D4.1 – COVID-19 misinformation tracking

Grant agreement number:	101003606	Due date of Deliverable:	M06
Start date of the project:	1 April 2020	Actual submission date:	30 Sept 2020
Duration:	36 months	Deliverable approved by the WPL/CO : ✓	

Lead Beneficiary:	THE OPEN UNIVERSITY (OU)
Contributing beneficiaries:	TECHNISCHE UNIVERSITEIT DELFT (TUD)
	CENTRUM BADAN KOSMICZNYCH POLSKIEJ AKADEMII NAUK (CBK)

Keywords
misinformation, fact-checking, spread patterns, co-spread

Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

History

Author	Date	Reason for change	Release
Tracie Farrell	August 2020	First Draft	v1
Grégoire Burel Harith Alani	September 2020	Second Draft	v2
Tracie Farrell	September 2020	Third Draft	v3
Gyöngyi Kovács / Hanken	September 2020	Review	v4
All	September 2020	Revision	v5

Executive Summary

In this deliverable, we present the state of the art in **measuring the impact of fact-checking**, highlighting a gap in which our knowledge of misinformation spread patterns is disconnected from how we approach the diffusion of fact-checking information. We show how current approaches that use holistic or aggregate measures may not provide the level of granularity needed to budge persistent claims. We outline some of the important features of the current COVID-19 pandemic, such as the companion “infodemic”, values and culture, that make measuring impact even more difficult. We highlight the necessity for understanding the **“co-spread” of both misinformation and fact-checking information**, to be able to measure the impact of fact-checking on specific misinforming claims temporally and, potentially, at the geographic or platform level. Through our initial analysis of the co-spread of misinformation and fact-checking information during the initial period of the COVID-19 pandemic, we can demonstrate that **fact-checking spread has a positive impact in reducing misinformation about specific claims**. In addition, we are able to provide insight about temporal factors such as the amount of shared misinformation (which is disproportionately higher than fact-checking content), the **different communities of fact-check sharers versus misinformation sharers**, and the short period of time in which fact-checks are likely to spread. In particular, we demonstrate that the **amount of shared misinformation is disproportionately higher for particular misinformation URLS compared to fact-checking content**, and that users that share each type of content do not mix. Finally, we show that the **impact of fact-checking tends to be short-lived** as spread in fact-checking information collapses. To overcome this, we argue that it will be necessary to build interaction bridges between fact-checking and misinformation spreaders, and create fact-checking content that is more appealing. This will help create an **engaging, sustainable fact-checking information spread over time**.

Table of content

Table of figures

List of Acronyms and Definitions

Abbreviation / acronym	Description
IFCN	**International Fact Checking Network**
Co-spread	**The spread patterns of fact-checking and misinformation involving the same claim.**
Fact-checking	**This refers to articles written by signatories to the IFCN checking claims.**
Claim	**This refers to a statement of fact, not opinion or interpretation, that can be fact-checked.**

1 Introduction

We know that **vast amounts of misinformation about COVID-19 proliferate on social media** as a result of both human and technological factors [Brennen et al., 2020; Cinelli et al., 2020; Vaezi & Javanmard, 2020]. Despite the efforts of dozens of fact-checking organisations, working globally to debunk and correct misinformation, misinformation about COVID-19 continues to emerge (and re-emerge) daily. The term **"infodemic"** has been used to describe the current climate in which misinformation about COVID-19 is proliferating faster than we can cope [Cinelli et al., 2020].

In this task, we aimed to (a) automatically collect fact-checks from legitimate fact-checkingers (fact-checkers that are registered and verified by the International Fact-Checkers Network –IFCN40), and (b) produce a visualisation of the spread of specific claims, alongside the corrective information published by the fact-checkers. We refer to this as **"co-spread"**. To achieve this task rapidly, we were able to rework a general misinformation tracking proof-of-concept tool developed in a previous H2020 project, Co-Inform. The goal of task 4.1 is to establish a better understanding of which misinformation is spreading, how rapidly it is spreading, and how effective are the fact-checks in combating this misinformation.

Previous research indicated that misinformation spreads much faster than true information, exploiting emotions and network characteristics to proliferate [Vosoughi et al., 2018]. This understanding has helped to highlight the features of misinformation that are appealing to users. However, **fact-checking is a different type of information from both true and false claims** [Jiang & Wilson, 2018]. As such, fact-checks may exhibit different spread patterns that illuminate new interdependencies and factors that play into the persistence of misinformation on the Web. **Fact-checks are responses to specific claims, often in specific contexts**. Disarticulating the fact-check from the claim it is evaluating may result in loss of granularity around which topics persist over time, for whom. In this deliverable, we explore **the co-spread of misinformation and fact-checks about COVID-19 from social, temporal, topological and typological perspectives**. We explore the impact of fact-checking on key misconceptions or falsehoods arising during the COVID-19 pandemic. We present our methodology for **tracking these claims at scale** and produce some early indications of the **mitigating issues that may help or hinder the spread of corrective information** online.

1.1 Measuring General Impacts of Fact-checking

At the time of writing, over 8.5K COVID-19 disinformation fact-checks have been published by IFCN (International Fact-Checking Network)[1] registered fact-checking organisations. The impact of these efforts can be measured in a number of ways, depending on the domain involved.

From **socio-psychological perspectives**, for example, **qualitative examination or self-report** may provide information about changes in belief or sharing behaviour over time. In addition, modelling techniques allow researchers to draw an abstraction of complex socio-cognitive processes from social and psychological theory. **Modelling user authentication**, for example, can show the internal and external processes that take place during a "primary encounter" with misinformation, which may be triggered by

[1] https://www.poynter.org/ifcn/

misinformation that one encounters unintentionally (through friends and family or on social media) or intentionally, through searching for news or directly asking people one knows [Safieddine & Ibrahim, 2020]. Challenges can arise if the internal authentication process is satisfied (if the claim seems true to the user based on the message, the source, the style or any other credibility indicators). Information consumers may not move on to an external verification process, at which point information literacy activities can be utilised. Our future work in HERoS hopes to illuminate how those authentication processes work in relation to fact-checking articles, in particular with regard to explainability, credibility and perceived bias.

Where you have **journalists and other media personalities** weighing in on the facts of science, lack of scientific experience may lead them to judge scientific quality on the basis of journalistic values (balance, novelty and conflict), rather than scientific rigour and criticality [Dornan, 2020]. Newsroom models of fact-checking may also be looking to appeal to their readership, whereas NGO models of fact-checking may be connected with the desire to promote democratic principles [Cherubini and Graves, 2016]. These types of fact-checkers may look at **how fact-checking improves the overall information environment** to assess impact. To provide one example from the UK context, the UK fact-checking organisation Full Fact (https://fullfact.org) argues that fact-checking should contribute to providing the public with valuable information, hold public figures accountable for what they say and build an "evidence base" of how misinformation emerges and spreads. To track the impact of their efforts, Full Fact also considers "reach" - who will see fact-checks? Research indicates that their audience is more educated, politically involved and more likely to be male, something the organisation is seeking to shift. When looking at how that audience uses their work, Full Fact reports that 41% of their audience uses fact-checking to develop their own position, while 27% claim they use it to prove a point. Government officials report relatively low levels of perceived bias (2-3% from both left and right parties), and even use corrections by Full Fact as a KPI for party officials. Two popular UK daily newspapers, the Daily Mail and the Sun have initiated corrections columns for their publications as a result of Full Facts intervention [Sippitt & Moy, 2020]. In HERoS, we will be examining **demographic trends** as well, attempting to automatically detect and study at scale the individuals helping to share and spread fact-checking efforts in their networks.

Computational approaches, as a compliment to the above, may measure **the level of misinformation in a network** or model the necessary interventions to see what tipping points may assure elimination or limitation of misinformation. For example, scientists have studied the impact of **segregation of networks**, which results from homophily and may encourage the spread of misinformation and make it more difficult for fact-checks to break through [Tambuscio et al., 2018]. Several authors have used **epidemiological modelling** as the basis for exploring misinformation using the susceptible -infected - recovered (SIR) and susceptible - infected - susceptible (SIS) models [Jin et al., 2013; Tambuscio et al., 2015]. Researchers have experimented with the inclusion of other features such as polarisation, "forgetting" information or the presence of debunkers and "immune" individuals [Saxena et al., 2020]. Use of such modelling techniques can illuminate how misinformation spreads through different nodes and jumps from community to community via connections referred to as "weak ties". The role of fact-checking in such models is the "remedy", where the assumption is that the presence of fact-checks can limit either the exposure to misinformation or the effectiveness of corrections on belief. This field is incredibly dynamic. Through such modelling, scientists can also infer the structural characteristics of the network that can encourage the spread of correct information and limit the spread of misinformation [Tambuscio & Ruffo, 2019]. Chains or groups of nodes may accelerate the spread of misinformation [Sarkar et al., 2019] and, as Xian *et al.* [2019] demonstrate, individuals can be exposed to and share misinformation across platforms. In the context of the current crisis, Cinelli *et al.* [2020] analysed spread patterns of different COVID-19 related misinformation across several platforms. The authors noted different diffusion patterns for different types of misinformation on each platform.

In HERoS, we are striving to ensure that these models are more closely aligned with real-life features of information consumption in a network and the socio-technological factors involved. For this we need to consider **mitigating features of health crises** that may impact information consumption or processing.

1.2 COVID-19 Mitigating Factors

As a health-related crisis, COVID-19 has created an environment of uncertainty and lack of control that allows misinformation to thrive. The kinds of misinformation that may appeal to the public can shed light on what information gaps are most critical. Newsguard reported that by April 2020, the most popular misinformation about COVID-19 circulating online involved its origins [Gregory and McDonald, 2020][2]. They have since added a number of COVID-19 myths that emerged at later stages during the progression of the virus across Europe, Asia and the Americas. Brennen et al. [2020] showed similarly that public attention during the COVID-19 pandemic appears to be focused on how this all happened, (virus origins and conspiracy theories), the risks (including how bad the situation is currently and how bad we expect it to be), and how we can fix the situation (in terms of preventing transmission, testing or vaccines). The public clearly also has information needs around **how government or prominent public figures may be involved** or **how the rest of the public may react**. Governments and authority figures may feature prominently because the public already knows that there will be some information they will not get. Experts in managing public health crises have admitted that communicating with the public during times of crises uncovers ethical issues around how much information to provide and in which tone, communicating information that may stigmatise or violate the privacy of an individual or group, and inciting panic [Timothy Coombs & Jean Holladay, 2014]. By looking at how misinformation related to these topics is spreading, we can understand when the public needs information and why, depending on a number of factors related to their information environment. We argue that the temporal patterns can illuminate which misinformation is persistent, despite the availability of evidence that refutes it.

