CSA
CONSEIL SUPÉRIEUR DE L'AUDIOVISUEL

FOCUS

THE SPREAD OF FAKE NEWS ON SOCIAL MEDIA:

A STUDY OF THE TWITTER SERVICE

November 2020

Les collections CSA

Contents

1. Abstract

Following the promulgation of the French law on the fight against information manipulation (2018), the *Conseil supérieur de l'audiovisuel* sought to improve its understanding of the mechanisms of propagation of fake news by initiating a specific study. The Covid-19 pandemic has made the problem of fake news even more acute in an era in which information is accessed through a wide range of media, including social media.

This study focuses more specifically on the mechanisms by which fake news is propagated and corrected on the Twitter social networking service. For this study, data concerning French-speaking information accounts active on Twitter and data concerning Tweets on certain topics of fake news were collected using Twitter's API. For the purposes of this study, information is considered to be false if it has been designated as such by journalists specialising in fact-checking. Similarly, Twitter accounts will be deemed to be either reliable or likely to share fake news per their classification by certain information auditors (in particular the *Le Monde* newspaper's *Décodeurs* web page, which forms part of the "Décodex" initiative).

The first part of the study is devoted to an examination of the least reliable information accounts. These accounts have a significantly lower number of subscribers on Twitter than the majority of reliable news accounts. However, in terms of ReTweets, accounts that are known to share fake news are on an equal footing with reliable accounts. Subscribers to unreliable accounts have a much higher propensity than subscribers to reliable accounts to contribute to the dissemination of information shared via these accounts (with 10 to 20 times more ReTweets per subscriber, depending on the indicator used).

To better understand the drivers and the impact of the virality of these accounts, the study proposes a qualitative analysis of their most shared Tweets. It emerges that the accounts that have been identified as unreliable focus strongly on current issues and divisive topics. Tweets relating to politics, immigration, health, religion and terrorism account for more than half of the corpus studied. The quantitative analysis shows an over-representation of terms related to delinquency, immigration, Israel and Palestine, paedophilia, Islam and even freemasonry. The tone adopted is mainly based on **informing their subscribers and criticising the cited people or institutions**. The use of non-text content (images, videos, links, etc.) also seems to characterise a strategy for increasing the legitimacy of published content, based mainly on sites linked to these accounts but also on sources drawn from traditional media.

Individuals exposed to an "echo chamber" phenomenon[1] are in the minority among the subscribers to the least reliable Twitter accounts: less than 20% of subscribers to these accounts do not follow any reliable account. Indeed, on average, these individuals mainly follow reliable accounts. Yet some (minority) individuals' subscription habits prevent them from coming into contact with sources of information that could correct fake news that they have obtained

[1] This is understood within the context of this study as being exposed to sources known to share fake news, without at the same time following reliable accounts**.**

elsewhere. A graph-based analysis also shows that many unreliable accounts have a significant proportion of common subscribers or follow each other on Twitter.

The first part of the study concludes with an analysis of the Twitter accounts of fact-checkers affiliated with traditional media. On average, the accounts of journalism units specialising in fact-checking have more followers than accounts categorised as unreliable, but their publications generate fewer interactions than the latter. In terms of the topics mainly covered, these two types of accounts only have politics and religion in common. Fact-checkers are more interested in topics related to their profession, such as media news or content related to media and information literacy.

The study then presents a more detailed analysis of collected Tweets that relate to preselected fake news, irrespective of whether those Tweets propagated or refuted that information.

All the fake news in the study shows a high concentration of Tweets over a very short period. This observation highlights the difficulty of the role of fact-checkers, who must react very quickly if they are to have an impact on the spread of fake news. Moreover, the fake news in the study is evenly divided between topics for which most of the information shared is fake news, and topics for which, conversely, most of the information shared is verified. However, for the topics with the highest volume of Tweets, fake news prevails in all cases. A study of fake news spreading chronology shows that, contrary to what might have been expected or hoped for, **"genuine" news information does not displace fake news**. For topics where fake news is in the majority, discussion on these topics quickly fades away on Twitter before the verified information can ever achieve a majority position.

This analysis is supplemented by an analysis of the number of subscribers to accounts that have shared fake news or verified information. As verified information is frequently shared by accounts with a high number of subscribers (especially when such accounts are attached to traditional media), the audience for verified information is larger than an analysis based on the volume of Tweets alone would suggest. For some topics, however, fake news remains more visible than verified information.

By analysing the most viral Tweets in the corpus under study, we can highlight several interesting features. Fake news is mainly a vehicle for criticising the authorities or expressing a sense of panic; for example, on sensitive health issues. It is based on information that has not been fact-checked or has been described as false by journalists, who in turn make corrections using practices specific to social media. However, these corrective measures have a low level of engagement (ReTweets, comments) compared to accounts that are likely to disseminate fake news. It should nevertheless be noted that in several of the cases studied, the corrections were made directly by internet users, drawing on information in these media articles.

2. Observations by Twitter

As part of its co-operation with digital platforms, the *Conseil* shared the objectives and results of this report with Twitter before its publication. Twitter representatives were given the opportunity to submit their remarks to the *Conseil*, which took into account those comments that it considered relevant in the final drafting of the report.

It also suggested that Twitter be given a space in the report to express itself and provide any information that the company considered useful, particularly on its efforts to combat the manipulation of information. This can be found below:

Twitter contribution

Our approach to facilitating positive public conversation

By facilitating public conversation, our goal is to make it easier to search for credible information on Twitter and to limit the spread of potentially harmful and misleading content.

Twitter recognises that disinformation is a complex social problem that requires a concerted solution. To this end, we are committed to working with the European Commission, governments, regulators, civil society and academic institutions to develop collaborative and workable solutions. We are signatories to the EU Code of Practice on Disinformation, we have established a partnership with UNESCO on media literacy, and we publish datasets on disinformation campaigns abroad to promote a shared understanding of the tactics and threats of bad actors.

We do not solely rely on interaction from account holders. We take proactive steps to try to prevent the spread of malicious accounts and Tweets, which involves trying to stay ahead of the evolving tactics of the bad actors. In an attempt to achieve this goal, we have made a number of changes, including enhanced security policies, better tools and resources to detect and stop malicious activity, stricter advertising rules and increased transparency to promote public understanding in all these areas.

