets of members of the *Disinfo12*

the researches investigating the *Disinfo12*. Several studies have investigated the spread of Covid-19 questionable narratives on OSNs focusing either on its impact on preventing the spread of the disease or on people's response to Covid-19, or even on proposing strategies to cope with misinformation spread. [7] is one of the early works studying misinformation on Covid-19 but not strictly limited to online social networks. The study shows that most websites reporting scientific content related to Covid-19 were not health-related websites, contributing to the spread of misinformation over the Internet. In [30], authors investigated whether the use of social media is helping diffusing information or misinformation with regard to the Covid-19 outbreak. Their findings show that social media users, unknowingly, contributed to the spread of misinformation [30]. [2] investigated the effect of misinformation on Covid-19 on individual responses and showed, through self-administered online surveys, that misinformation circulating on OSNs have impacted people's responses against Covid-19 and their belief in vaccines. [38] also discussed the impact of Covid-19-related false information in OSNs on the prevention and cure of the disease. [38] further suggested measures to be adopted by social media based on the use of advanced technologies like natural language processing and data mining approaches to detect and remove information with no scientific basis. Other works such as [8] investigated the sources of Covid-19 misinformation on articles and on Twitter. In particular, the study identifies the most prominent topics of Covid-19 misinformation. The analysis shows that Trump mentions on Twitter was the most linked topic to Covid-19 misinformation, compromising around 40% of the overall misinformation conversation. The study also shows that the majority of the information covered by media was not double-checked before publishing. In [14], the author also examines the source of misinformation showing that, while the *Disinfo12* were the primarily sources of misinformation on Covid-19, state-sponsored actors and political extremists from the far right significantly contributed to the spread of misinformation.

In addition to these efforts, very few recent works investigated the misinformation linked to *Disinfo12* however presenting limited research. For example, [20] analyzes three accounts of the *Disinfo12* showing examples of misinformation through tweets and highlights the use of their websites to sell own products, which they falsely claim to heal from Covid-19, while [33] gave examples of censorship of claims made by *Disinfo12* related to Covid-19. To the best of our knowledge, no previous works have examined in detail the behavior of the *Disinfo12*. Due to *Disinfo12*'s heavy involvement in spreading misinformation, we consider of pivotal importance to investigate their online activity and strategy in detail. To this aim, in this work, we make the first attempt in characterizing and analyzing the behavior of *Disinfo12* and we hope that our study serves as a starting point for deeper analyses on *Disinfo12*.

4 Data Collection and Processing

This section provides a description of the collection process executed to build our dataset, which we make available to the research community[4]. The data collection process consists in the gathering of:

- The tweets created by the *Disinfo12* and *GoodInfo* accounts. This data collection was carried out by using the Twitter API and could gather up to 3,200 users' most recent tweets.
- The retweets received by the *Disinfo12* and *GoodInfo* from September 2020 to August 2021. This data collection was carried out by using the Twitter API for Academic Research, which allowed us to search for the retweets received by the accounts under investigation.

Table 2 reports statistical information relative to the activity of the *Disinfo12* and *GoodInfo*. In particular we focus on tweets posted by the account and on retweets received by an account (i.e., when another account retweets a tweet posted by the user in consideration). Table 2 shows that *GoodInfo* have higher number of average tweets per account (2,887) than the *Disinfo12* (1,596). Regarding the retweets however, the *Disinfo12* have significantly higher average retweets per account (around 4 times more) and per tweet (10 times more) than *GoodInfo*, revealing that tweets of *Disinfo12* were re-shared significantly more than those of *GoodInfo*.

Figure 2 portrays the interaction network of the users retweeting the *Disinfo12*. The yellow nodes represent the *Disinfo12* while the nodes in red are the users that re-shared their content. The node size is proportional to the in-degree, which corresponds to the number of retweets received. Interestingly, the most prominent accounts, such as *RobertKennedyJr*, are endorsed by a large base of accounts that do not retweet any other member of the *Disinfo12*, while less prominent *Disinfo12* users share their followers. Furthermore, in the superimposed insets in Figure 2, we display some of the websites owned by four members of the *Disinfo12*, which are heavily used in the misinformation diffusion (as detailed in Section 5.2).

5 Exploratory Analysis

In this Section, we address the RQs posed in the introduction and discuss the main findings of our exploratory analysis.

5.1 RQ1: Sharing Activities

To address *RQ1*, we start by inspecting the online activities of the accounts under scrutiny. We disentangle tweets in diverse sharing activities as follows:

- Original Tweet: posts containing original content. May contain text, images, hyperlinks, and mentions to other accounts.
- Retweet: Re-sharing a tweet of another account. A retweet contains all information of the original post.
- Quote: A quote is a specific type of retweet containing additional comments added by the author of the quote.
- Reply: A reply is a comment to an existing tweet.

Figure 3 shows the distribution of the different sharing activities of *Disinfo12* and *GoodInfo*. First, we notice that there is a significant difference in the percentage of *original tweets* generated by the *Disinfo12* and the *GoodInfo*. Specifically, *original tweets* makes up around 70% of the *Disinfo12* activity while consists of only around 20% of the *GoodInfo* shared messages. This suggests that the *Disinfo12* primarily tend to act as sources of information, generating new content to circulate on social media. Also in terms of *retweets*, there is a notable difference between the behavior of the two sets of users. *Retweets* make up only 15% of the activity of *Disinfo12* while around 40% of that of the *GoodInfo*. This means that *GoodInfo* tend to retweet other accounts, from users to newspapers, suggesting that they rely on, and allegedly endorse, other sources of information. *Disinfo12*, on the contrary, do not massively leverage the retweet activity, which may represent a strategy to reinforce their role of primary information source. Another interesting point can be seen looking at the percentage of *replies*. *GoodInfo* tend to engage in discussions, with 30% of their activity as *replies*, showing their willingness of interacting with other accounts, mainly in health-related questions and inquiries. Unlike-wise, *Disinfo12*, barely reply to other accounts (less than 10% of their tweets), which might suggest a lack of interest in engaging in discussions.