Health topics, in general, are often accompanied by misinformation [Vaezi & Javanmard, 2020; Xie et al., 2020], but is all misinformation equal to all audiences? Spence *et al.* [2007] examined the information seeking activities of individuals during the September 11th attacks on the World Trade Center in New York City (Spence et al., 2005) and during Hurricane Katrina. The authors demonstrated that, under conditions of uncertainty with threat of danger, **people will seek information continuously**, updating it often. They noted some **differences related to potentially vulnerable communities** (such as those with disabilities, people of colour and, more generally, women) in information seeking early on at the crisis preparation stage, with those who are disabled having more personally relevant informational needs (in comparison to needing information about the scope of the crisis or its impact on others). Information seeking across different communities may also be impacted by the **terminology used**. In an early study on COVID-19 misinformation on Twitter, Kouzy *et al.* [2020] collected and examined tweets (n= 673) containing certain

[2] https://www.newsguardtech.com/covid-19-myths/

relevant hashtags (such as #corona, #nCov and #COVID-19). The authors found that certain hashtags were more correlated with misinformation (#corona, #nCov).

In their analysis of misinformation on Twitter during the Ebola health crisis, Jin *et al.* [2014] found that rumours were more localized topologically, with certain rumours spreading more prolifically in certain geographic areas. The authors argue that understanding the **connection of the context to the misinforming claim** will aid in providing corrective information to that specific audience. In their study of communication to the public during the 2014 West African Ebola outbreak, Allgaier and Svalastog [2015] found similarly that traditions or cultural norms may interfere with recommended advice and interventions, influencing the spread of misinformation. In addition, the **characteristics of the media** in a given area of the world can influence how misinformation connects with existing narratives. Harman [2020] argued that journalism can detract from and misrepresent real facts, especially in global health crises, contributing to the spread of misinformation.

One significant hurdle has been growing **scientific skepticism**, which makes it difficult to get the public to accept expert advice and guidance. In his study on science related misinformation, Dornan [2020] remarked that it is necessary to dig deeper into the reasons why this skepticism and rejection of expertise may be aligning with the political right. The rejection of mainstream media as purveyors of facts that are true, regardless of one's political leaning, feeds into this. Dornan relates this to how we communicate about science, focusing on the fact that science can illuminate truth, which it sometimes does, but that this is not its strength. Rather, **the important quality of science is that it can be revised with better information**. The process of contestation is where the real power of the scientific method can be realised. He also argues that **simply removing content may make it seem more appealing** to certain viewers, especially if it touches upon the exact fear that certain groups may have of deep-state conspiracies to prevent citizens from discovering the truth. This gives pause to those of us who are attempting to categorise stories according to the level of truth that they have been determined to contain. State-of-the-art approaches that seek to eliminate misinformation in the network as a success measure, may be missing the mark and **transferring the problem to another more isolated network**, as we have seen with "free speech" platforms like Parler and Gab. Moreover, research indicates that missteps related to the **inaccurate labeling of stories as misinformation may erode trust** in fact-checking efforts and encourage mistrust in mainstream media [Freeze et al., 2020].

Studying how misinformation about COVID-19 spreads facilitates our understanding of the public's information needs during a health crisis. It also helps to identify patterns in the people, timing and topics that are significant in the spread of misinformation. However, these **misinforming claims tell only part of the story**. Looking at the **co-spread of corrective information** is also critical for stemming the flow of misinformation. The presence of fact-checks on a network may be indicative of **what information is most important to clarify**. It may also tell us whether or not the presence of this information has influenced the spread patterns of misinforming claims, helping us to better understand the impact of fact-checking.

1.3 Fact-checking Spread Patterns - What we know now

A fact-check is **assessing claims for accuracy**. The assessment can be that the claim is true, false, or somewhere in between [Vlachos & Riedel, 2014]. Therefore, it represents a third category of information that is addressing **information already in the network** [Jiang & Wilson, 2018]. Fact-checks will reach some individuals who have not been exposed to the original misinforming claim, as well as those who have, and who may have strong internal or external motivations to believe it. As mentioned above, fact-checking is also subject to skepticism that may be influenced by perceptions of the media or an individual's priors. Those interdependencies make general ways of measuring fact-checking impact difficult for prescribing interventions without knowing **which claims persist and why**. In this section, we review what is already known about the temporal, topological and typological patterns that appear to impact the spread of fact-checks online, relating these patterns back to social and psychological factors that may influence the spread of fact-checks in online networks.

Temporal patterns indicate that fact-checks and **corrections need to be supplied early and often** [Aird et al., 2018; Starbird et al., 2018]. This is most likely because misinformation proliferates during times of conflict and war [Lewandowsky et al., 2013], political events [Kuklinski et al., 2000], breaking news developments, disasters [Starbird et al., 2018] and global health crises [Harman, 2020], in which humans need information and fear is elevated. Unfortunately, the **biggest impacts appear to happen shortly after initial circulation** [Starbird et al., 2018], which means that mitigation measures need to be **preventative or very efficient**. As we will see below, topology and typology may influence the temporal patterns observed.

Topological patterns may be visible with regards to the **demographics that are interested in certain topics and able to communicate about them.** Amazeen et al. found that the "**need for orientation**", a combination of interest and uncertainty, was the greatest predictor of sharing fact-checks on Facebook or Twitter, and that fact-checking can be associated with **age, ideology and political participation** [Amazeen et al., 2019]. Several papers implicate conservatism as being linked with skepticism toward fact-checks [Kahne & Bowyer, 2017; Shin & Thorson, 2017, Robertson et al., 2020]. However, this may be somewhat correlational rather than causational, based on media perceptions or polarisation in the political field in-country. For example, Lyons et al. [2020] found that views on fact-checking are more polarising in the US in comparison with Northern European counterparts, and more similar to Spain and Italy. Rather than a left-right dynamic, Lyons et al. found that "anti-elite" sentiment was a better predictor of resistance to fact-checking. Such sentiments can be aggravated by perceptions around transparency and trust in the media [Humprecht, 2020]. Amazeen et al. [2019] argued that fact-checks are most appreciated when they **reinforce existing beliefs**, something that they feel is a key aspect of understanding users' potential motivations to share them. In a US political context, Wintersieck [2017] found similarly that fact-checks supporting a political candidate's statement may impact voter behaviour more than fact-checks that show a candidate was incorrect or dishonest [Wintersieck, 2017]. From the Japanese context, however, Fujishiro et al. [2020] demonstrated that left-right polarisation does not dominate fact-checking dynamics in Japan as much as an anti-xenophobia dynamic, in which liberal Japanese fact-check xenophobic comments about China and Korea. The authors argue that failing to understand the real issues behind xenophobia make **fact-checking ineffective for the purposes of correcting misconceptions that are deeply rooted in racism**. These differences will be important to model as part of seeking patterns in fact-checking and fact-checking acceptance. In their qualitative examination of **COVID-19 misinformation topics**, for example, Brennen et al. [2020] identified misinformation about public authority action, community spread, general medical information, prominent actors, conspiracy theories, transmission fears, virus origins, public preparedness and vaccine development. The authors found that **public authorities, including government and health agencies, are often the target of misinformation** and that misinformation about them was often quite simply false, rather than manipulated in sophisticated ways.

Typology is important to consider, as certain types of misinformation may be more or less resistant to fact-checking. Conspiracy theories, for example, may be difficult to fact-check because of the nature of their **foundational epistemology**. The very nature of conspiracies is that they are secret, which means that the absence of evidence may be seen as evidence. Studies have proposed that education and information literacy may play a role in conspiracy belief [Craft et al., 2017]. Hoaxes and other **information that is demonstrably false** may be easier to correct. Tambuscio *et al.* [2015] used agent-based simulations to develop a dual-epidemiological model for assessing the impact of fact-checkers on viral hoaxes. Their work defines the “minimal reaction” of users (the number of users posting fact-checks) that would be necessary to get rid of a viral hoax. However, this model was not transferred to a real dataset. Later work by Kim *et al.* [2018] used real-world datasets from Twitter and Weibo to model how the network could be mobilised to spread corrective information effectively. Still, these models are meant to **predict how future fact-checks may diffuse and cannot verify causational relationships**. Additionally, Oeldorf-Hirsch *et al.* [2020] found that fact-checking labels were not necessarily effective in credibility assessments of news posts, in particular with regard to memes. The **type of correction** may also play a role in the spread of fact-checks. Research indicates that **short format refutations** may be more effective than a simple retraction [Ecker et al., 2020], possibly because this provides additional space to explain how an evaluation of accuracy was made, without overwhelming the user.

Other **social patterns** have been observed in the proliferation of fact-checks on online networks. Research on misinformation spread across online networks shows that **emotional proximity** among users can aid in misinformation spread, particularly in a crisis [Huang et al., 2015]. Indications are that fact-checking may function similarly. Hannak *et al.* found that users were more likely to engage with fact-checks posted by people they know, rather than strangers [Hannak et al., 2014]. Some researchers have also studied the impact of fact-checking as having a **chilling effect on information sharing in general**. Research on the **“nudging effect”** of fact-checks to shape user behaviour indicated that fact-checks may discourage users from sharing news stories in general, particularly from non-mainstream sources [Margolin et al., 2018; Nekmat, 2020].

The review of existing work shows a **gap in understanding the relation and interaction in the co-spread of corrective information propagated by fact-checking and misinformation**. With a more general understanding of how fact-checks spread, and knowledge of the potential impact of temporal, topological and typological factors, we feel it is important and even necessary to examine some of these phenomena at scale and over time, to be able to devise effective, **sustainable treatment** for the COVID-19 infodemic.

2 Co-Spread of Fact-Checking and Misinformation

To help us explore the interaction between fact-checking and misinformation spread, **we look at the spread of fact-checking articles and their associated misinforming claims side by side**. We compare the diffusion of 2,830 misinformation and 734 fact-checking URLs about COVID-19 on Twitter from early December 2019 to early May 2020 to capture this interaction over time. We perform the co-spread analysis at **two different granularity levels** during the pandemic to better understand if spread patterns vary depending on the referential period studied. First, we analyse how spread differs during the pandemic by observing changes between the initial pandemic onset, the ramping up phase and late pandemic period (COVID-19 level). Second, we study the relative misinformation and fact-checking diffusion patterns by aligning individual URL spreads and analysing how individual misinformation spread after their initial appearance (relative level).