What we are doing about disinformation during the COVID-19 pandemic

As the global community grapples with the COVID-19 pandemic, Twitter helps people find reliable information, connect with others and follow what is happening in real-time.

Throughout this unprecedented period, Twitter has continued to adapt and update its policies and actions, and to increase transparency and share more data so that experts and the public are better able to analyse the evolution of the debate around COVID-19**.**

We have expanded our policy guidance to address content that directly contradicts COVID-19 guidance from authoritative global and local public health information sources. We have introduced new labels and warning messages that provide context and additional information on certain Tweets containing disputed or misleading information related to COVID-19. This will make it easier to find facts and make informed decisions about what people see on Twitter.

Our work to protect the civic conversation

The conversation that takes place on Twitter is never more important than during elections. Twitter is a place where people come for direct information from elected officials and candidates. It is where they find the latest news and, increasingly, it is an important source of information on when and how to vote in elections. As more and more people seek ways to vote safely and exercise their basic civil rights during the COVID-19 pandemic, the need for this type of information has only increased.

Our civic integrity policy, therefore, targets the types of content that are most directly harmful; that is, content relating to information or misrepresentation about how to participate in civic processes; content that may intimidate or suppress participation, and false affiliation. In September 2020, in light of changes in how people will vote in 2020, and in keeping with our commitment to protect the integrity of the electoral conversation, we expanded this existing framework. The aim is to enhance protection against content that could distort voting and help stop the spread of harmful fake news that could compromise the integrity of an election or other civic process. We are now also labelling or removing false or misleading information intended to undermine public confidence in an election or other civic process. In line with our current approach to enforcement, Tweets labelled under this extended policy will have reduced visibility throughout the service. Reducing the visibility of Tweets means that we will not amplify Tweets on several surfaces through Twitter.

Transparency

Transparency is at the heart of the work we do at Twitter. The open nature of our service has led to unprecedented challenges in protecting freedom of expression and the right to privacy as governments around the world increasingly attempt to intervene in this open exchange of information. We believe that transparency is a key principle in our mission to protect the open internet and to advance the Internet as a tool for progress.

A fundamental belief in the power of open public conversation inspired Twitter to launch one of the first transparency reports in the sector in 2012.

Access to data

Since 2006, Twitter's APIs have given researchers and developers access to news and events happening around the world. Twitter strongly believes in free access to data to study, analyse and contribute to the public conversation, which is why we continue to maintain a public-available API. Researchers are using Twitter data to provide valuable feedback on how online conversations and interactions evolve on and off Twitter. We continue to provide more accessible means of making data and information available to the public and researchers.

Our service is the largest source of real-time data on social media, and we make this data freely available to the public through our public API. No other major service does this.

To further support our ongoing efforts to protect the public conversation and help people find authoritative health information relating to COVID-19, Twitter has published a specific COVID-19 API access point in Twitter developer labs to allow approved developers and researchers to study the public conversation about COVID-19 in real-time. It is a unique dataset that covers tens of millions of Tweets per day and provides a snapshot of the evolution of the global public conversation around an unprecedented crisis. Making this access available for free is one of the most unique and valuable things Twitter can do as the world comes together to protect our communities and seek answers to pressing challenges.

On 12 August, we also presented the new Twitter API. Rebuilt from the ground up to deliver new features faster, this release includes the first set of new devices and features to enable developers to help the world connect to the public conversation taking place on Twitter.

Information operations

Since 2019, we have been publishing all accounts and content associated with potential disinformation operations that we have found on our service since 2016. The archives are now the largest of their kind in the sector. In 2019, new bodies of data have been added to the archives, enabling members of the public, governments, academics and researchers to study these behaviours, learn and build capacity for future media literacy to combat manipulation and disinformation. The data includes more than 160 million Tweets and more than eight terabytes of content.

3. Introduction

The *Conseil supérieur de l'audiovisuel* (CSA), the French public authority in charge of regulating audiovisual communication, is the guarantor of **freedom of expression, in the interest of the public and professionals of the sector**. The law of 22 December 2018 on the fight against information manipulation gives it new powers to respond to some of the challenges for democracy that have emerged in the digital world.

The legislation establishes a duty of co-operation by operators of online platforms in the fight against the dissemination of fake news[2]. More specifically, this concerns platforms with more than five million unique visitors per month[3]. The CSA has sent them recommendations[4] which, per Article 11 of the above-mentioned law, concern issues such as the establishment of an accessible and visible reporting system, the transparency of their algorithms, the promotion of content from companies and press agencies as well as audiovisual communication services, the fight against accounts responsible for the widespread propagation of fake news, informing users on the promotion of information content, and media and information literacy. The CSA has also set up a committee of experts on the fight against online disinformation, which brings an economic, scientific, technological and ethical perspective to its work.

The **Covid-19 pandemic** has made the problem of fake news even more acute in an era in which access to information is provided through an unprecedented variety of media, including social media[5]. The World Health Organisation has thus spoken of fighting not only the Covid-19 pandemic but also an *"infodemic"*. According to its managing director, "*fake news spreads faster and more easily than this virus, and is just as dangerous*" [6,7]. It is also for this reason that the European Commission has published the consultation document outlining the Digital Services Act, a package that should result in a proposal by the end of 2020. Its primary purpose is to change the liability regime for platform operators[8].

This is the situation that has prompted this study on the phenomena of propagation of fake news. However, the study does not aim to verify Twitter's compliance with its obligations or to

[2] The definition of platform operators used in the decree implementing the law of 22 December 2018 is that of Article L. 111-7 of the French consumer code: "any natural or legal person offering, on a professional basis, whether for payment or not, an online communication service to the public based on: 1) The classification or referencing, by means of computer algorithms, of content, goods or services offered or put online by third parties; or 2) the bringing together of several parties with a view to the sale of a good, the provision of a service or the exchange or sharing of content, goods or services".

[3] This threshold was set by Decree no. 2019-297 of 10 April 2019 on the information obligations of online platform operators promoting information content related to a debate of general interest.

[4] Recommendation no. 2019- 03 of 15 May 2019 by the *Conseil supérieur de l'audiovisuel* to online platform operators as part of the duty to co-operate in the fight against the dissemination of fake news.