We now investigate the hashtags and the content published by both *Disinfo12* and *GoodInfo* in their tweets with the aim of identifying the topics covered by the different set of users under observation. We start by inspecting the most used hashtags by each class of accounts. Figures 4 and 5 show a treemap displaying the 10 most used hashtags for *GoodInfo* and *Disinfo12*, respectively. The treemap has three dimensions: the position of the boxes (from left to right, highly occurring to the left while least occurring to the right), the size of the box (the larger the more occurring), and the color opacity (the darker the more occurring). The three dimensions are based on the number of occurrences within the set of collected tweets[5]. Analyzing the treemap for each set of users, we see that all hashtags used by *GoodInfo* refer to Covid-19 discussions (e.g., *#COVID19*, *#vaccine*, etc.). In particular, we see that *#COVID19* dominates the hashtags used by *GoodInfo*. On the contrary, the treemap of the *Disinfo12* shows that several hashtags do not relate to Covid-19 discussions in a direct way, such as *#climatechange* (ranked third) and *#Trump* (ranked fourth). This shows that *GoodInfo* focus is primarily on facts related to the coronavirus pandemic situation, while *Disinfo12* tend to share trending content not necessarily related to the discussion about Covid-19. It is worth noting that the *Disinfo12* heavily utilize the hashtags *#Monsato* and *#TheDefender*. The former is an American agrochemical and agricultural biotechnology corporation, while the latter is tied with the website Children's Health Defense managed by the *Disinfo12* Robert Kennedy Jr. Several media cases were systematically reported on Twitter by referencing *Disinfo12's* websites. Among those, Children's Health Defense represents a prime example of a web page tied with *Disinfo12's* that diffuses controversial messages against pharmaceutical companies, in general, and vaccines, in particular.

To further investigate the topics that differentiate *Disinfo12* and *GoodInfo*, we examine the content of their published tweets. For this purpose, we use the Sparse Additive Generative Models of Text (SAGE) [29], a text differentiation method that, starting from two or more corpuses of text, finds the distinctive words that most characterize each corpus. SAGE assigns a score to each word, which is calculated based on the occurrence of the word in each corpus. If

[4]The IDs of the collected tweets are available at the following link: https://github.com/gi-ux/Disinformation-Dozen-TweetIDs

[5]Note that we aggregated hashtags referring to same topic such as, e.g., *Covid19* and *covid19*.

Table 2: Statistics relative to the activity of the *Disinfo12* and *GoodInfo* in the collected dataset

Users	Nr. Accounts	Nr. Tweets	Nr. Retweets	Avg. Tweets per Account	Avg. Retweets per Account (per Tweet)
Disinfo12	12	19,159	771,746	1,596	64,312 (40.3)
GoodInfo	260	750,820	3,467,039	2,887	13,334 (4.6)