We address the following research questions:

1. ***Are misinformation and fact-checking information shared similarly?***
2. ***How do misinformation and fact-checking co-spread patterns vary at the pandemic level and relative level?***
3. ***How does fact-checking spread affect the diffusion of misinformation about COVID-19?***

We conduct an analysis on the co-spread of fact-checking information and misinformation on Twitter based on the sharing of misinforming URLs that were collected from claim reviews collected from fact-checking websites.

In our approach, first, we collect Twitter data by looking for the appearance of misinforming URLs that we have collected. Second, misinformation and fact-checks spread is aggregated for three different time periods at two different granularity: 1) From the COVID-19 worldwide spread perspective (COVID-19 level analysis), and; 2) From the initial emergence of a misinforming URL (relative level analysis). This allows for a better understanding of spread at different levels. Third, we perform multiple analyses to investigate how fact-checks and misinformation spread behaviour differs. This analysis allows the identification of significant relations between misinformation spread and fact-checking information, which can be used for designing better methods for spreading fact-checking information on social media. Finally, weak causation and impulse response analyses are performed between fact-checks and misinformation in order to identify if fact-checking information diffusion impacts misinformation spread.

3 Data

For our analysis, we need to create a dataset that contains both misinformation and fact-checking information. We focus our work on Twitter due to its popularity and its accessibility. We rely on COVID19-related reports from fact-checking websites that identify misinforming content by their URLs, and search the occurrences of these URLs in user posts on Twitter. This approach has two advantages:

first, we can be assured that experts have assessed the accuracy of claims and second, we can look specifically at misinformation and fact-checks that link back to information we know is proliferating online and is not simply the opinion of single individuals.

3.1 Fact-Check Dataset

The dataset of fact-checks comes from the misinfo.me tool that collects URLs that have been fact-checked, labelled and provided with a fact-checker review. The reviews are published by multiple fact-checking websites belonging to the International Fact-Checking Network (IFCN) using the standard ClaimReview schema [IFCN, 2020; Schema.org, 2020], which was defined appositely for the purpose of annotating reviews of claims. We use IFCN signatories because they follow certain principles in their fact-checking efforts that we feel best represent good practice in the field. The data collection is primarily based on the Data Commons ClaimReview public feed [DataCommons.org, 2020]. From this public feed, ratings are extracted and normalised between -1 and +1, depending on their credibility [Mensio & Alani, 2019]. Using these ratings, we only select misinforming URLs (ratings <=0). We also keep the URLs of the original fact-checking articles and then filter all the URLs to only get the COVID-19 fact-checks by using a set of relevant COVID-19 keywords [Twitter, 2020], based on the title and content of the fact-checks. The final URL dataset includes fact-checks published until the 4th of May 2020, with a total of 2,830 distinct misinforming URLs and 734 fact-checking URLs.

3.2 Twitter Dataset

Using the misinformation and fact-checking URLs, we create the Twitter dataset by searching their occurrences on Twitter in all languages by adapting an existing Twitter Hashtag crawler that collects posts using Twitter's mobile interface [Upreti, 2020]. Out of all the seed URLs, we find posts for only 1,190 distinct URLs for a total of 21,394 posts from 16,308 different users. On average, there are 17.54 posts for each URL (sigma = 28.35, min = 1, max = 232).

Figure 1 shows the cumulative spread of misinforming and fact-checking information URLs shared over time in our dataset. The figure also shows the number of COVID-19 casualties and cases over the same period as well as the COVID-19 *initial, early and late* periods (vertical dashed lines).

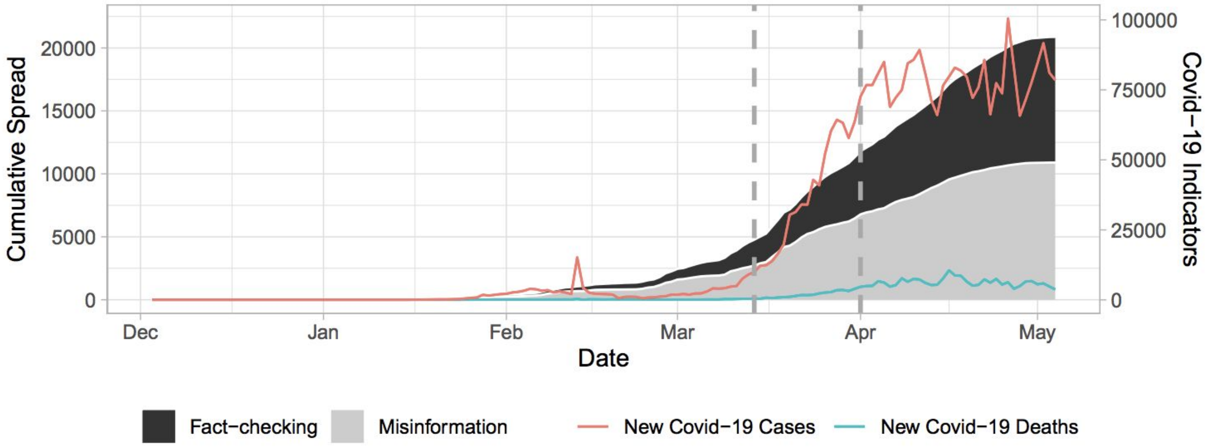


Figure 1- Stacked cumulative spread of misinforming and corrective information URLs over time and global number of COVID-19 casualties and cases over the same time period.

3.3 COVID-19 Cases Dataset

To generate the different time periods at the COVID-19 pandemic granularity, we use the data produced by the European Center for Disease Prevention and Control (ECDC) [2020]. The ECDC collects daily statistics about the number of COVID-19 cases and casualties worldwide for multiple countries. Although the data is continuously updated, for this deliverable, we focus on the 1st Dec. 2019 to 4th May 2020 period since it matches the data we collected on Twitter so far. The beginning date is selected as the 1st Dec. 2019 since this date tends to be associated with the first traceable case of the pandemic [Wang et al., 2020].

3.4 Analysed Periods Generation

We analyse the behaviour of fact-checking information and misinformation at two different granularity levels. First, at the pandemic level (COVID-19 level analysis), we are interested in understanding if spread behaviour varies during three different time periods within the pandemic based on the amount of worldwide cases. Second, at the URL level (relative analysis), we are interested in understanding how behaviour differs based on the number of days since the first occurrence of a misinformation-related URL (i.e., a particular misinforming content or its associated fact-checking information). This would show how misinformation and fact-checks spread over time independently from when they were posted.

3.5 COVID-19 Periods

We generate three *initial, early and late* time periods to analyse fact-checking information and misinformation spreads at the level of the COVID-19 pandemic. We fit multiple linear regression models for the daily worldwide COVID-19 cases curve in order to identify inflection points in the amount of COVID-19 cases [Muggeo, 2003].

Looking for two inflection points in the curve, the *initial* time period is specified as any tweet posted before Saturday, Mar 14, 2020. The *early* period corresponds to any tweet between Saturday, Mar 14, 2020 and Thursday, Apr 2, 2020. The *late* period is for any posts after Thursday, Apr 2, 2020.

3.6 Relative Periods

To understand sharing behaviour independently from when each URL has been initially shared, we align the initial sharing of each URL so that all the URLs shared in the dataset always start from the same initial time (i.e., we normalise the dates for each analysed URL). We identify the first occurrence of each URL and then obtain the number of times it has been shared per day for each day following its initial appearance.

Following the same approach outlined in the previous section, we use the daily aggregated curve containing all the shared URLs (i.e., misinforming and fact-checking URLs) for identifying the *initial, early* and *late* relative time periods by obtaining the inflection point in the daily shared URLs. While the COVID-19 periods are obtained from the unaligned misinforming and fact-checking URLs shares and COVID-19 cases amounts (Figure 1), the aligned amount of misinformation and fact-checks shares is used for dividing the relative periods. Using this method, the *initial* time period is specified as any URL shares happening within the first 2 days after its first occurrence. The *early* period corresponds to shares between day 2 and day 14. Finally the *late* period is for any shares happening after 14 days.

4 Analysis

Multivariate Spread Variance Analysis

The first part of the analysis is to identify the different patterns of appearance of misinformation and fact-check URLs over varying periods of time. In order to perform such analysis, we use the one way Multivariate ANalysis Of VAriance (MANOVA) and the one way ANalysis Of VAriance (ANOVA) methods. This approach allows us to determine if there are significant differences in information spread between the fact-checking information and misinformation groups in each initial, early and late periods.

We focus on the MANOVA and ANOVA methods since they allow us to determine if there are significant differences in information spread between the corrective information and misinformation groups in the initial, early and late spread periods we have defined in the previous section.

Experimental Setup

MANOVA and ANOVA rely on the definition of independent variables and dependent variables. For our analysis, the amount of information spread associated with each information type is our dependent variable whereas each information type (i.e., misinformation and corrective information) is an independent variable.

Since our data does not follow all the assumptions required for the standard ANOVA and MANOVA methods (i.e., multicollinearity, normality and homogeneity), we use non-parametric versions of MANOVA and ANOVA for the analysis, using F-approximations permutation tests. The F-approximation of ANOVA's test, as well as Wilks' Lambda Type Statistic are obtained with their p-value and the associated permutation test p-value.

Our analysis is divided into two different parts for the COVID-19 and relative level analyses: 1) A Non-parametric MANOVA analysis is performed for identifying if there are differences in spread between the different periods and information types, then; 2) Non-parametric ANOVA analysis is then performed if the MANOVA results are significant for each individual time period for determining in which sub-period (i.e., *initial*, *early* and *late*) the pattern differs.

For the non-parametric ANOVA analysis, the Kruskal-Wallis test is used and the p values are adjusted using Bonferroni correction (since multiple dependent variables are analysed). Significant results mean that the behaviour of corrective information and misinformation are significantly different whereas a non-significant result means that the distribution of spread for each time period is non-significant.

4.1 Results of Co-Spread Analysis

In the following section we report the results for each analysed period. We only report significant results for brevity.