[5] According to the Digital News Report 2019 of the Reuters Institute for the Study of Journalism, the use of social media for obtaining information is on the rise in France, reaching 42%. The French now obtain almost as much information online (69%), including social media, as they do via television (71%).

[6] WHO. *Munich Security Conference*. 15/02/2020.

[7] Some of the most common fake news about coronavirus includes the claim that 5G waves are responsible for the pandemic, that coronavirus contains inserts of artificial HIV, that chlorine spray can kill the virus, that "Bill Gates' vaccine" contains a tracking chip, and that vitamin C is an effective treatment for the virus. Digimind, July 2020, "*Covid : les Fake News les plus répandues sur les médias sociaux en France " (Covid: the most widespread Fake News on social media in France)*.

[8] European Commission. *The Digital Services Act package*. 11/06/2020.

assess its actions during the pandemic. Nor does it aim to produce a cross-cutting report of all the issues raised by the dissemination of fake news. Based on **quantitative and qualitative analyses,** it sheds light on the **phenomena of the propagation of fake news** on a social media platform that is widely used in France.

This work enhances the *Conseil*'s knowledge in the performance of its duties and reflects its growing technical expertise in data processing tools and methods. It also aims to contribute to a better understanding of the phenomena of fake news dissemination by the general public and by the public sphere.

3.1. Definitions and scope of the study

The increasing attention paid to the spread of fake news has introduced several concepts into the common vocabulary. These are sometimes used in equivalent ways but nevertheless describe sometimes different realities.

Disinformation includes "information that *is verifiably false or misleading, that is created, presented and disseminated for profit or with the deliberate intention of misleading the public, and is likely to cause public harm"*[9]*.* It is a technique used, for example, in campaigns to destabilise a group of people, a company or even an election. This term, along with the expression "***fake news***", has been popularised by investigations into possible Russian interference in the 2016 US presidential election campaign. The concept of "fake news" does not, however, encapsulate all the facets of the phenomenon of information manipulation because *"most of the time, manipulators do not adopt a position in relation to the truth: they simply seek to achieve an effect"*[10].

Misinformation is information whose inaccuracy *"is not intentional"*[11], but is—for example—the result of incorrect data interpretation. Internet users may also share content from their contacts in good faith, without having verified its origin, which contributes to the spread of false information.

These various concepts were also accompanied by the concept of **fact-checking**, which has been made necessary by *"an ever-stronger wave of fake news"*[12]. This approach consists of *"identifying and processing items of information, one by one, which [...] appear to require clarification"*[13], particularly by contextualising them and providing additional information to Internet users. Units consisting of fact-checking journalists have been formed over recent years in French newsrooms; for example, "FakeOff" for *20 Minutes*, "AFP Factual" for *Agence France-Presse*, "Les Observateurs" at *France 24*, "Vrai ou fake" at *France Info*, "CheckNews" at *Libération* and "Les Décodeurs" at *Le Monde*.

[9] European Commission. *A Multi-Dimensional Approach to Disinformation, Report of the Independent High-Level Group on Fake News and Online Disinformation*. 03/2018

[10] CAPS and IRSEM. *Les Manipulations de l'information : un défi pour nos démocraties (Information Manipulations: a challenge for our democracies).* 08/2018

[11] Ibid.

[12] *Le Monde*. *Le Décodex, un outil de vérification de l'information (The Décodex: a fact-checking tool)*. 23/01/2017.

[13] Ibid.

It is neither within the *Conseil*'s competence nor responsibility to assess the veracity of information. This study will therefore focus on fake news in general without attempting to identify whether it intends to manipulate, or whether the errors are unintentional. **Fake news will be considered here as information that has been analysed by journalists—particularly those specialising in fact-checking—and found to be incorrect.** Sites, media and social network accounts identified by these journalists as having already spread fake news will also be considered here as sources conducive to the dissemination of such information.

3.2. The strong challenge posed by social media in terms of access to information and fake news

According to the Digital News Report 2019 of the Reuters Institute for the Study of Journalism, **the use of social media for obtaining information is on the rise in France, reaching 42% of the population**. The French now obtain almost as much information online (69%), including social media, as they do via television (71%)[14]. According to the same report, in addition to the widespread use of social media for accessing information, there is a growing mistrust of the media. As a result, **French people's confidence in information in the broadest sense is now the lowest in Europe**, at 24% (down 11 points from the previous report). Social media, which are still widely used, are not immune to this crisis of confidence: 14% of those surveyed felt that they "*disseminate conspiracy theories, share biased information and filter information by means of algorithms*".

There are therefore strong democratic issues underlying fake news and its propagation on social media, all the more so as they are not generated by a single category of the population but may stem from various categories of individuals[15].

The choice was made to carry out this study using data collected on the Twitter social networking service. Twitter provides access to the data via an API[16], a prerequisite for implementing the analysis methodology planned by the *Conseil*. Conversely, Facebook, which has a considerable number of users in the world (2.5 billion[17]) and in France (37 million), has taken the decision, since 2018 and the "Cambridge Analytica" case, to restrict access to its data, in particular by modifying the terms and conditions of use of its APIs. The discussion initiated by the *Conseil*'s services with Facebook's management to obtain means of access to the platform's public content for the purposes of this study was not conclusive.

The *Conseil* is aware of the limitations of this approach, particularly as the sociological profile of Twitter users is not representative of internet users as a whole. According to several researchers

[14] Ibid.

[15] Andrew Guess, Jonathan Nagler, Joshua Tucker. Science Advances *Less than you think: Prevalence and predictors of fake news dissemination on Facebook.* 09/01/2019.

[16] An API is a programming interface provided by a service to establish connections between several software programs in order to exchange data. The provision of an API allows administrators to define the terms of use (paid or free, threshold of extractable data, etc.) and/or the information to which users will have access (inaccessibility of personal data, inability to access results beyond a certain date, etc.).

[17] Lefigaro.fr. *Facebook a franchi la barre des «37 millions d'utilisateurs» en France en 2019 (Facebook crossed the "37 million users" mark in France in 2019)*. 10/02/2020.

at SciencesPo's MediaLab, "*Twitter is an **important sounding board for media stories**. All studies show that the dominant audience on Twitter (unlike Facebook) is very specific: it is mainly made up of individuals of a high social and cultural level, who—among those who share news—have strong political convictions"*[18]*.* Moreover, according to these authors, who analyzed the links shared by the accounts of nearly 400 media, Twitter is characterised by *"blurrier boundaries, especially between right-wing and extreme right-wing media, which can help certain marginalised information to be relayed beyond its natural audience*."