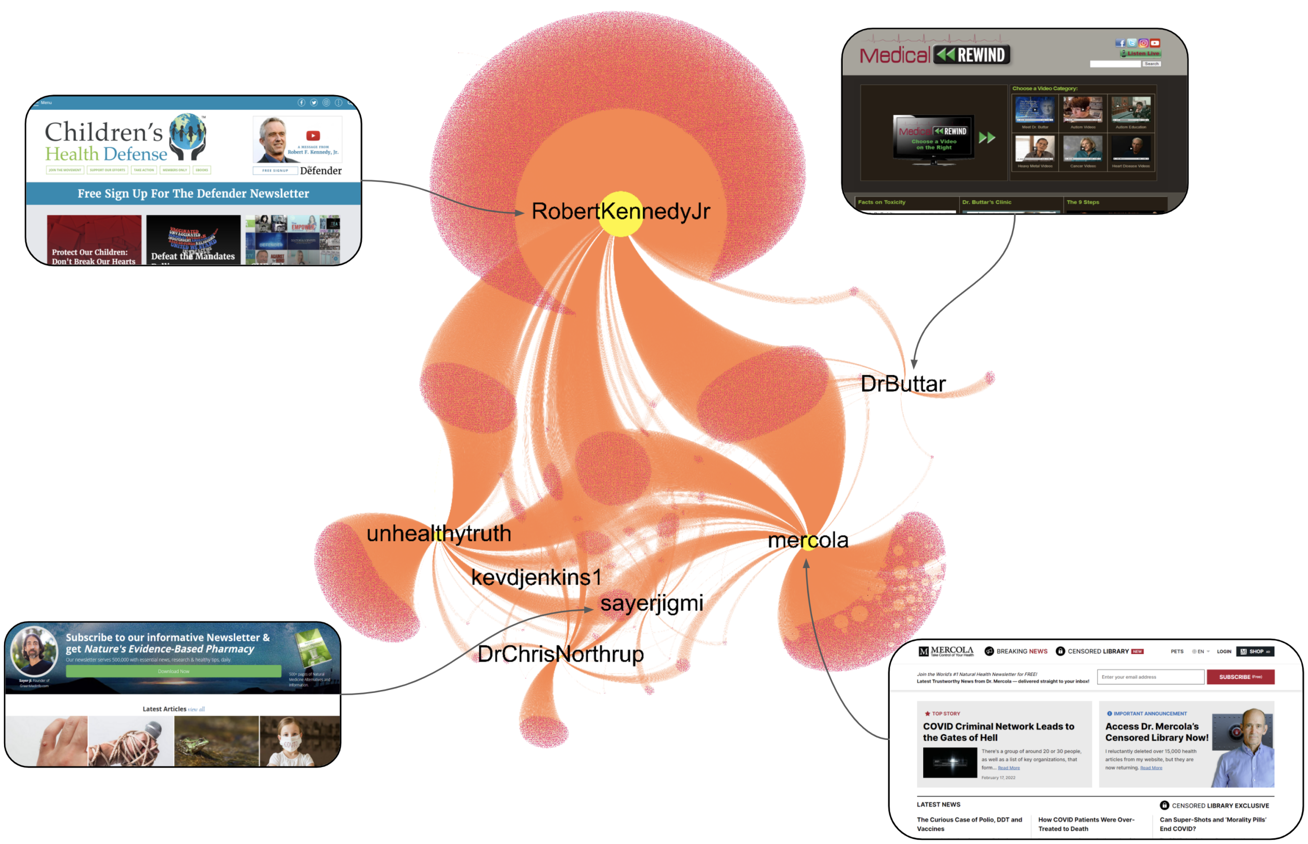

Figure 2: Interaction network among users retweeting the *Disinfo12*. The yellow nodes represent the *Disinfo12* while the users that re-shared their content are represented in red. Node size is proportional to the in-degree. Websites of some *Disinfo12* members are in insets.

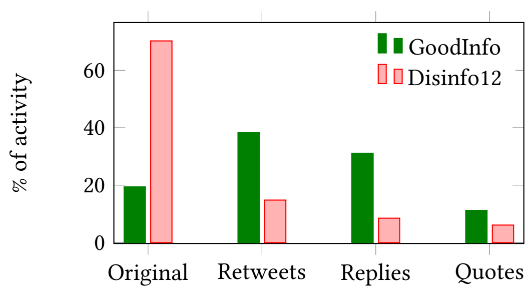

Figure 3: Distribution of sharing activities for *GoodInfo* and *Disinfo12*

a word occurs frequently in both corpuses, it is given a low score. On the contrary, words occurring frequently only in one corpus are given a high score. Before applying SAGE, we first clean the text within each tweet by removing stop-words, mentions, and special characters. Then, we generate two corpuses of texts, one containing tweets shared by *GoodInfo* and the other including tweets generated by *Disinfo12*. Finally, we apply SAGE to identify the distinctive topics that mainly characterize the two categories of users. We also perform the same operation by only considering the hashtags in the tweets, i.e., we build a corpus of hashtags for every set of users. The rationale is to narrow our previous analysis by excluding common hashtags (e.g., all the generic hashtags related to Covid-19) to further highlight the peculiar hashtags that differentiate the two groups.

Figure 4: Treemap of the 10 most used hashtags by *GoodInfo*

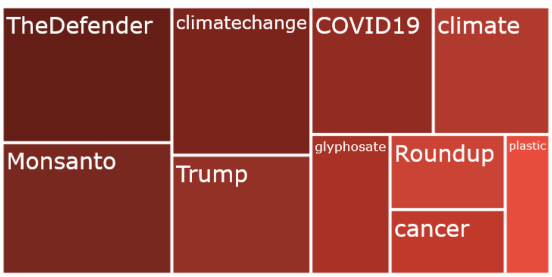

Figure 5: Treemap of the 10 most used hashtags by *Disinfo12*

Table 3 reports SAGE outcomes by listing the most distinctive keywords and hashtags shared by every set of users. Comparing the two lists, we see that topics shared by *GoodInfo* are related to Covid-19, e.g., *covax* and *sarscov2*, with also some topics such as *africa* that links to Covid-19 discussions in Africa. In contrast, the most discussed topics by *Disinfo12* do not reveal a direct relationship with Covid-19 discussions or the health sector. For instance, *thedefender*, *zook* and, *orthostatic* are not directly connected with Covid-19 discussions. As for the hashtags, those used by the *GoodInfo* relate to different contexts including Covid-19 discussion (*#deltavariant*) and politics (*#bidenharris2020*, *#rgblegacy*, *#scotusnominee*, and *#trumpinsurrection*). As for the hashtags used by the *Disinfo12*, we observe that they refer to a range of topics such as censorship (e.g., *#censoring*, *#censorship*), chemical/pharmaceutical companies (e.g., *#bayer*, *#monsanto*) up to controversial events related to vaccines (e.g., *#gates*, *#hankaaron*), which might suggest that the *Disinfo12* strategically leverage controversial events to raise fear and distrust in the health system within their community.

5.2 RQ2: Information Sources

To address *RQ2*, we inspect the links shared in their tweets by the accounts under investigation. Specifically, we explore the URL domains mainly published by both the *GoodInfo* and the *Disinfo12*.

Table 3: List of distinctive topics and hashtags used by *GoodInfo* and *Disinfo12* extracted with SAGE

GoodInfo		Disinfo12	
Topics	Hashtags	Topics	Hashtags
edr	bidenharris2020	thedefender	gates
rasmussen	georgia	zook	hankaaron
drc	rgblegacy	classen	censorship
covax	russia	dressen	thedefender
epi	scotusnominee	orthostatic	monsanto
africa	deltavariant	miller12345	gardasil
preprint	italy	enlarged	bayer
sarscov2	russian	aldous	epa
postdoc	gopsedition	bitches	censoring
malaria	trumpinsurrection	bicycled	sharks