4.1.1 COVID-19 Period Analysis

The one way MANOVA analysis comparison at the COVID-19 level URL shares for misinforming URLs and fact-checking URLs shows a significant permuted p-value of 0.01. This means that at the COVID-19 pandemic level, there are significant differences in how misinforming URLs and fact-checking URLs spread and that the type of shared URLs has an effect on the number of shared URLs during the pandemic. Following the significant result of the MANOVA analysis, a one way ANOVA analysis is performed for each COVID-19 time period. The Bonferroni adjusted Kruskall-Wallis tests are only significant for the *initial* ($p = 0.00558$) and *late* ($p = 0.0234$) periods. This result means that sharing behaviour does not differ fundamentally during the *early* COVID-19 period ($p = 1$) whereas sharing behaviour differs in the *initial* and *late* periods.

Looking at the individual URLs shares for each time periods, we observe higher deviations in sharing behaviour for misinformation ($\sigma \in \{20.5, 24.9, 26.3\}$) compared to fact-checking information ($\sigma \in \{7.52, 6.94, 11.3\}$). It also appears that fact-checked information is shared less often than the corresponding misinforming URLs in terms of means with lower means for all the time periods ($2.42 < 5.88$, $3.60 < 8.73$ and $6.34 < 10$). This suggests that perhaps the types of users that share misinformation is more varied than the types of users that share fact-checks. Similarly, there may be a variation in what misinforming topic attracts the most shares compared to the fact-checking content.

4.1.2. Relative Period Analysis

The one way MANOVA analysis comparison at the relative URL shares level for misinforming URLs and fact-checking URLs shows a significant permuted p-value of 0. This means that at the relative URL level, there are significant differences in how misinforming URLs and fact-checking URLs spread and that the type of shared URLs has an effect on the amount of spread at different relative time periods.

Following the significant result of the MANOVA analysis, a one way ANOVA analysis is performed for each relative time period. The Bonferroni adjusted Kruskall-Wallis tests are only significant for the *early* ($p = 4.74 \times 10^{-4}$) and *late* ($p = 1.338 \times 10^{-5}$) periods. This means that sharing behaviour during the *initial* ($p = 0.552$) relative period does not differ during that period whereas differences exist when looking at the *early* and *late* periods.

The individual distribution of misinforming and fact-checking URLs for each time period show that the amount of shares tends to be similar across the URL types with a slightly higher spread for the misinforming URLs in general. Interestingly, the highest difference in term of mean and standard deviation between the different URL types appears to be mostly during the initial phase with a more important standard deviation for the misinforming URLs (sigma = 12.6 for misinforming content and sigma = 3.31 for fact-checks). Compared to the COVID-19 levels analysis, this shows that the difference in spread appears to be highly related to the initial amount of shares of a given URL and to external contextual factors rather than simply the intrinsic properties of the shared URLs.

4.2 Fact-checking Misinformation Impact Analysis

In the previous section we have compared how fact-checking URLs and misinforming URLs spread and if each information type spreads in a similar fashion at different levels.

In this section we investigate how the two types of information (fact-checking URLs and misinforming URLs) impact each other. In particular, we are interested in understanding if the spread of fact-checking information has a beneficial impact in reducing the diffusion of misinformation. For this analysis, we focus on modelling the spread of URLs as a Vector AutoRegression (VAR) model using the misinformation and fact-checking URLs as endogenous variables. We perform this analysis at the relative level (i.e., the relative number of days since the first appearance of a URL related to a particular misinformation) and determine if weak causation relations between each information type exists.

4.2.1 Experimental Setup

Although it is not simple to identify causation relations between each information type, it is possible to estimate if the spread of a given information type can be used to predict the spread of another information type using a Granger causality test. In order to compute the Granger causality test we first build a Vector AutoRegression (VAR) model using the combined misinformation spread and fact-checking information for the analysed period. However, since our data is non-stationary, we first integrate each analysed information type so that the spread amount for each day is represented as the difference between the current day value and the previous day value.

A 14 days order value is used for the VAR model based on Akaike's information criterion. Using the VAR(14) model, we perform a bootstrapped Granger causality test for determining if misinformation spread can be associated with fact-checking URL spread or if fact-checking spread can be inferred from misinformation spread.

In order to better understand the dynamics that relate fact-checking information and misinformation, impulse response analysis is performed as well as Forecast Error Variance Decomposition (FEVD). For the impulse response analysis, we use orthogonal impulse responses in order to evaluate the spread response of the different types of URLs for 14 days steps. This approach allowed us to determine how a particular sharing behaviour may affect other types of URLs shares in future. We are particularly interested in determining **if an increase in fact-checking information shares trigger a reduction in misinformation diffusion**. We run the FEVD with the same 14 days periods in order to obtain the contribution importance of each information type on both misinforming URLs and fact-checking URLs spread.

4.2.2 Results of Impact Analysis

Using the VAR(14) model, we are able to observe a Granger causality relation showing that fact-checking spread has predictive causality over misinformation spread (p = 0.02). This observation is not found in the opposite direction (p = 0.93). This result suggests that at the relative-level, change in fact-checking information spread may cause a change in misinformation spread and therefore **fact-checking articles have an impact on misinformation spread**. Surprisingly, the opposite result shows that **fact-checking spread may not be influenced by misinformation spread**.

The impulse response for the orthogonal shock in the amount of shared fact-checking URLs (Figure 2) shows an initial drop in misinformation shares (first day) but mixed results afterwards. Despite this observation, a general downward misinformation spread trend can be observed. This suggests that **fact-checking tends to have a short significant impact on the spread of misinformation**. A shock in misinformation leads to a sharp drop in misinformation spread. This confirms our previous observation that misinformation spreads tend to occur mostly after its initial spread and decrease quickly in the following days.

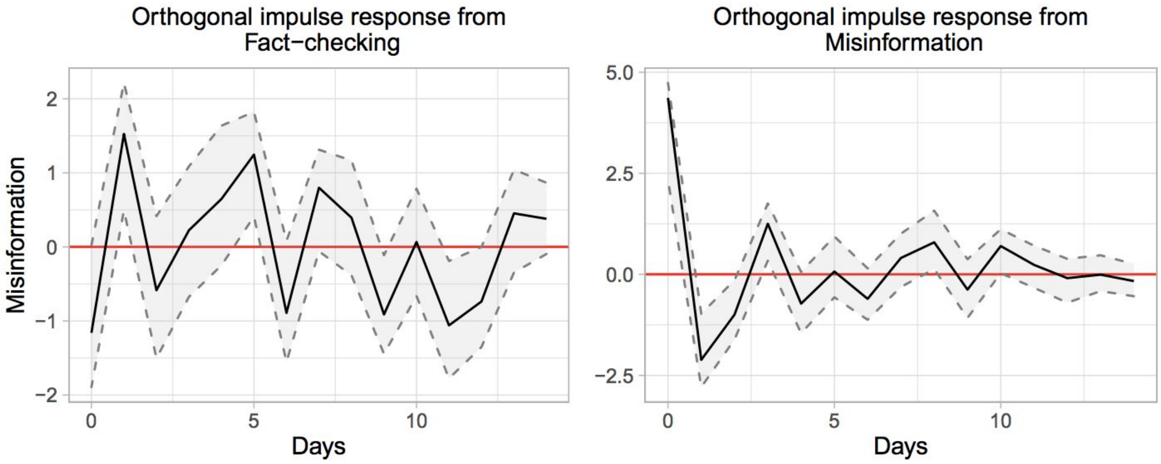


Figure 2- Bootstrapped relative-level orthogonal impulse response from fact-checking shock (95% confidence interval).

The impulse response for the orthogonal shock in the amount of shared misinforming URLs (Figure 3) shows a delayed fact-checking increase two days after the initial misinformation spread. This result suggests that fact-checking spread tends to follow misinformation spread despite a lack of causal relation (i.e., fact-checking articles are created as a response to misinformation). As with the misinformation sharing behaviour, we observe a **sharp decrease in fact-checking sharing behaviour after the initial shock** as initial sharing behaviour reduces.

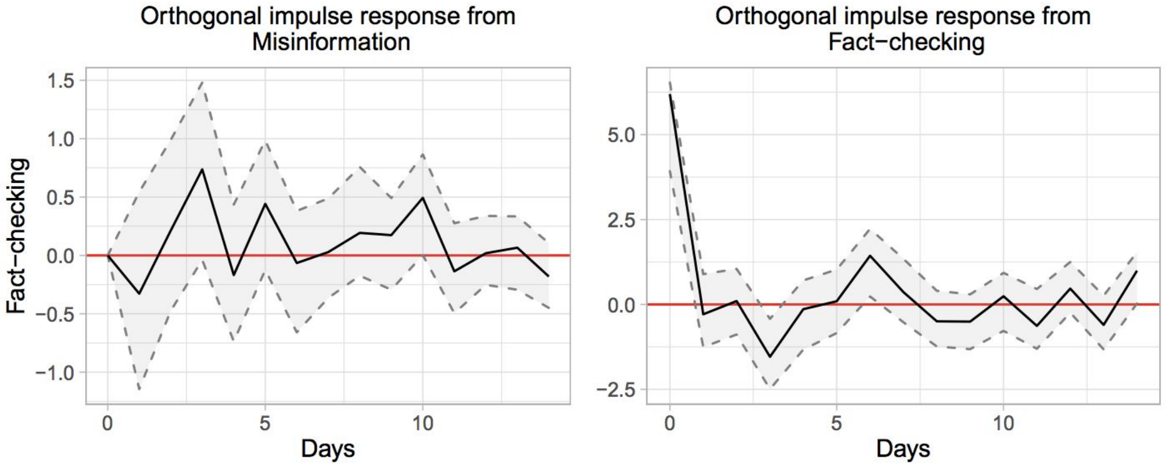


Figure 3- Bootstrapped relative-level orthogonal impulse response from misinformation shock (95% confidence interval).

The FEVD results displayed in Figure 4 show that misinformation spread predictions are directly affected by the spread of fact-checking information with **misinformation prediction getting more affected by fact-checking spread as time goes by** whereas fact-checking spread appears to be unaffected by past misinformation spread. This result adds to our previous causality observation between fact-checking information and misinformation spread.