3.3. Description of the Twitter service

Twitter is a **social networking service for publishing short messages—*Tweets***— of no more than 280 characters. The service describes itself as "*the showcase of what's happening in the world and the current topics of conversation*"[19]. It is accessible free of charge, for publishing (with an account) and viewing (without an account) Tweets from public accounts.

These Tweets can contain different types of media and hashtags ("#" followed by a word or phrase written without spaces, such as #BlackLivesMatter). The use of these **hashtags** allows Tweets to be grouped together on the same subjects, which are highlighted on the social networking service—as "trending topics"—in cases where, for example, a large number of messages are concentrated into a short space of time. It is thus possible to follow "conversations" that are popular on the social networking service at a given point in time, as their display method can be customised for each user (by location, for example).

All Tweets are collected on a **profile** linked to the user who publishes them, which may be either public or private, and also within a separate **timeline**—displayed upon login—which consists of an aggregation of all posts from profiles to which a given user subscribes. In this case, the user is said to "follow", or be a "follower" of, one or more users. Every registered user on Twitter has a profile containing his or her Tweets and a news feed. This news feed has been offered in reverse chronological order of content since the beginning of the service. Since 2016, the user can also choose an alternative version of this news feed with an **algorithmic component**[20]: the display of Tweets is then personalised according to several criteria, such as the number of user interactions on the content. Twitter offers the ability to switch at any time from a news feed with algorithmic input to a feed that is reverse chronological only, which always exists and remains available for each user.

Users do in fact have several options for interacting with each Tweet, whether it comes from an account to which they are subscribed or from a public account that they do not follow. They can then republish it so that it appears on their profile and in their subscribers ' news feeds (a **ReTweet**). They can also republish it by **commenting** on it**,** always subject to the 280-character limit (**QuoteTweet**), add the Tweet to their **favourites** list and keep it on their profile in a dedicated section called "Likes") or **share** it, either by copying the link to the Tweet or by private message. This last feature makes it possible to send messages to another user that are neither accessible to others nor published on your profile or in your followers' news feeds.

[18] Dominique Cardon, Benjamin Ooghe-Tabanou, Guillaume Plique, Jean-Philippe Cointet. SciencesPo Médialab. *Les nouveaux circuits de l'information numérique (The new digital information channels).* 2019.
[19] Twitter, About.
[20] Slate. *Twitter's New Order. Inside the Changes That Could Save Twitter's Business – and Reshape Civil Discourse*. 05/03/2017.

3.4. Objectives of the study

This study aims at **better understanding the mechanisms by which fake news is propagated on the Twitter social networking service**. It takes the form of two analyses based on largely distinct methodologies. The first is based on an analysis of Twitter accounts labelled according to their reliability. The second is based on the analysis of a certain number of topical issues that have given rise to fake news dissemination.

In its first part, the study therefore addresses the issue of disinformation through the labelling of news accounts active on Twitter. The labelling adopted by the *Conseil* for this analysis uses a classification made by *Le Monde*'s "Décodex" based on the reliability of the information shared by these accounts. The analysis methodology was then to compare the characteristics of the different categories of accounts (in particular accounts providing reliable information, and accounts known to share fake news). This comparison includes the number of subscribers to the different types of accounts and the propensity for Tweets from these different types of accounts to be ReTweeted. This section also presents an analysis of the issues addressed by accounts known to propagate fake news. It also includes an analysis of the diversity of information accounts followed by Twitter users to identify the presence of so-called "echo chamber" phenomena. Finally, this section will also present an analysis of the activity of certain fact-checker accounts found on Twitter.

A second part of the study presents the results of an analysis based on the labelling of Tweets linked to certain fake news items identified by fact-checkers. This section includes a quantitative analysis of the chronology of the spread of fake news and its refutation and also provides a qualitative analysis of the content of the Tweets with the highest virality.

The study focuses on French-language news accounts and Tweets published in French. It is therefore based on the assumption that French-speaking users will mainly use French-language information sources[21]. Its conclusions are therefore not necessarily transposable to the non-French-speaking world.

3.5. Data collection on Twitter

This study is based on data collection carried out on Twitter through the APIs offered free of charge to the general public by the platform. As far as the *Conseil* is aware, no platform similar to Twitter offers an open and free API providing such comprehensive access to the data on its platform.

The API provided by Twitter allows external developers to use the platform's data to develop programs or applications, or to carry out projects for statistical purposes. As such, the data made available via the API include the Tweets sent by an account, and also all the interaction statistics specific to each Tweet (number of ReTweets, number of "Likes", etc.) or the identity of the accounts that follow or are followed by a given account. The API also makes it possible to search for Tweets containing a given keyword. The versions of the API made available free of charge by

[21] For example, the Digital News Report 2019 of the Reuters Institute for the Study of Journalism confirms that the sixteen online news sites most consulted by the French are all French-speaking.

Twitter and used by the *Conseil* have many restrictions in terms of the number of queries that can be made to the API over a given period of time, and also in terms of the depth of the history in the case of keyword searches of Tweets.

In order to query the Twitter API, the *Conseil*'s services used the Tweepy and Twython "libraries" (for Python-language code) and the rtweet library (for R-language code). By using these libraries to issue queries, it has been possible to collect data and set up databases, the contents of which will be presented in detail in the following sections.

3.6. Precautions taken by the CSA to ensure compliance with the European General Data Protection Regulation

The above-mentioned data has been extracted and processed in compliance with the General Data Protection Regulation (GDPR) and the law of 6 January 1978 (the *Informatique et libertés* Data Protection law). These regulations cover any processing, whether computerised or not, which, outside the domestic context, relates to "information relating to identified or identifiable people". While the study did not involve any examination of specific individuals, it did involve the processing of certain personal information. Firstly, the processing of personal data was necessary because of the way the Twitter API works. For example, the study looked at the diversity of accounts followed by some Twitter users. This analysis initially involved collecting the user IDs of these individuals, which constitute personal data. These IDs were then used to send queries to the Twitter API to collect lists of accounts followed by these users. The collection of identifying data is purely incidental and is not intended for individual treatment. This identifying data was therefore deleted at the end of the collection. Secondly, the text content of some Tweets was also collected. The Tweets concerned by this collection are those issued by information accounts and those containing certain keywords referring to specific fake news items. The very nature of the Tweets collected reduces the likelihood that they contain personal information, although this possibility cannot be excluded.