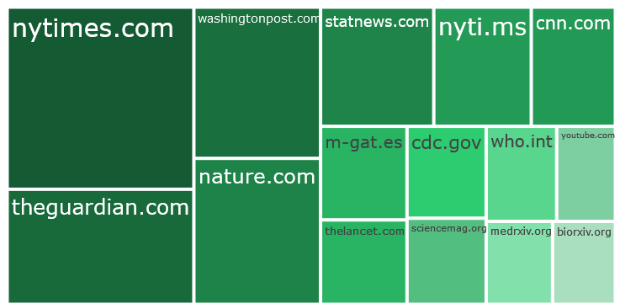

Figure 6: Treemap of the top 15 URL domains shared by *GoodInfo*

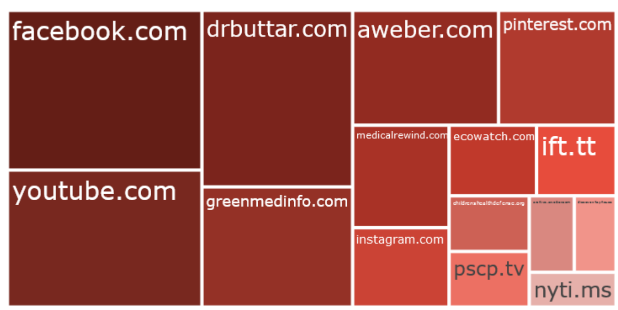

Figure 7: Treemap of the top 15 URL domains shared by *Disinfo12*

The sharing of links to web pages is considered an important tool for broadcasting information as it allows to refer to media outlets, videos, or even personal blogs and posts on social media. Figures 6 and 7 show a treemap of the URL domains mainly shared by the

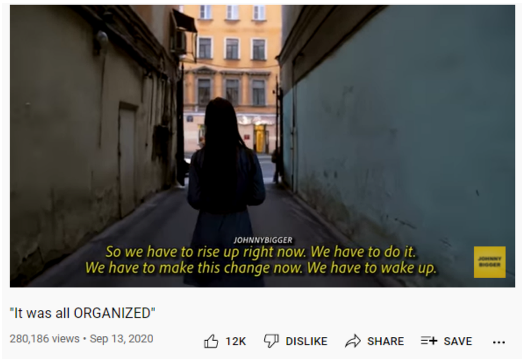

Figure 8: An example of YouTube video shared by *Disinfo12* on Twitter

Table 4: Titles and descriptions of videos shared on YouTube by *GoodInfo* and *Disinfo12* extracted by SAGE

GoodInfo		Disinfo12	
Title	Description	Title	Description
covid	virology	right	amend
coronavirus	covid	told	dore
health	infectious	byron	fox
live	director	repartitions	cable
science	viruses	truth	fnc
vaccines	coronavirus	agenda	valuetainment
watch	outbreaks	election	awakening
virus	development	americans	sears
public	vaccines	pleidian	nrule
vaccine	virus	heated	censorship

GoodInfo and the *Disinfo12*, respectively. Comparing the treemaps, we see that the most shared URL domains are of diametrically opposed origins. For *GoodInfo*, the primarily sources of information are newspapers, e.g., *The New York Times*, *The Guardian*, and *The Washington Post*, and scientific journals such as *Nature*. *Disinfo12*, on the contrary, primarily rely on *Facebook* posts and *Youtube* videos and on web-sites owned by *Disinfo12* such as *drbuttar.com*, *medical-rewind.com*, and *greenmedinfo.com*. This shows a clear distinction between the sources and types of information *Disinfo12* and *GoodInfo* leverage to spread content.

It is worth noting that the treemaps also show that *YouTube* is the only domain in common between the two sets of users, however with a lower degree for *GoodInfo* (ranked 14^{th}) than for *Disinfo12* (ranked 2^{nd}). Based on this finding, we investigate in more detail the use of *YouTube* for the proliferation of information on Twitter performing two analyses.

The first analysis consists in checking the status of the *YouTube* videos. A video that is currently not available on YouTube means that it has been either removed by *YouTube* or its owner, or that the account of the video has been banned, which may indicate that the video does not respect YouTube's policies, e.g., it might contain questionable or misleading information. We inspect the URL links of videos shared by each set of users and report if a video is still online or if it has been removed. The findings reveal that 45% of the videos shared by *Disinfo12* are unavailable (as banned videos or accounts) while only 5% of videos shared by GoodInfo are removed. This suggests that the content shared by *Disinfo12* has been considered not appropriate for the YouTube community and, allegedly, subject to banning to a larger extent if compared to *GoodInfo*'s shared videos.

As a second analysis, we collect the title and description of each *YouTube* video shared by each set of users and then analyze, using SAGE, the topics treated by each set of users. Table 4 reports the list of distinctive words that are used in the titles and descriptions of videos shared by both *GoodInfo* and *Disinfo12*. The list of *GoodInfo* contains words clearly tied with health-related discussions (abundance of words such as *health*, *vaccines*, *science*, *infectious*, and *outbreaks*). In contrast, the title and descriptions of videos shared by *Disinfo12* barely contains words that can directly relate to health discussions and are rich of controversial and suspicious words, such as *agenda*, *truth*, *awakening*, and *censorship*. As an example, in Figure 8, we show a screenshot of a YouTube video shared by *Disinfo12*, with the caption *"It was all ORGANIZED"*, hinting to a conspiracy theory behind the Covid-19 outbreak.

To further characterize the information sources used in the Covid-19 discussion and, in particular, by the communities around the *Disinfo12* and *GoodInfo* users, we categorize the shared URL domains based on their credibility. Similarly to [9, 13, 34, 43], we classify the URLs in low- and high-credibility information sources by leveraging the reviews provided by the Media Bias/Fact Check website (mediabiasfactcheck.com). A high-credibility source is assigned a value of 1, while a low-credibility source is assigned a value of 0. We then compute for each user a credibility score ranging between 0 and 1, which represents the proportion of the shared high-credibility URLs over all the published URLs.

Figure 9 portrays the credibility score of accounts that exclusively endorse (by means of retweets) *Disinfo12* or *GoodInfo* users. We refer to these users as *Disinfo12 followers* and *GoodInfo followers*, respectively. In this categorization, we do not consider accounts that re-share content generated by both *Disinfo12* and *GoodInfo* users. From Figure 9, it can be noticed how *Disinfo12 followers* exhibit a much lower credibility score with respect to *GoodInfo followers*, which means they rely on low-credibility information sources to a larger extent if compared to *GoodInfo followers*. This, in turn, suggests how *Disinfo12 followers* play a role in the diffusion of questionable information, further contributing to, and allegedly exacerbating, the spread of misinformation. The role of such secondary actors and their position within the *Disinfo12* community is further explored in the next section.