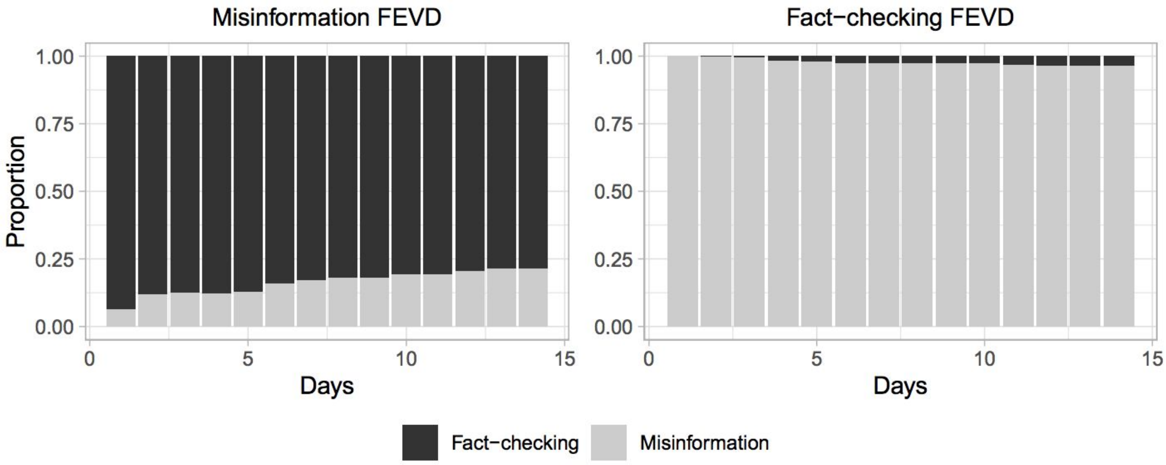


Figure 4- Forecast Error Variance Decomposition (FEVD) for the relative-level misinformation and fact-checking spread.

4.3 Misinformation Sharing Behaviour

To better understand the behaviour of users on Twitter with regards to sharing misinformation or fact-checks, and to understand community overlap in spreading both misinformation and fact-checks, we ran a preliminary analysis on the Twitter timelines of a subset of users. Aim of this analysis was to observe **if, and how often, users share misinformation, fact-checks, and the combination of the two**. For each user included in this study, we retrieved their individual timelines using the standard user-timeline API provided by Twitter, which allows us to retrieve a maximum of 3200 tweets from the timeline of individual users. Next, we compared the misinformation and fact-check URLs in our database against the URLs appearing in the individual tweets collected for all the users. This will tell us how many times a particular user posted any of these URLs.

To select the users for this study, we narrowed down the dataset used in our earlier analysis, and filtered in users who shared at least 2 unique URLs in their timeline that match the URLs in our dataset. This resulted in a list of 504 users. Each URL belongs either to a claim reviewed by a fact-checker and given a score in the range [-1; +1]; or they belong to a fact-check article.

For each user, as shown in Figure 5, we count how many *Negative Information* were shared (i.e., URLs that were reviewed by a fact-checker, and the result was negative) and how many *Positive Information* were shared (i.e., URLs that were reviewed by a fact-checker and the result was positive, in addition to URLs published by a fact-checker).

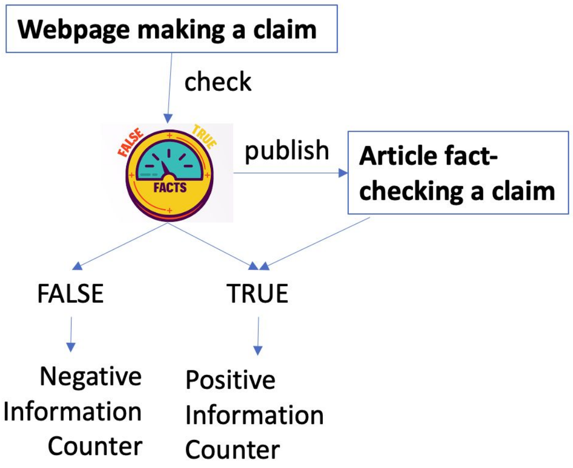

Webpage making a claim
check
FALSE
TRUE
FACTS
publish
Article fact-checking a claim
FALSE
TRUE
Negative Information Counter
Positive Information Counter

Figure 5- method for calculating Negative and Positive Information Counts for a given Twitter user.

Figure 6 shows the plot of these two counts for each user in the dataset. X-axis represents the Negative Information counts and y-axis represents the Positive Information counts. the size of the dot indicates the number of users that have the given numbers. Hence the larger the dot is, the more users with the same negative and positive information counts.

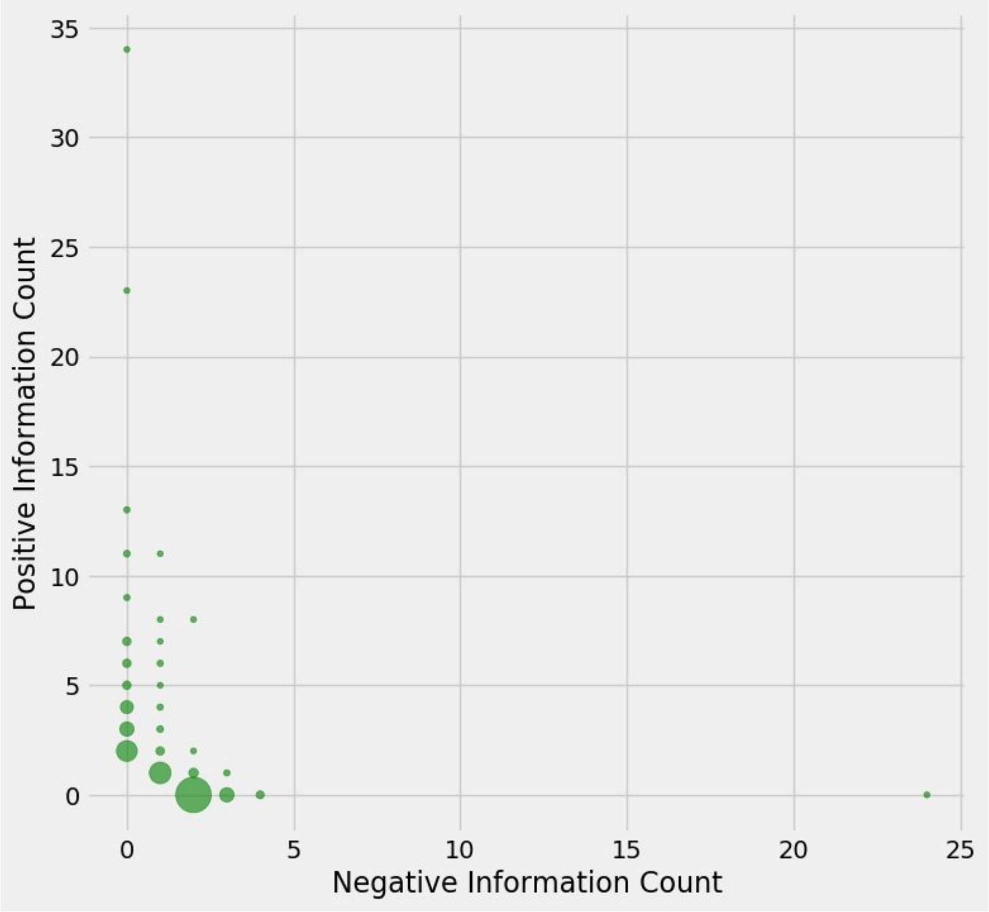


Figure 6- Analysing the positive and negative information sharing behaviour by individual users.

What the plot suggests, given that most users fall either on the y-axis or the x-axis, is that the majority of users are either sharing misinformation, or factual information, with **only a handful of users sharing both types of information**. In other words, what this indicates is that those who share misinformation are much less likely to share correct or fact-check information, and vice versa.

Although this is a preliminary study, it demonstrates the need for researching such behaviours, and discovering ways to **encourage sharing of corrective and fact-check information more**. Further analysis on user behaviour will be carried out later in the project. In particular, we are looking to differentiate co-spread on the basis of different topics, or potentially types of misinformation, to observe further useful patterns.