Although the aim of the study was therefore to minimise the collection and processing of personal data, it would incidentally be possible to collect such data. A notice was therefore published on 13 September 2019 on the CSA's website to inform the public of the existence of a study on the phenomenon of information propagation on social media[22]. In accordance with GDPR principles, this notice stated the purposes of the data processing, the legal basis of the processing, the categories of personal data processed, the source of the data, the recipients of the data and the data storage period. This study is consistent with the CSA's stated purpose for processing this information, which indicated that the *Conseil* was seeking to better understand motivations for disseminating information on the Twitter platform.

The data collected by the CSA for this study will not be reused for other purposes. The GDPR provides for six separate legal bases for the collection and processing of personal data. In the present case, the CSA has used as its legal basis the exercise of the public authority vested in it pursuant to law no. 86-1067 of 30 September 1986, and specifically Article 17-2 thereof. Since the

[22] *Conseil supérieur de l'audiovisuel. Données personnelles : le CSA mène une étude sur le phénomène de propagation des fausses informations sur les réseaux sociaux (Personal data: the CSA is studying the phenomenon of the spread of fake news on social media)*. 13/09/2019.

entry into force of Law no. 2018-1202 of 22 December 2018, the CSA is indeed required to contribute to the fight against the dissemination of fake news and has in particular been assigned the responsibility, where necessary, of making recommendations to online platform operators to fight more effectively dissemination of such information. To carry out these new responsibilities, the CSA needs to have an adequate understanding of the phenomenon of fake news dissemination on online platforms, especially social media such as Twitter. This study contributes to the *Conseil*'s expertise on this subject.

4. Analysis of the activity of reliable and unreliable information accounts and fact-checking accounts

This first section presents an analysis of the characteristics of different categories of Twitter accounts: Twitter accounts categorised as reliable or unreliable by *Le Monde*'s “Décodex” initiative and the accounts of certain information auditor units operated by journalists (fact-checkers) active on the Twitter platform.

4.1. Introduction to the Décodex

In order to carry out its analysis, the *Conseil* did not categorise the accounts itself, but used the categorisation from *Le Monde*'s “Décodex” without modification. This categorisation of information sources according to their reliability is available at https://www.lemonde.fr/webservice/decodex/updates.

This directory of Décodex sources is intended, according to *Le Monde*[23], to help Internet users find their way around in a context where many fake news items circulate on the internet and where the sources of such fake news are frequently the same. The classification is based on an analysis of the site's sources, authors, their identification, the presence of corrections in case of errors, links to dubious sites, etc. The tool is most prominently available as a search engine on the *Le Monde* website, where an Internet user can enter the address of a site about which they would like to obtain information and a browser extension which, while browsing, shows information about the sites being viewed. The classification uses "blue" sites (categorised "1" in the database made freely available by *Le Monde*) for parody sites, "red" sites (categorised "2" in the database) that regularly disseminate fake news, and "orange" sites (categorised "3" in the database) for sites whose reliability or approach is questionable. Sites that are considered reliable do not have a specific colour and are placed in category "4" in the database.

As will be detailed below, the classification used by *Le Monde* journalists has also been adopted by several researchers studying disinformation phenomena. In addition, in a 2019 study, 14,000 hypertext links exchanged between 391 French media between April and November 2018 were mapped, providing an automatic display of a site classification that was "*very close to the classification carried out by hand by the Le Monde newspaper's “Décodeurs” team to rate the reliability of information sources using professional criteria"*[24].

[23] *Le Monde*. *L’annuaire des sources du Décodex : mode d'emploi (Décodex directory of sources*:*user manual)*. 23/01/2017.

[24] Dominique Cardon,, Benjamin Ooghe-Tabanou, Guillaume Plique, Jean-Philippe Cointet. SciencesPo Médialab. *Les nouveaux circuits de l’information numérique (The new digital information channels)*. 2019.

No data concerning the socio-demographic characteristics of subscribers to accounts categorised by the "Décodex" were collected for this study. However, information published by *Le Monde*[25] indicates that sites in the "red" and "orange" categories are not the exclusive preserve of younger and less-educated people. The age group that seems to stand out in its appetite for dubious sources is the 25 to 49-year-old age group. People from higher socio-professional categories are also more likely to view less reliable sources.

4.2. Statistical analysis of the various categories of accounts

The statistical analysis of the various categories of "Décodex" accounts is based on the last 1,000 Tweets collected from the timeline of each account (collection made on 13 September 2019). The data relating to these Tweets (number of ReTweets[26], number of favourites, etc.) was also collected. For each of these accounts, data was also collected on the number of subscribers, the number of accounts followed and the total number of Tweets published since the account was created. The following table presents the characteristics of the accounts analysed at the collection date. These statistics exclude Tweets that were not issued by the information account in question but are instead ReTweets of Tweets from another account[27].

DÉCODEX CATEGORY	MEAN NUMBER OF SUBSCRIBERS	MEAN NUMBER OF RETWEETS BY TWEET	MEAN NUMBER OF FAVOURITES	NUMBER OF ACCOUNTS
1 (parody)	152,004	47	173	12
2 (circulating fake news)	14,239	18	17	29
3 (doubtful)	85,335	20	22	21
4 (apparently reliable)	516,962	14	29	157

[25] *Le Monde. La désinformation ne touche pas seulement les jeunes et les personnes peu diplômées (Disinformation has a wider audience than just young people and people with little education). 05/08/2019.*

[26] ReTweets in the form of quotes are also taken into account in this analysis. However, these quoted Tweets are sometimes used not to help propagate the original Tweet, but to criticise it. The analysis carried out in this section could therefore lead to an overestimation of the virality of the different types of accounts (and in particular, of accounts propagating fake news).

[27] Using the data supplied by the Twitter API regarding the number of ReTweets, it is not possible to distinguish whether the ReTweets have been made from the information account in question; incorporating ReTweets from Tweets that are not published by the information account would therefore distort the analysis of the virality of these accounts (if information account B ReTweets a Tweet from account A, the statistics for the Tweet from account B include all the ReTweets made from account A or from other third-party accounts).