5.3 RQ3: Network and Communities

To address *RQ3*, we evaluate the position of the two classes of users within the social network. For this purpose, we consider the *retweet network* and the communities within this network. A *retweet* refers to the scenario when a user re-shares a tweet of another user. This sharing activity is also usually tied with the concept of social endorsement [23, 37]. In this context, examining to which degree certain users are retweeted and by who allows us to understand to what extent these users are endorsed and supported within

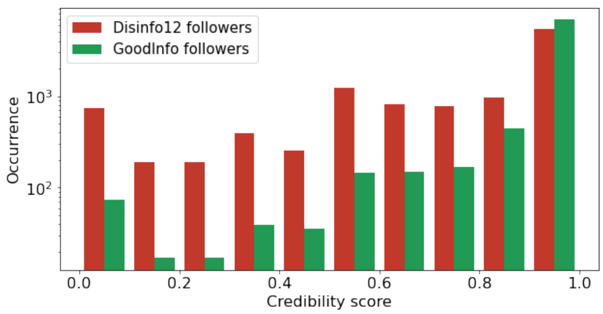

Figure 9: Credibility score of *Disinfo12 followers* and *GoodInfo followers*

their communities. To capture these relationships between users, a *retweet network*, where users are represented by nodes and retweets are represented by edges, can be constructed and analyzed. Notice that an edge originates at the node who retweets and terminates at the node whose tweet has been retweeted.

We build the retweet network by using data from an existing repository [5], which consists of tweets related to general discussions about Covid-19. The rationale is to inspect the sets of users under analysis in a more general and broad discussion. Note that our dataset (used so far in the study) is specifically built around the *Disinfo12* and *GoodInfo* users and includes only their and their *followers* activities, thus, disregarding any other actor involved in the Covid-19 discussion.

As the dataset in [5] only specifies the tweet IDs, we applied a hydration process to get all the information fields relative to every tweet. From the gathered set of tweets, we considered only the retweets and built a retweet network. Overall, the retweet network consists of almost 13 million nodes representing users and about 48 million edges representing user interactions in the form of retweet. Within the retweet network, we recognize three classes of nodes: *i)* the *Disinfo12* and the *Disinfo12 followers* (74,231 nodes), *ii)* the *GoodInfo* and the *GoodInfo followers* (153,551 nodes), and *iii)* the accounts that either retweet both *Disinfo12* and *GoodInfo* or none of them (12,285,720 nodes).

To investigate how these users are connected and grouped together within the retweet network, we apply the Louvain method to extract communities within our network. From the community extraction, a large number of communities was found (around 300,000). In particular, our findings show that only 10 of these communities group 74% of the users in our dataset. According to this result, we consider the largest 10 communities in our analysis, and we manually inspected the actors mainly involved and the topics primarily discussed to characterize each of them.

Table 5 describes these communities in terms of topic and language used in the discussion, and also reports the number of users that retweeted *Disinfo12* and *GoodInfo* users. Among the ten communities, several communities can be of interest to our analysis. For instance, besides being the largest communities, community 1 and community 2 include the highest number of *Disinfo12 followers* and *GoodInfo followers*, respectively, along with the *Disinfo12* and *GoodInfo* users. In addition, within these two communities,

Table 5: The 10 most populated communities identified in the retweet network (sorted by number of users who retweet the *Disinfo12*)

Nr.	Main Topic (Language)	*Disinfo12*	*GoodInfo*
1	USA news and politics (EN)	48,402	34,942
2	USA news and politics (EN)	7,156	113,490
3	UK news (EN)	3,735	24,994
4	Spain and South America news (ES)	2,782	4,400
5	African News (several)	1,619	5,203
6	Global Discussion (EN)	1,600	7,386
7	India news (HI)	1,151	2,766
8	Japan news (JP)	702	862
9	Indonesia news (IND)	234	961
10	Thailand news (TH)	99	540

we found several users involved in political discussions, accounts of politicians, and media outlets tied with both the Democratic and Republican parties. Interestingly, community 2, which is the community mainly populated by *GoodInfo followers*, includes left-leaning political figures and media outlets.

Community 1 represents the community with the largest body of *Disinfo12 followers* (~48k accounts), but still with an important presence of *GoodInfo followers* (~35k accounts). To have a better understanding of how *Disinfo12* and *GoodInfo* users interact and are grouped together, we further apply the Louvain method to extract sub-communities within community 1. The outcome of the community extraction shows that 72% of the accounts within community 1 are concentrated in only 5 sub-communities. Table 6 reports the distribution of *Disinfo12 followers* and *GoodInfo followers* within these 5 sub-communities. In particular, sub-communities 1 and 2, mainly consist of *Disinfo12 followers* (~25k and ~10k, respectively) while containing only a small fraction of users retweeting only *GoodInfo* content. It is worth-noting that these two sub-communities consist of roughly 50% of users that retweet only the *Disinfo12*. In contrast, sub-community 3 mainly consists of *GoodInfo followers* with only ~4k users that re-share only *Disinfo12* tweets. Finally, sub-communities 4 and 5 show a more balanced distribution of the two sets of users with respect to the other sub-communities. Based on this distribution, we conclude that there exist two macro-groups (sub-communities 1 and 2) that primarily contribute to broadcasting *Disinfo12*'s tweets within the social network, representing therefore the secondary actors responsible for the spreading of misinformation. In contrast, sub-community 3 group together accounts that mainly diffuse accurate information.