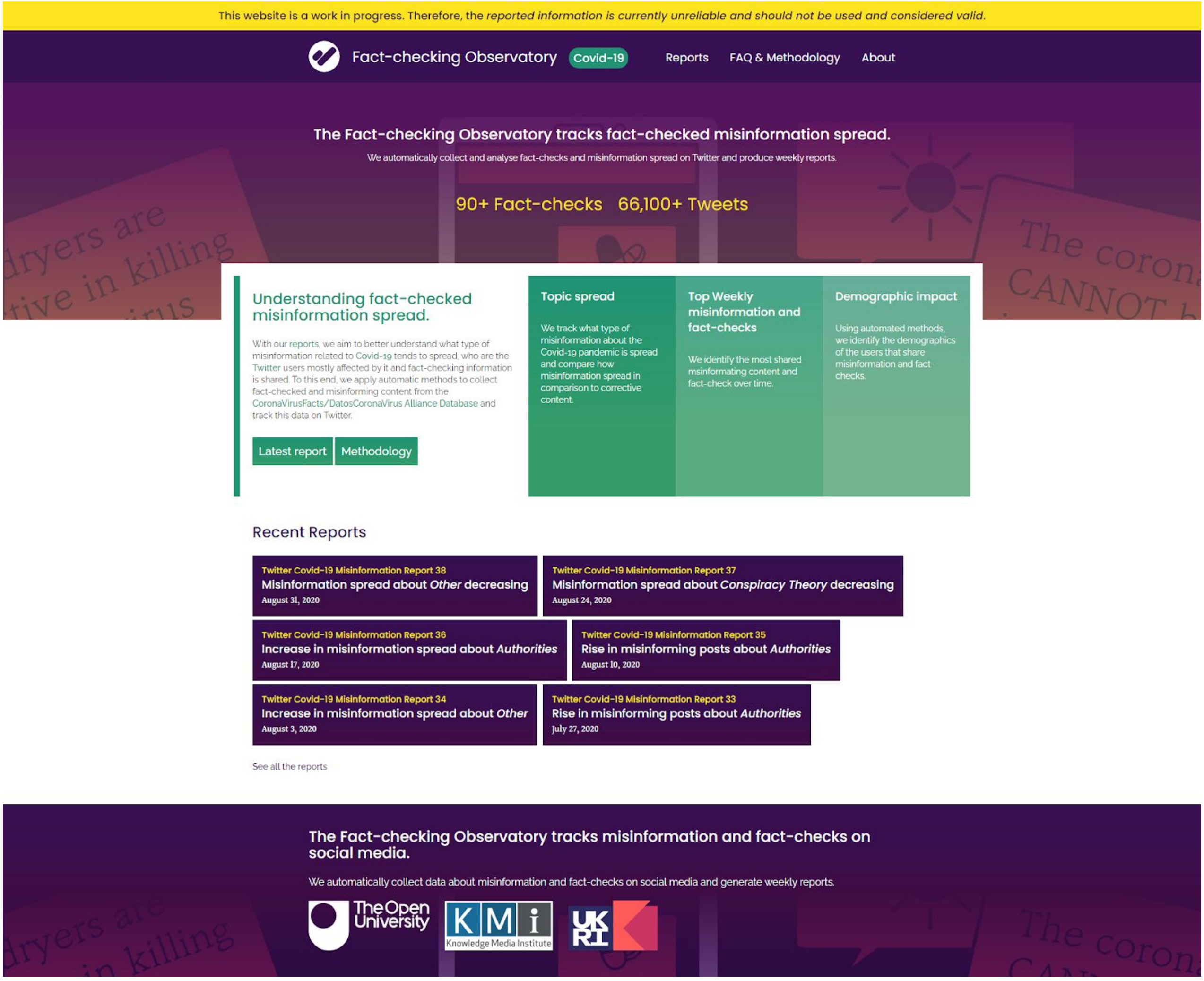
4.4 Continuous Misinformation Tracking

For the previous analyses, data collection was created using a static list of misinforming and fact-checking URLs and a static crawl. In order to scale our analysis, it is necessary to be able to update the collected data continuously. To accomplish this, we designed and developed a tool, with additional support from the UK Higher Education Innovation Fund (HEIF)[3], in the form of a proof-of-concept website **(FC observatory) for automatically generating reports about the co-spread of misinformation and**

[3] https://re.ukri.org/knowledge-exchange/the-higher-education-innovation-fund-heif/

fact-checks on Twitter. The website automatically produces human readable reports every week about the topical and demographic spread of misinformation and fact-checks on Twitter.

The **new Twitter crawler** developed as part of the project is based on two different existing Twitter crawlers and is an improvement compared to the crawler used in our previous analysis as it is able to retrieve more tweets. Although the website is not made completely public yet as we testing is being finalised and data collection is still catching up (i.e., hosting it on a dedicated URL and advertising it), a **temporary demo website** can be accessed at the following address: https://evhart.github.io/fc-observatory/ (Figure 7).



This website is a work in progress. Therefore, the *reported information is currently unreliable and should not be used and considered valid.*

The Fact-checking Observatory tracks fact-checked misinformation spread.

We automatically collect and analyse fact-checks and misinformation spread on Twitter and produce weekly reports.

90+ Fact-checks 66,100+ Tweets

Understanding fact-checked misinformation spread.

With our reports, we aim to better understand what type of misinformation related to Covid-19 tends to spread, who are the Twitter users mostly affected by it and fact-checking information is shared. To this end, we apply automatic methods to collect fact-checked and misinforming content from the CoronaVirusFacts/DatosCoronaVirus Alliance Database and track this data on Twitter.

Topic spread

We track what type of misinformation about the Covid-19 pandemic is spread and compare how misinformation spread in comparison to corrective content.

Top Weekly misinformation and fact-checks

We identify the most shared msinformating content and fact-check over time.

Demographic impact

Using automated methods, we identify the demographics of the users that share misinformation and fact-checks.

Recent Reports

Twitter Covid-19 Misinformation Report 38
Misinformation spread about *Other* decreasing
August 31, 2020

Twitter Covid-19 Misinformation Report 37
Misinformation spread about *Conspiracy Theory* decreasing
August 24, 2020

Twitter Covid-19 Misinformation Report 36
Increase in misinformation spread about *Authorities*
August 17, 2020

Twitter Covid-19 Misinformation Report 35
Rise in misinforming posts about *Authorities*
August 10, 2020

Twitter Covid-19 Misinformation Report 34
Increase in misinformation spread about *Other*
August 3, 2020

Twitter Covid-19 Misinformation Report 33
Rise in misinforming posts about *Authorities*
July 27, 2020

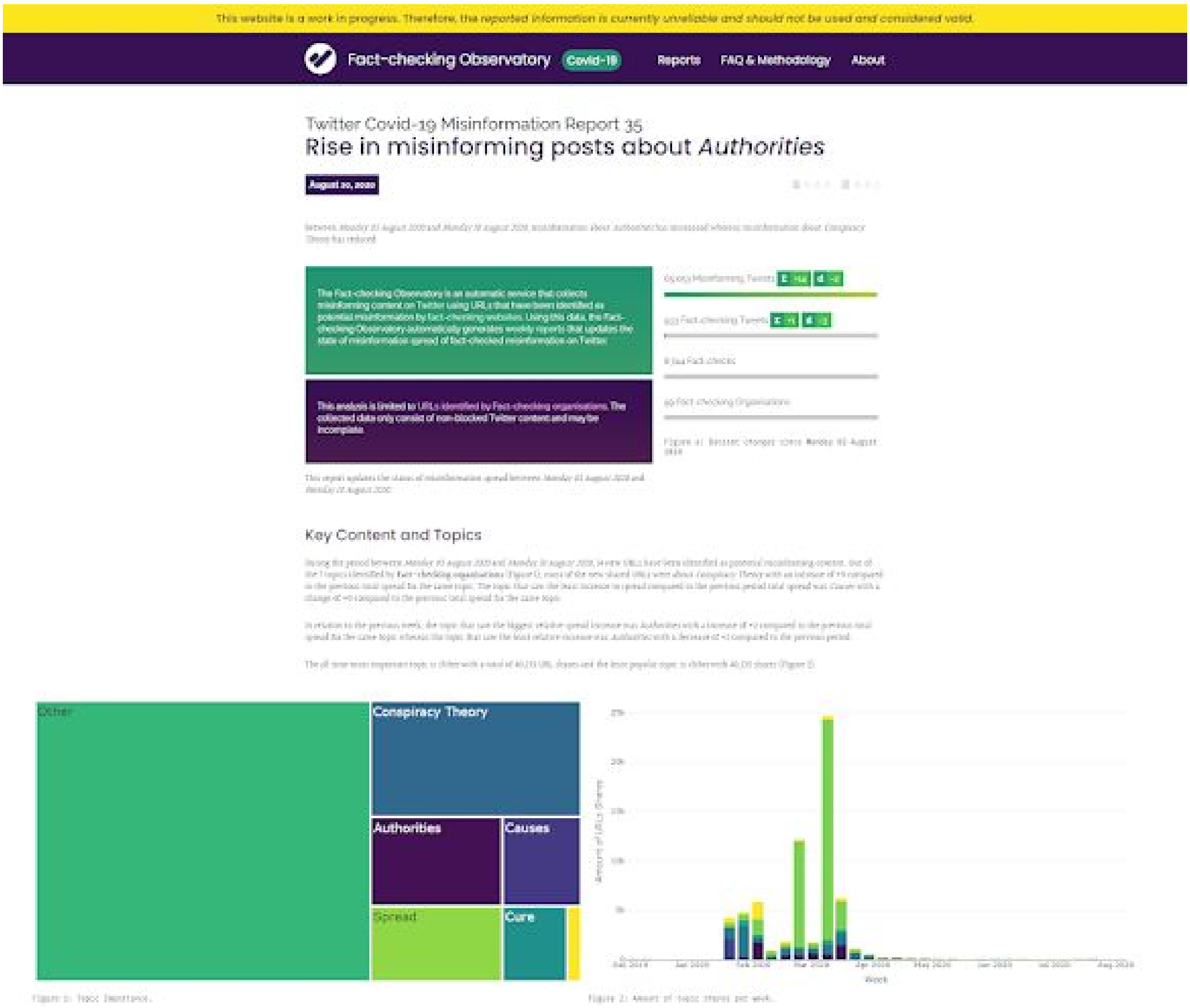
See all the reports

The Fact-checking Observatory tracks misinformation and fact-checks on social media.

We automatically collect data about misinformation and fact-checks on social media and generate weekly reports.

Figure 7- The temporary fact-checking observatory homepage displaying the project aims and recent reports.

A report example is presented in Figure 8. The report shows the amount of fact-checking misinformation and fact-check spread for a given week compared to previous periods with insights about the amount of fact-checking organizations, topical spread and demographics.



This website is a work in progress. Therefore, the reported information is currently unreliable and should not be used and considered valid.

Twitter Covid-19 Misinformation Report 35

Rise in misinforming posts about *Authorities*

August 20, 2020

Between *Monday [illegible] August 2020* and *Monday [illegible] August 2020*, misinformation about *Authorities* has increased whereas misinformation about *Conspiracy Theory* has reduced.

The Fact-checking Observatory is an automatic service that collects misinforming content on Twitter using URLs that have been identified as potential misinformation by fact-checking websites. Using this data, the Fact-checking Observatory automatically generates weekly reports that updates the state of misinformation spread of fact-checked misinformation on Twitter.

This analysis is limited to URLs identified by Fact-checking organisations. The collected data only consist of non-blocked Twitter content and may be incomplete.

[illegible] Misinforming Tweets [illegible] [illegible]

[illegible] Fact-checking Tweets [illegible] [illegible]

[illegible] Fact-checks

[illegible] Fact-checking Organisations

Figure 4: Recent changes since Monday [illegible] August 2020

This report updates the status of misinformation spread between *Monday [illegible] August 2020* and *Monday [illegible] August 2020*.

Key Content and Topics

During the period between *Monday [illegible] August 2020* and *Monday [illegible] August 2020*, [illegible] new URLs have been identified as potential misinforming content. Out of the [illegible] topics identified by fact-checking organisations (Figure 1), most of the new shared URLs were about *Conspiracy Theory* with an increase of [illegible] compared to the previous total spread for the same topic. The topic that saw the least increase in spread compared to the previous period total spread was *Causes* with a change of [illegible] compared to the previous total spread for the same topic.

In relation to the previous week, the topic that saw the biggest relative spread increase was *Authorities* with a increase of [illegible] compared to the previous total spread for the same topic whereas the topic that saw the least relative increase was *Authorities* with a decrease of [illegible] compared to the previous period.