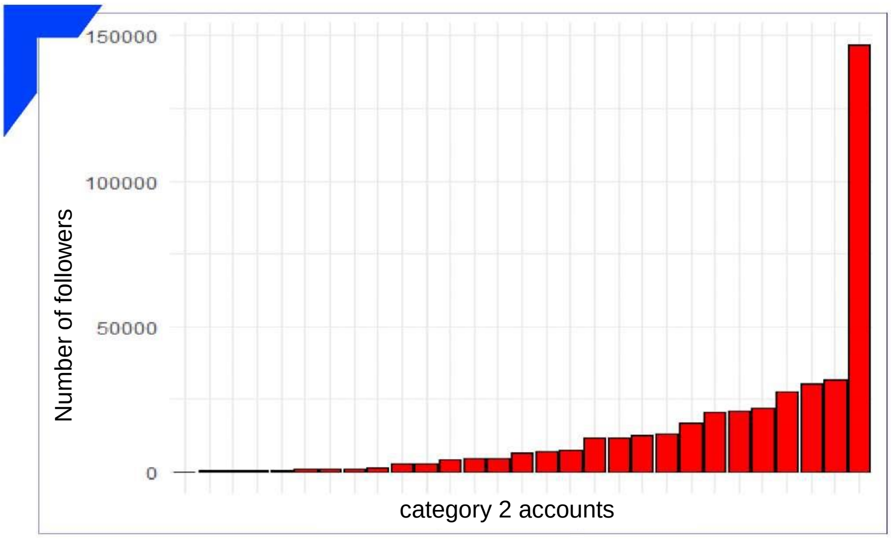
Due to the high dispersion of indicators between the different accounts, this analysis of mean averages can usefully be complemented by a median analysis[28]:

DÉCODEX CATEGORY	MEDIAN NUMBER OF SUBSCRIBERS	MEDIAN NUMBER OF RETWEETS BY TWEET	MEDIAN NUMBER OF FAVOURITES	NUMBER OF ACCOUNTS
1	18,968	10	17	12
2	6,409	2	2	29
3	21,596	8	5	21
4	73,945	4	8	157

A more detailed analysis of the number of subscribers and the number of ReTweets per Tweet for the different account categories is given in the following sections.

4.2.1. Reliable accounts have a significantly higher subscriber base than other types of accounts

Whether the analysis focuses on median or mean averages, it appears that the least reliable accounts (category 2 or 3) have a number of subscribers—and therefore a direct audience—that is very significantly lower than for category 4 accounts considered as reliable. However, this analysis of median and mean averages hides significant dispersion in the number of subscribers. For example, category 2 accounts have from 285 to nearly 150,000 subscribers. In addition, there is a very large gap between the most widely followed category 2 account and other accounts:

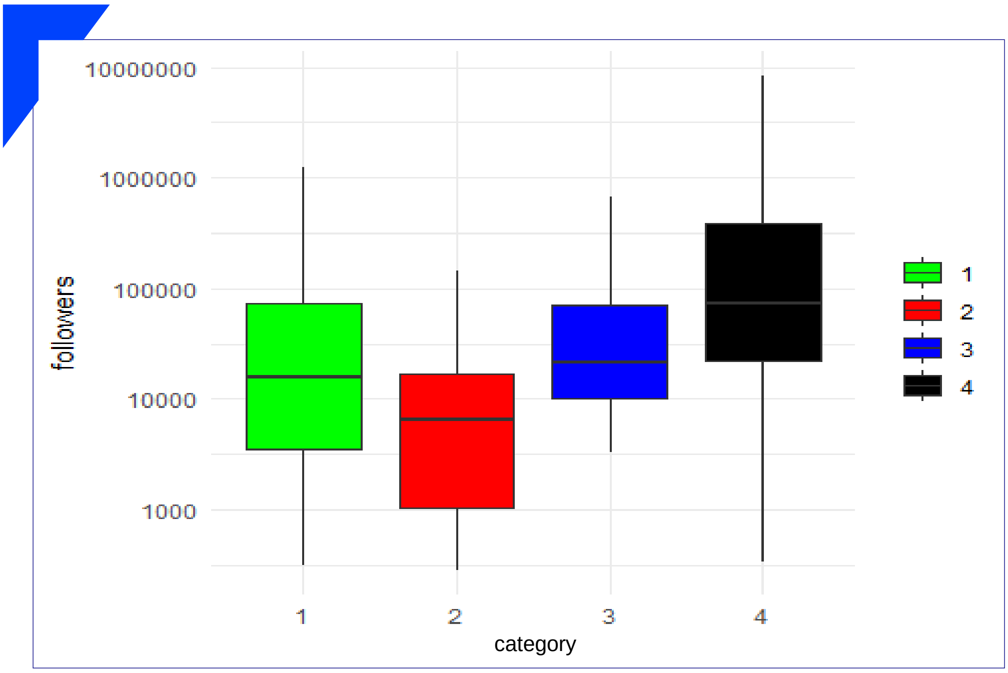


[28] Specifically, for statistics on the median number of ReTweets and the median number of favourites, this involves calculating the mean average number of ReTweets per Tweet, or number of favourites for each account, and then calculating the median of these mean averages.

Accounts classified as category 4 show an even greater dispersion, with between 334 and over 8 million subscribers. A significant number of category 4 accounts have a very significantly higher number of subscribers than the largest category 2 account. There are 59 accounts identified as reliable with more than 150,000 subscribers and 26 accounts with more than 1,000,000 subscribers.

Category 3 accounts are in an intermediate position. Some of them have a substantially higher number of subscribers than category 2 accounts, but without competing with the audiences of the largest category 4 accounts.