To further explore this community and its sub-communities, in Figure 10, we display the interaction in the form of retweets among the nodes in community 1. Edges with weight (i.e., frequency of occurrence) less than 2 and nodes with in-degree less than 10 are hidden to minimize visual clutter. For this reason, sub-communities 4 and 5 cannot be observed in Figure 10, while the most populated sub-communities can be appreciated. In Figure 10, red nodes represent *Disinfo12* and *Disinfo12 followers* while green nodes represent *GoodInfo* and *GoodInfo followers*. The edge takes the same color of the author of the retweet (source node), whereas node size is proportional to the in-degree of the node. The graph is visualized using

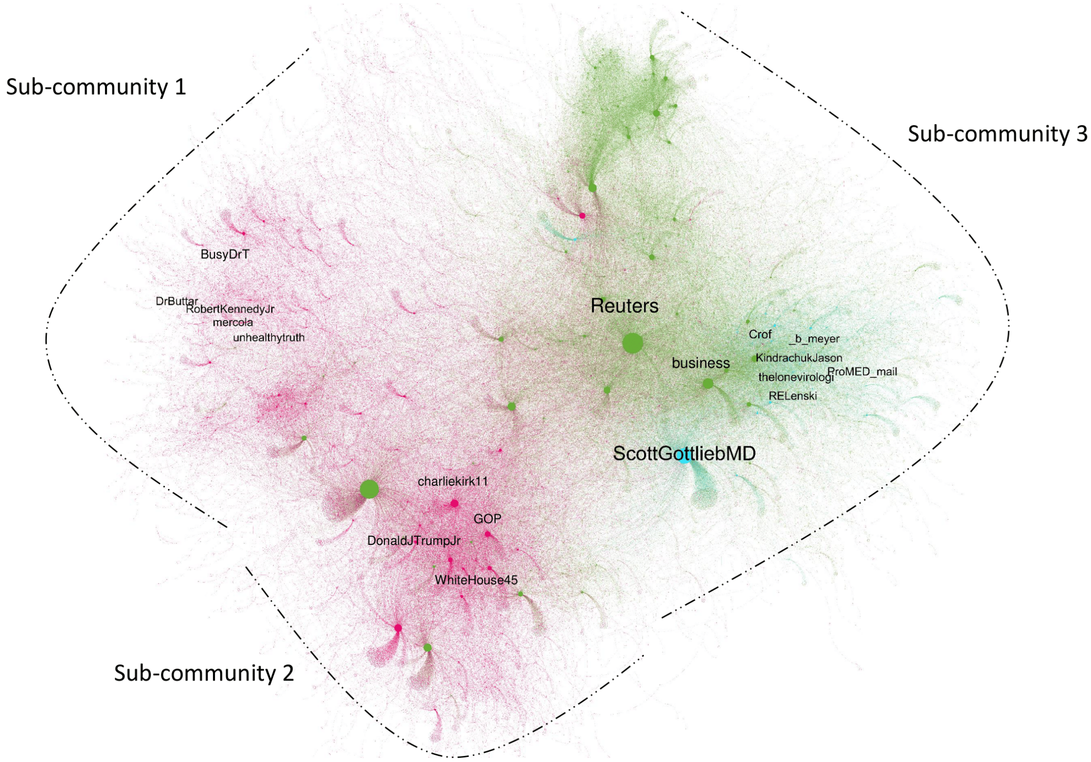

Figure 10: Community 1 and its sub-communities. Red nodes represent *Disinfo12* and *Disinfo12 followers* while green nodes represent *GoodInfo* and *GoodInfo followers*

Table 6: The five most populated sub-communities identified within community 1

Nr.	Description	Disinfo12 Retweeters	GoodInfo Retweeters
1	*Disinfo12*	10601	972
2	Conservative users	25203	3207
3	Liberal users	3796	19001
4	News	1159	2036
5	News	1106	4137

a force-directed layout [15], where nodes repulse each other, while edges attract their nodes. In our setting, this means that users are spatially distributed according to the amount of retweets between each other.

Examining in more detail the three sub-communities, we found that sub-community 2 (almost entirely populated by *Disinfo12 followers*) include accounts relative to right-leaning political figures (e.g., GOP, Rudolph Giuliani, WhiteHouse45), and conservative journalists and TV reporter. Interestingly, we also observe that several accounts within the sub-community 3 (almost entirely populated by *GoodInfo followers*) are left-leaning political elites and media outlets. As for sub-community 1, it includes the *Disinfo12*, their followers, several low-credibility media outlets, and their journalists.

6 Discussion and Conclusion

The unfortunate outbreak of the Covid-19 pandemic was shortly followed by an infodemic, with conspiracy theorists massively diffusing misinformation over social media. In particular, twelve users, referred to as Disinformation Dozen (*Disinfo12*), were identified for being responsible for the spread of two-thirds of Covid-19-related misinformation, which represented an unprecedented case in the spread of false and questionable narratives on social media.

In this work, we performed an analysis on the activity of *Disinfo12* on Twitter, comparing their online behavior to a selected set of users who were known to proliferate accurate and helpful information in response to Covid-19 outbreak (we referred to this set of users as *GoodInfo*). Specifically, we examined the strategies the *Disinfo12* have adopted to proliferate misleading and false information. We observed how *Disinfo12* acted as sources of information by creating original content while interacting much less with other users if compared to *GoodInfo*. By contrasting the most used hashtags by *Disinfo12* to those of *GoodInfo*, we noticed that *Disinfo12* aimed at increasing uncertainty and promoting distrust in health and pharmaceutical organizations by constantly pointing to the involvement of these organizations in a global conspiracy.