The all time most important topic is *Other* with a total of [illegible] URL shares and the least popular topic is *Other* with [illegible] shares (Figure 2).

Figure 1: Topic Importance.

Figure 2: Amount of topic shares per week.

Figure 8- A partial view of a report displaying insights about misinformation and fact-checking content on Twitter for a particular week.

The processing pipeline and continuous crawling is implemented in a newly designed realtime processing framework that can be used for diverse tasks. The data collection and analysis pipeline for the report generation is displayed in the following figure (Figure 9).

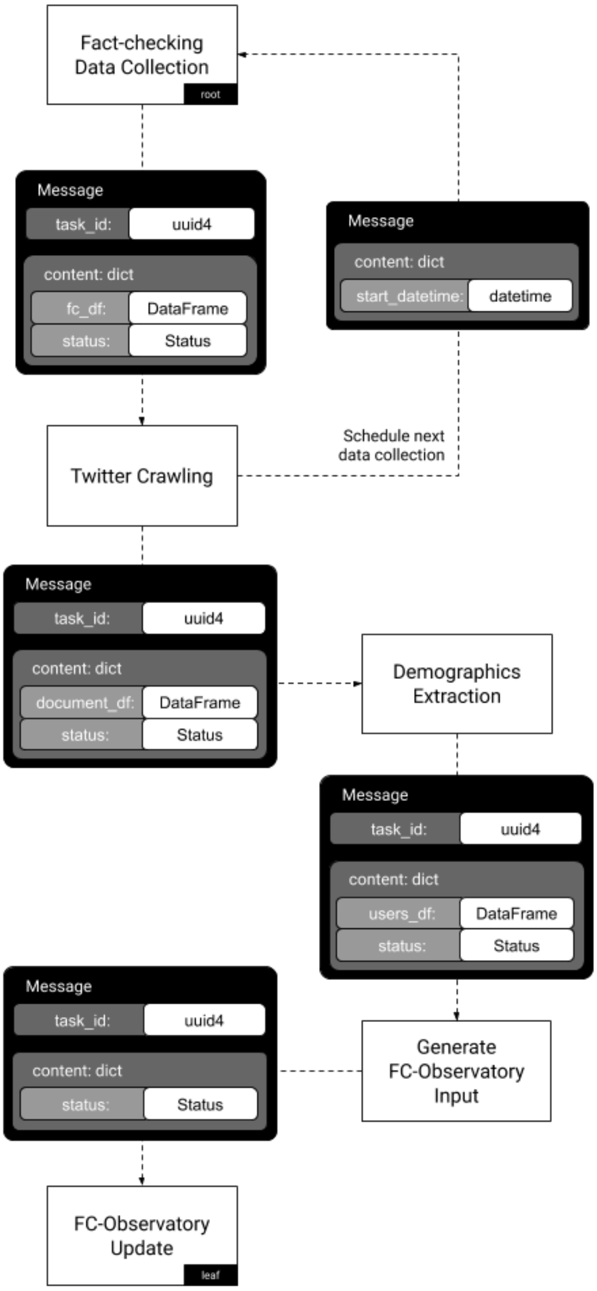


Figure 9- Continuous Fact-checking and misinformation Twitter data collection and report generation.

Since the website is only based on COVID-related misinformation, the list of URLs used for collecting misinformation and fact-checked URLs is obtained directly from the IFCN through an agreement with the Poynter institute.[4] Collecting the list of URLs directly from Poynter allows the reliable identification of COVID-related topics as they are labeled from IFCN organisations (i.e., Conspiracy theories, Authorities, Spread, Causes, Cure and Symptoms). This opens **future analysis of misinformation in topical terms, looking** . Similarly, as part of the website processing pipeline user demographics are automatically extracted this may be used in future for **understanding spread impact on gender, age and account types**.

[4] Poynter COVID-19, https://www.poynter.org/covid-19-poynter-resources/

As part of the HERoS project, we aim to reuse the data-collection framework developed for the HEIF project in order to obtain **more current data and continuous information** about the state of misinformation on social media during the COVID-19 pandemic. Based on feedback from fact-checkers, we plan to add the ability to **obtain reports about particular misinformation URLs** besides the weekly reports. We also hope to allow the retrieval of the misinformation report in a computer-readable format so the insights can be used by other tools developed for the project.

5 Discussion and Future Work

Our results show that at the COVID-19 level, fact-checked URLs are less shared (compared to misinformation in term of mean) during all the periods and that the standard deviation and mean are much higher for misinforming URLs. This indicates that there may be some **intrinsic features of misinforming URLs, potentially related to topic or sentiment, for example, that make them more shareable than fact-checks**. This echoes previous work that describes the enticement of emotion and novelty in misinformation [Vosoughi et al., 2018]. Likewise, this also indicates that the communities sharing fact-checks and those sharing misinformation are likely different, indicating that previous agent-based models that address the impact of fact-checkers on a network [Jin et al., 2013; Tambuscio et al., 2018; Tambuscio & Ruffo, 2019; Saxena et al., 2020] may need to be adjusted for **lower-than-expected inter-community contact**. This may also be related to the topic of misinformation and the culture or degree of polarisation apparent in the region in which the misinforming claim or its fact-check emerged, something we intend to explore in future deliverables. Finally, significant differences in sharing behaviour appears mostly during the ramping up period of the pandemic (the “early” phase) with large variations in deviation and means toward misinformation. This may be explained by the heightened fears and extreme uncertainty concerning the pandemic during that particular period in which the public need for information is outweighing the authority's ability to provide it [Huang et al., 2015; Spence et al., 2005, 2007].

5.1 Lessons Learned

At the relative level, we confirm previous findings showing the initial stage of circulation is generally associated with highest information spread in general [Starbird, 2018]. The absence of significance in general behaviour during the initial spread period and the observed high difference in standard deviation during that period shows that **most difference in spread behaviour happens in the later periods and may be associated with the virality of misinforming content** and its ability to spread deeper compared to fact-checking posts [Vosoughi et al., 2018]. This result also highlights that the difference in spread may be highly **related to the initial amount of shares of a given URL and to external contextual factors** rather than simply the intrinsic properties of the shared URLs (e.g, the relation between the pandemic state and the misinforming URLs topics rather than simply the misinforming URLs topics).

Causality analysis confirms that **misinformation spread can be predicted from fact-checking spread**. This relation is also confirmed by the FEVD analysis where misinformation spread is partially influenced by fact-checking (Figure 4). However, the opposite relation is not observed, meaning that **fact-checking spread behaviour is not causally related to misinforming behaviour** even though impulse analysis shows that to some extent misinformation spread shocks tend to lead to an initial increase in fact-checking spread.

Although the previous observation is encouraging, our results show that **the reduction in misinformation spread associated with an increase in fact-checking information is mostly temporary**. This indicates that the misinformation reduction power of fact-checking is impeded by its apparent inability to be shared over long periods of time. This echoes previous research that suggested that the amount of corrective information may play an essential role in reducing misinformation [Aird et al., 2018; Starbird et al., 2018]. To this end, better **fact-checking campaigns may be required to increase the virality of fact-checking content for increasing its shareability**.

5.2 Limitations and Future Work

Although **our approach is really *accurate***, since it does not depend on automatic annotations for identifying misinformation and fact-checked URLs, our data **covers only a small amount of misinforming content and does not contain variations of the same misinforming posts**. A relatively simple approach for future work would be to use automatic misinformation detection methods coupled with semantic similarity measures to detect content that is already fact-checked but associated with different URLs.

As already highlighted in the previous section, another area of improvement is in the ability to collect data more effectively by adopting a **continuous approach to data collection**. We plan to integrate the data collection and processing pipeline developed as part of our HEIF funded project to increase the amount and quality of collected social media data (at the time of the writing of this report, we estimate the new data collection approach to collect around 300k misinforming and fact-checking tweets compared to the 21,394 posts collected for the analysis presented in this deliverable).

In addition, as our results have shown, additional **topological and community analysis is required to better characterise the deviations and mean differences** observed in the multivariate spread analysis. We plan to increase the granularity of our analysis by obtaining more fine grained information about the users (e.g., demographics) that share misinformation as well as intrinsic misinformation and fact-checking content features such as topical information. Our involvement in the HEIF project gives us access to topical information from Poynter fact-checked claims as well as demographics information. we aim to update our analysis with such new information. Finally, for our analysis we focused only on Twitter, additional work should investigate **additional popular online communities**.

6 Conclusion

In this deliverable, we have presented the state of the art in **measuring the impact of fact-checking**, which is somewhat disarticulated from our knowledge of how misinformation spreads. We showed how current approaches rely on more holistic or aggregate measurements to understand the impact of fact-checking, such as information literacy or reductions in misinformation online. We outlined some of the important **features of the current COVID-19 pandemic,** such as the ballooning “infodemic”, fear and differing cultural norms, that confound efforts to measure the impact of fact-checking on specific misinforming claims using current approaches. We highlight the necessity for understanding the **"co-spread" of both misinformation and fact-checking information**, to be able to measure the impact of fact-checking on **specific misinforming claims** temporally and, potentially, at the geographic or platform level. We offer an initial analysis of the co-spread of misinformation and fact-checking information during the initial period of the COVID-19 pandemic. Although our results show that **fact-checking spread has a positive impact in reducing misinformation**, we have found that the impact of fact-checking is seriously impeded by three different factors: the amount of shared misinformation (which is disproportionately higher than fact-checking content), the different communities of fact-check sharers *versus* misinformation sharers, and the short period of time in which fact-checks are likely to spread. First, the amount of **shared misinformation is disproportionately higher for particular misinformation URLS compared to fact-checking content**; Second, it appears that users that share each type of content do not mix, meaning the **users tend to not correct themselves**; Third, **the impact of fact-checking tends to be short-lived** as spread in fact-checking information collapses. To overcome this, it will be necessary to build **interaction bridges between fact-checking and misinformation spreaders**, and create **fact-checking content that is more appealing**. This will help create a sustainable fact-checking information spread over time.