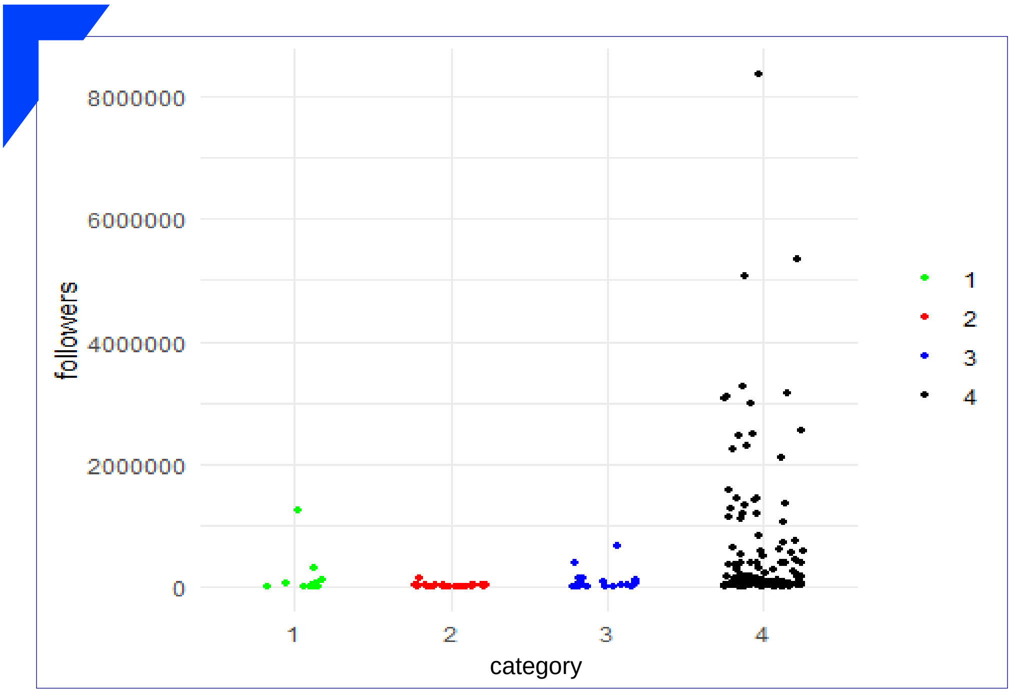
The following graph shows the dispersion of the number of subscribers for each of the account categories. Note that the graph uses a logarithmic scale that compresses the differences.



Reading: for each category, 25% of the accounts are below the bottom edge of the rectangle and 25% above the top edge. The horizontal line in the rectangle represents the median. The lines also extend up to the minimum and maximum values but excluding outliers[29]*.*

The following graph shows the same dispersion without using a logarithmic scale and representing all the accounts. This representation also makes it possible to illustrate the disproportion between the audience for category 2 accounts and the audience for category 4 accounts.

[29] Points are considered to be outliers if they are located at a distance from the edge of the rectangle that exceeds 1.5 times the height of the rectangle.



The results in this section present similarities with findings highlighted in previous studies. For example, the analysis of Fletcher et al. (2018)[30] of audience data for websites surveyed by the "Décodex" showed that the audience for websites categorised as sharing fake news is small and that these sites were viewed by less than 1% of French Internet users. In particular, they were significantly less widely accessed than the sites of the main French newspapers (*Le Monde* and *Le Figaro*) or the *France Info* site.

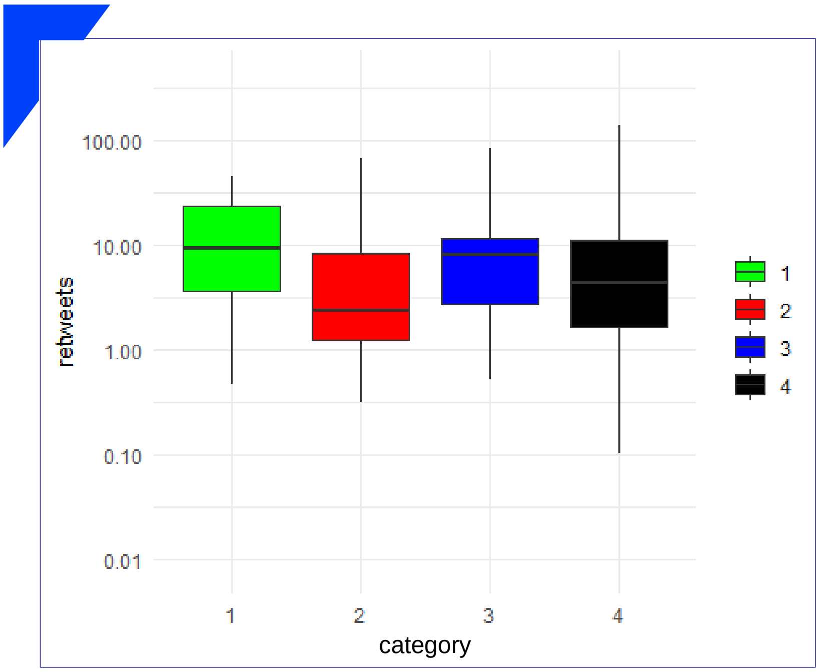
4.2.2. Unreliable accounts show significant numbers of ReTweets, especially in relation to their number of subscribers

An analysis of the number of subscribers to an account only provides a partial view of its audience and visibility on Twitter. When a user follows an account, the Tweets published by that account are indeed displayed in their timeline. However, due to recent modifications to the Twitter algorithm (see description above), it is not known how visible the Tweet is in the timeline. In addition, the degree of attention paid by the user to a Tweet appearing in their timeline is also uncertain. Last but not least, the number of subscribers to an account gives only a partial view of the visibility of the Tweets published by that account. Tweets can indeed be visible beyond the subscribers of the account that published the Tweet if they are the subject of ReTweets. In such a case, subscribers to accounts that have carried out ReTweets will also see the Tweet appear in their timeline. In other words, an account audience is impacted not only by its number of subscribers but also by its ability to be picked up as ReTweets by third-party accounts.

From the tables presented above, it appears that Tweets issued by category 4 accounts attract fewer ReTweets than category 1, 2 and 3 accounts. If, on the other hand, the analysis is carried out using the median average, category 4 accounts are still subject to fewer ReTweets than

[30] Fletcher, Richard; Cornia, Alessio; Graves, Lucas; Nielsen, Rasmus Kleis. *Measuring the reach of "fake news" and online disinformation in Europe*. 2018.

category 1 and 3 accounts but this time exceed category 2 accounts. The dispersion analysis is again interesting, and it appears that some category 2 accounts are on an equal footing with category 4 accounts:



This result is also consistent with the findings highlighted by previous studies. For example, Fletcher et al. (2018) showed that some accounts known to share fake news on Facebook were able to generate[31] levels of engagement similar to, or even higher than, the levels enjoyed by some of the larger media outlets categorised as trustworthy. *Le Monde* (2018)[32] also analysed the interactions on Facebook accounts indexed by “Décodex”: it shows that reliable accounts account for 72.3% of interactions on Facebook. *Le Monde* notes that while disinformation accounts have never overtaken reliable accounts, some disinformation accounts compete with well-established news accounts. For example, the two most popular disinformation accounts have generated more interactions than *Le Point* or *Libération*.