Moreover, we analyzed the information sources (e.g., media outlets, websites, other social media, etc.) exploited by *Disinfo12* and *GoodInfo* in their tweets. We noticed how *Disinfo12* heavily rely on proprietary websites and YouTube videos to spread pieces of information and conspiracies. Interestingly, we inspected these videos and found out that nearly 50% of the videos were removed, supposedly for being conspiracy-based. We then inspected the title and description of the remaining 50% of the videos, i.e., those which were not removed by YouTube, and discovered that they mostly have a conspiracy nature.

Finally, we explored the communities floating around the *Disinfo12* and *GoodInfo* to identify potential contributors to the diffusion of misinformation. In particular, we identified two categories of account directly involved in the re-sharing of content generated by the *Disinfo12*. The first category is linked to low-credibility sources, media outlets, and influential users who retweet *Disinfo12*'s content without any apparent political involvement. A second category is composed of accounts politically active and linked to the Republican party, such as right-leaning political figures, journalists, and media outlets. This result is in line with previous findings, which observed how conservative accounts are more susceptible to misinformation [6, 18, 42]. To the contrary, we found that liberal accounts stand in the communities characterized by a vivid endorsement towards *GoodInfo*.

Overall, our work represents a starting point in the understanding and characterization of prominent misinformation influencers as the *Disinfo12*. Our findings pose novel questions and challenges in the fight against information manipulation on social media, which will be further explored in our future works.

References

[1] Edmund Andrews. 2019. How fake news spreads like a real virus. https://engineering.stanford.edu/magazine/article/how-fake-news-spreads-real-virus
[2] Zapan Barua, Sajib Barua, Salma Aktar, Najma Kabir, and Mingze Li. 2020. Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. *Progress in Disaster Science* 8 (07 2020), 100119. https://doi.org/10.1016/j.pdisas.2020.100119
[3] Staci L Benoit and Rachel F. Mauldin. 2021. The "anti-vax" movement: a quantitative report on vaccine beliefs and knowledge across social media. *BMC Public Health* 21 (11 2021). https://doi.org/10.1186/s12889-021-12114-8
[4] Zack Budryk. 2020. Conspiracy theorists who claim 5G linked to coronavirus believed to burn cell towers in Europe. https://thehill.com/policy/technology/493927-arsonists-falsely-linking-5g-to-coronavirus-burn-cell-towers-in-europe
[5] Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health and Surveillance* 6 (05 2020), e19273. https://doi.org/10.2196/19273
[6] Wen Chen, Diogo Pacheco, Kai-Cheng Yang, and Filippo Menczer. 2021. Neutral bots probe political bias on social media. *Nature Communications* 12 (09 2021), 5580. https://doi.org/10.1038/s41467-021-25738-6
[7] Jose Yunam Cuan-Baltazar, Maria José Muñoz-Perez, Carolina Robledo-Vega, Maria Fernanda Pérez-Zepeda, and Elena Soto-Vega. 2020. Misinformation of COVID-19 on the Internet: Infodemiology Study. *JMIR Public Health and Surveillance* 6 (04 2020), e18444. https://doi.org/10.2196/18444
[8] Sarah Evanega, Mark Lynas, Jordan Adams, and Karinne Smolenyak. 2020. Coronavirus misinformation: quantifying sources and themes in the COVID-19 'infodemic'. https://allianceforscience.cornell.edu/wp-content/uploads/2020/10/Evanega-et-al-Coronavirus-misinformation-submitted_07_23_20-1.pdf
[9] Emilio Ferrara, Herbert Chang, Emily Chen, Goran Muric, and Jaimin Patel. 2020. Characterizing social media manipulation in the 2020 U.S. presidential election. *First Monday* 25 (10 2020). https://doi.org/10.5210/fm.v25i11.11431
[10] Emilio Ferrara, Stefano Cresci, and Luca Luceri. 2020. Misinformation, manipulation, and abuse on social media in the era of COVID-19. *Journal of Computational Social Science* 3 (11 2020). https://doi.org/10.1007/s42001-020-00094-5
[11] Center for Countering Digital Hate. 2021. The disinformation dozen: Why platforms must act on twelve leading online anti-vaxxers. https://www.counterhate.com/_files/ugd/f4d9b9_b7cedc0553604720b7137f8663366ee5.pdf
[12] Sheera Frenkel, Davey Alba, and Raymond Zhong. 2020. Surge of Virus Misinformation Stumps Facebook and Twitter. *The New York Times* (03 2020). https://www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html
[13] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on Twitter during the 2016 U.S. presidential election. *Science* 363 (01 2019), 374–378. https://doi.org/10.1126/science.aau2706
[14] Peter Hotez. 2022. Communicating science and protecting scientists in a time of political instability. *Trends in Molecular Medicine* 28 (03 2022), 173–175. https://doi.org/10.1016/j.molmed.2022.01.001
[15] Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE* 9 (06 2014), e98679. https://doi.org/10.1371/journal.pone.0098679
[16] EM Lederer. 2020. UN Chief Antonio Guterres: Misinformation About Covid-19 is the New Enemy. https://time.com/5811939/un-chief-coronavirus-misinformation/
[17] Luca Luceri, Felipe Cardoso, and Silvia Giordano. 2021. Down the bot hole: Actionable insights from a one-year analysis of bot activity on Twitter. *First Monday* (2021).
[18] Luca Luceri, Ashok Deb, Adam Badawy, and Emilio Ferrara. 2019. Red bots do it better: Comparative analysis of social bot partisan behavior. In *Companion proceedings of the 2019 World Wide Web conference*. 1007–1012.
[19] Luca Luceri, Ashok Deb, Silvia Giordano, and Emilio Ferrara. 2019. Evolution of bot and human behavior during elections. *First Monday* (2019).
[20] Benjamin E. Mabrey. 2021. *The Disinformation Dozen and Media Misinformation on Science and Vaccinations*. Ph. D. Dissertation. https://ir.library.oregonstate.edu/concern/honors_college_theses/n009w8842?locale=en
[21] Thomas Macaulay. 2020. Social media firms will use more AI to combat coronavirus misinformation. https://thenextweb.com/neural/2020/03/17/social-media-firms-will-use-more-ai-to-combat-coronavirus-misinformation-even-if-it-makes-more-mistakes/
[22] Raina M. Merchant and Nicole Lurie. 2020. Social Media and Emergency Preparedness in Response to Novel Coronavirus. *JAMA* 323 (03 2020). https://doi.org/10.1001/jama.2020.4469
[23] Panagiotis Metaxas, Eni Mustafaraj, Kily Wong, Laura Zeng, Megan O'Keefe, and Samantha Finn. 2015. What Do Retweets Indicate? Results from User Survey and Meta-Review of Research. *Proceedings of the International AAAI Conference on Web and Social Media* 9 (2015), 658–661. https://ojs.aaai.org/index.php/ICWSM/article/view/14661#:~:text=Our%20findings%20indicate%20that%20retweeting
[24] Sophie Mylan and Charlotte Hardman. 2021. COVID-19, cults, and the anti-vax movement. *The Lancet* 397 (03 2021), 1181. https://doi.org/10.1016/s0140-6736(21)00443-8
[25] Taylor Nelson, Nicole Kagan, Claire Critchlow, Alan Hillard, and Albert Hsu. 2020. The Danger of Misinformation in the COVID-19 Crisis. *Missouri Medicine* 117 (2020), 510–512. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7721433/
[26] Dimitar Nikolov, Alessandro Flammini, and Filippo Menczer. 2021. Right and left, partisanship predicts (asymmetric) vulnerability to misinformation. *Harvard Kennedy School Misinformation Review* (02 2021). https://doi.org/10.37016/mr-2020-55
[27] Jonathan Oppenheim. 2020. Jonathan Oppenheim - Covid-19 trusted tweeters. https://www.ucl.ac.uk/oppenheim/Covid-19_tweeps.shtml
[28] Francesco Pierri, Brea Perry, Matthew R. DeVerna, Kai-Cheng Yang, Alessandro Flammini, Filippo Menczer, and John Bryden. 2021. Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. https://ui.adsabs.harvard.edu/abs/2021arXiv210410635P/abstract
[29] Proceedings of the 28th International Conference on Machine Learning 2011. *Sparse Additive Generative Models of Text*. Proceedings of the 28th International Conference on Machine Learning. https://icml.cc/2011/papers/534_icmlpaper.pdf
[30] Jayaseelan R, Brindha D, and Kades Waran. 2020. Social Media Reigned by Information or Misinformation About COVID-19: A Phenomenological Study. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3596058
[31] Jon Roozenbeek, Claudia R. Schneider, Sarah Dryhurst, John Kerr, Alexandra L. J. Freeman, Gabriel Recchia, Anne Marthe van der Bles, and Sander van der Linden. 2020. Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science* 7 (10 2020), 201199. https://doi.org/10.1098/rsos.201199
[32] Marietje Schaake. 2020. Coronavirus shows Big Tech can fight 'infodemic' of fake news. *Financial Times* (03 2020). https://www.ft.com/content/b2e2010e-6cf8-11ea-89df-41bea055720b
[33] SERRC. 2021. Covid Information Struggles, Brian Martin. https://social-epistemology.com/2021/07/12/covid-information-struggles-brian-martin/
[34] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. *Nature Communications* 9 (11 2018). https://doi.org/10.1038/s41467-018-06930-7
[35] Clea Skopeliti and Bethan John. 2020. Coronavirus: How are the social media platforms responding to the 'infodemic'? https://firstdraftnews.org/articles/how-social-media-platforms-are-responding-to-the-coronavirus-infodemic/
[36] Reuters Staff. 2020. False claim: A COVID-19 vaccine will genetically modify humans. *Reuters* (05 2020). https://www.reuters.com/article/uk-factcheck-covid-19-vaccine-modify-idUSKBN22U2BZ