References

Aird, M. J., Ecker, U. K. H., Swire, B., Berinsky, A. J., & Lewandowsky, S. (2018). Does truth matter to voters? The effects of correcting political misinformation in an Australian sample. *Royal Society Open Science*, *5*(12), 180593. https://doi.org/10.1098/rsos.180593

Allgaier, J., & Svalastog, A. L. (2015). The communication aspects of the Ebola virus disease outbreak in Western Africa – do we need to counter one, two, or many epidemics? *Croatian Medical Journal*, *56*(5), 496–499. https://doi.org/10.3325/cmj.2015.56.496

Amazeen, M. A., Vargo, C. J., & Hopp, T. (2019). Reinforcing attitudes in a gatewatching news era: Individual-level antecedents to sharing fact-checks on social media. *Communication Monographs*, *86*(1), 112–132. https://doi.org/10.1080/03637751.2018.1521984

Brennen, J. S., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). *Types, Sources, and Claims of COVID-19 Misinformation*. 13.

Cherubini, F., & Graves, L. (2016). The rise of fact-checking sites in Europe. *Reuters Institute for the Study of Journalism, University of Oxford. http://reutersinsfitute. polifics. ox. ac. uk/our-research/rise-fact-checking-sites-europe*.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 Social Media Infodemic. *ArXiv:2003.05004 [Nlin, Physics:Physics]*. http://arxiv.org/abs/2003.05004

Craft, S., Ashley, S., & Maksl, A. (2017). News media literacy and conspiracy theory endorsement. *Communication and the Public*, *2*(4), 388-401.

DataCommons.org (2020). Fact Check Data, DataCommons.org, viewed 28 September 2020 from <https://www.datacommons.org/factcheck/download>

Dornan, C. (2020). *Science Disinformation in a Time of Pandemic*. http://www.deslibris.ca/ID/10104118

ECDC (2020). COVID-19 Data, ECDC, viewed 28 September 2020 from <https://opendata.ecdc.europa.eu/covid19/casedistribution/csv>.

Ecker, U. K., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2020). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology*, *111*(1), 36-54.

Freeze, M., Baumgartner, M., Bruno, P., Gunderson, J. R., Olin, J., Ross, M. Q., & Szafran, J. (2020). Fake Claims of Fake News: Political Misinformation, Warnings, and the Tainted Truth Effect. *Political Behavior*. https://doi.org/10.1007/s11109-020-09597-3

Fujishiro, H., Mimizuka, K., & Saito, M. (2020). Why Doesn't Fact-Checking Work?: The Mis-Framing of Division on Social Media in Japan. *International Conference on Social Media and Society*, 309–317. https://doi.org/10.1145/3400806.3400841

Gregory, J and McDonald, K. (2020). *Trail of Deceit: The Most Popular COVID-19 Myths and How They Emerged*, News Guard, viewed September 1 2020 <https://www.newsguardtech.com/covid-19-myths/>

Hannak, A., Margolin, D., Keegan, B., & Weber, I. (2014). Get Back! You Don't Know Me Like That: The Social Mediation of Fact Checking Interventions in Twitter Conversations. *ICWSM*.

Harman, S. (2020). The danger of stories in global health. *The Lancet*, *395*(10226), 776–777. https://doi.org/10.1016/S0140-6736(20)30427-X

Huang, Y. L., Starbird, K., Orand, M., Stanek, S. A., & Pedersen, H. T. (2015). Connected Through Crisis: Emotional Proximity and the Spread of Misinformation Online. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, 969–980. https://doi.org/10.1145/2675133.2675202

IFCN (2020). Verified signatories of the IFCN code of principles, IFCN, viewed 28 September 2020 from <https://ifcncodeofprinciples.poynter.org/signatories>

Jiang, S., & Wilson, C. (2018). Linguistic Signals under Misinformation and Fact-Checking: Evidence from User Comments on Social Media. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–23. https://doi.org/10.1145/3274351

Jin, F., Dougherty, E., Saraf, P., Cao, Y., & Ramakrishnan, N. (2013). Epidemiological modeling of news and rumors on Twitter. *Proceedings of the 7th Workshop on Social Network Mining and Analysis -*

SNAKDD ’13, 1–9. https://doi.org/10.1145/2501025.2501027

Jin, F., Wang, W., Zhao, L., Dougherty, E., Cao, Y., Lu, C.-T., & Ramakrishnan, N. (2014). Misinformation Propagation in the Age of Twitter. *Computer*, *47*(12), 90–94. https://doi.org/10.1109/MC.2014.361

Kahne, J., & Bowyer, B. (2017). Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. *American Educational Research Journal*, *54*(1), 3–34.

Kim, J., Tabibian, B., Oh, A., Schölkopf, B., & Gomez-Rodriguez, M. (2018). Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining - WSDM ’18*, 324–332. https://doi.org/10.1145/3159652.3159734

Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D., & Rich, R. F. (2000). Misinformation and the currency of democratic citizenship. *Journal of Politics*, *62*(3), 790–816.

Lewandowsky, S., Stritzke, W. G. K., Freund, A. M., Oberauer, K., & Krueger, J. I. (2013). Misinformation, disinformation, and violent conflict: From Iraq and the “War on Terror” to future threats to peace. *American Psychologist*, *68*(7), 487–501. https://doi.org/10.1037/a0034515

Margolin, D. B., Hannak, A., & Weber, I. (2018). Political Fact-Checking on Twitter: When Do Corrections Have an Effect? *Political Communication*, *35*(2), 196–219. https://doi.org/10.1080/10584609.2017.1334018

Mensio, M., & Alani, H. (2019). *MisinfoMe: Who is Interacting with Misinformation?* 5.

Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, *22*(19), 3055–3071. https://doi.org/10.1002/sim.1545

Nekmat, E. (2020). Nudge Effect of Fact-Check Alerts: Source Influence and Media Skepticism on Sharing of News Misinformation in Social Media. *Social Media + Society*, *6*(1), 2056305119897322. https://doi.org/10.1177/2056305119897322

Oeldorf-Hirsch, A., Schmierbach, M., Appelman, A., & Boyle, M. P. (2020). The Ineffectiveness of Fact-Checking Labels on News Memes and Articles. *Mass Communication and Society*, *23*(5), 682–704. https://doi.org/10.1080/15205436.2020.1733613

Safieddine, F., & Ibrahim, Y. (2020). *Fake News in an Era of Social Media: Tracking Viral Contagion*. Rowman & Littlefield.

Sarkar, S., Guo, R., & Shakarian, P. (2019). Using network motifs to characterize temporal network evolution leading to diffusion inhibition. *Social Network Analysis and Mining*, *9*(1), 14. https://doi.org/10.1007/s13278-019-0556-z

Schema.org (2020). ClaimReview, Schema.org, viewed 28 September 2020 from <https://schema.org/ClaimReview>

Shin, J., & Thorson, K. (2017). Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media. *Journal of Communication*, *67*(2), 233–255. https://doi.org/10.1111/jcom.12284

Sippitt, A., & Moy, W. (n.d.). Fact Checking is About What we Change not Just Who we Reach. *The Political Quarterly*, *n/a*(n/a). https://doi.org/10.1111/1467-923X.12898

Spence, P. R., Lachlan, Ken., Burke, J. M., & Seeger, M. W. (2007). Media Use and Information Needs of the Disabled During a Natural Disaster. *Journal of Health Care for the Poor and Underserved*, *18*(2), 394–404. https://doi.org/10.1353/hpu.2007.0047

Spence, P. R., Westerman, D., Skalski, P. D., Seeger, M., Ulmer, R. R., Venette, S., & Sellnow, T. L. (2005). Proxemic Effects on Information Seeking after the September 11 Attacks. *Communication Research Reports*, *22*(1), 39–46. https://doi.org/10.1080/0882409052000343507

Starbird, K., Dailey, D., Mohamed, O., Lee, G., & Spiro, E. S. (2018). Engage Early, Correct More: How Journalists Participate in False Rumors Online during Crisis Events. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–12. https://doi.org/10.1145/3173574.3173679

Tambuscio, M., Oliveira, D. F. M., Ciampaglia, G. L., & Ruffo, G. (2018). Network segregation in a model of misinformation and fact-checking. *Journal of Computational Social Science*, *1*(2), 261–275. https://doi.org/10.1007/s42001-018-0018-9

Tambuscio, M., & Ruffo, G. (2019). Fact-checking strategies to limit urban legends spreading in a segregated society. *Applied Network Science*, *4*(1), 116. https://doi.org/10.1007/s41109-019-0233-1

Tambuscio, M., Ruffo, G., Flammini, A., & Menczer, F. (2015). Fact-checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks. *Proceedings of the 24th International Conference on World Wide Web*, 977–982. https://doi.org/10.1145/2740908.2742572

Timothy Coombs, W., & Jean Holladay, S. (2014). How publics react to crisis communication efforts: Comparing crisis response reactions across sub-arenas. *Journal of Communication Management*, *18*(1), 40–57. https://doi.org/10.1108/JCOM-03-2013-0015

Twitter (2020) COVID-19 Stream Filtering Rules, Twitter Developer, viewed 28 September 2020 from <https://developer.twitter.com/en/docs/labs/covid19-stream/filtering-rules>

Upreti, A. (2020). Twitter Scrapper, viewed 28 September 2020 from <https://github.com/amitupreti/Hands-on-WebScraping>

Vaezi, A., & Javanmard, S. H. (2020). Infodemic and risk communication in the era of CoV-19. *Advanced Biomedical Research*, *9*(1), 10. https://doi.org/10.4103/abr.abr_47_20

Vlachos, A., & Riedel, S. (2014). Fact Checking: Task definition and dataset construction. *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, 18–22. https://doi.org/10.3115/v1/W14-2508

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wang, Z., Yang, B., Li, Q., Wen, L., & Zhang, R. (2020). Clinical Features of 69 Cases With Coronavirus Disease 2019 in Wuhan, China. *Clinical Infectious Diseases*, *71*(15), 769–777. https://doi.org/10.1093/cid/ciaa272

Wintersieck, A. L. (2017). Debating the truth: The impact of fact-checking during electoral debates. *American Politics Research*, *45*(2), 304-331.

Xian, J., Yang, D., Pan, L., Wang, W., & Wang, Z. (2019). Misinformation spreading on correlated multiplex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(11), 113123.

https://doi.org/10.1063/1.5121394

Xie, B., He, D., Mercer, T., Wang, Y., Wu, D., Fleischmann, K. R., Zhang, Y., Yoder, L. H., Stephens, K. K., Mackert, M., & Lee, M. K. (n.d.). Global health crises are also information crises: A call to action. *Journal of the Association for Information Science and Technology*, *n/a*(n/a). https://doi.org/10.1002/asi.24357