[37] Massimo Stella, Emilio Ferrara, and Manlio De Domenico. 2018. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences* 115 (11 2018), 12435–12440. https://doi.org/10.1073/pnas.1803470115
[38] Samia Tasnim, Md Mahbub Hossain, and Hoimonty Mazumder. 2020. Impact of Rumors or Misinformation on Coronavirus Disease (COVID-19) in Social Media. *Journal of Preventive Medicine and Public Health* 53 (04 2020), 171–174. https://doi.org/10.3961/jpmph.20.094
[39] Emily K. Vraga, Melissa Tully, and Leticia Bode. 2020. Empowering Users to Respond to Misinformation about Covid-19. *Media and Communication* 8 (06 2020), 475. https://doi.org/10.17645/mac.v8i2.3200
[40] David Wemer. 2020. Addressing the coronavirus "infodemic". https://www.atlanticcouncil.org/blogs/new-atlanticist/addressing-the-coronavirus-infodemic/
[41] Martin Whitehead, Niall Taylor, Alex Gough, Danny Chambers, Mike Jessop, and Phil Hyde. 2019. The anti-vax phenomenon. *Veterinary Record* 184 (06 2019), 744.1–744. https://doi.org/10.1136/vr.l4027
[42] Harry Yaojun Yan, Kai-Cheng Yang, Filippo Menczer, and James Shanahan. 2020. Asymmetrical perceptions of partisan political bots. *New Media & Society* 23 (07 2020), 146144482094274. https://doi.org/10.1177/1461444820942744
[43] Kai-Cheng Yang, Francesco Pierri, Pik-Mai Hui, David Axelrod, Christopher Torres-Lugo, John Bryden, and Filippo Menczer. 2021. The COVID-19 Infodemic: Twitter versus Facebook. *Big Data & Society* 8 (01 2021), 205395172110138. https://doi.org/10.1177/20539517211013861