Category 2 accounts, therefore, have a low number of subscribers compared to category 4 accounts, but they are on an equal footing with the latter in terms of the number of ReTweets. They generate particularly high levels of virality in relation to their number of subscribers. In order to study this property in greater detail, the following table shows the number of ReTweets per 1,000 subscribers for the various account categories:

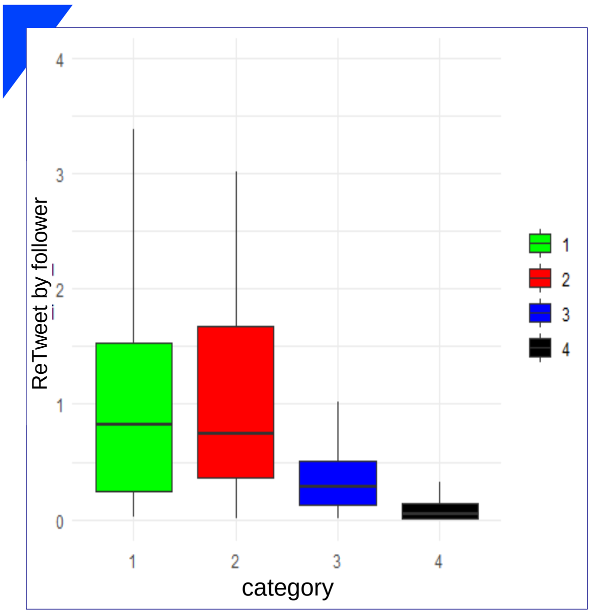
[31] On Facebook, engagement is calculated as the sum of the number of comments, shares and reactions to a post.

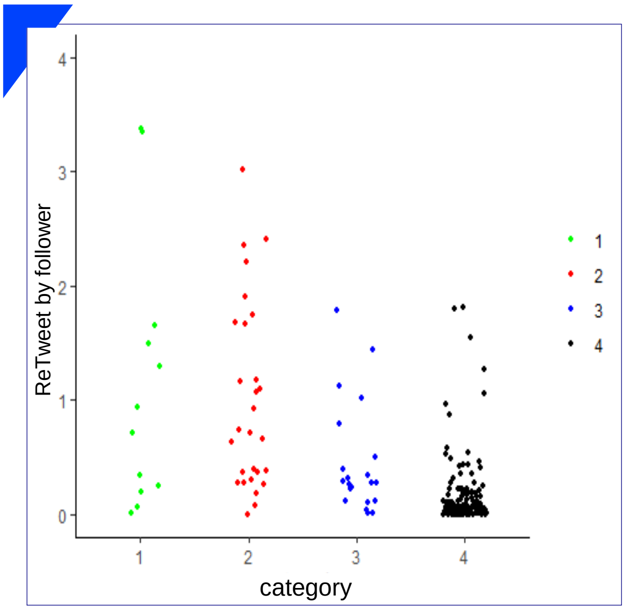
[32] *Le Monde. Les fausses informations circulent de moins en moins sur Facebook (The circulation of fake news is declining on Facebook).*

DÉCODEX CATEGORY	**MEAN NUMBER OF RETWEETS PER 1,000 SUBSCRIBERS**	**MEDIAN NUMBER OF RETWEETS PER 1,000 SUBSCRIBERS**
1	1.14	0.82
2	1.51	0.74
3	0.46	0.28
4	0.15	0.04

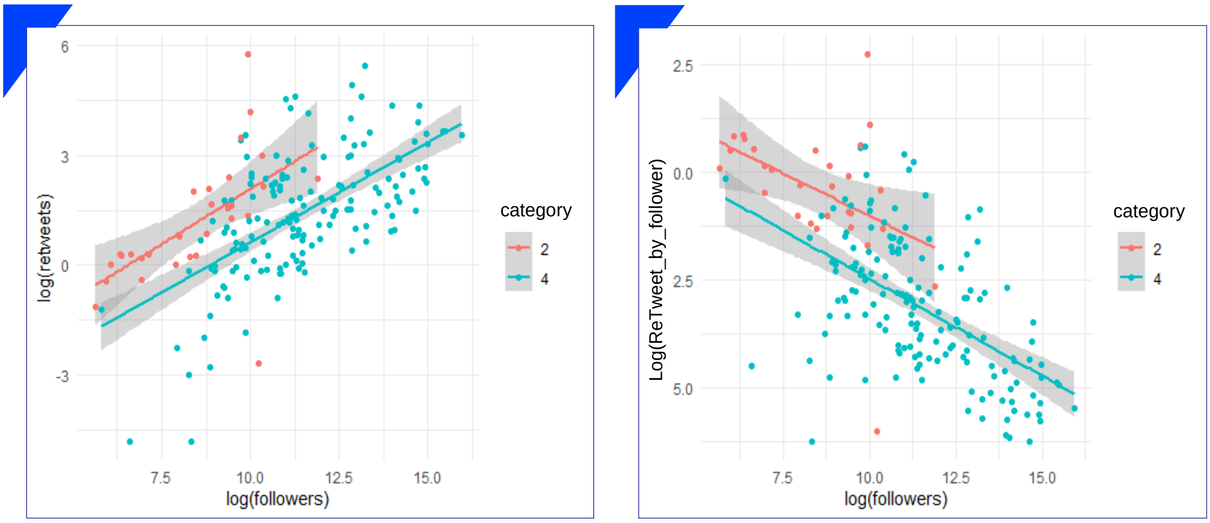
Reading: For a category 2 account, every 1,000 subscribers generate a mean average of 1.51 Tweets.

The mean or median number of ReTweets per Tweet and per subscriber is thus much higher for category 1, 2 and 3 accounts than for category 4 accounts that are considered reliable. Subscribers to accounts known to share fake news therefore have a much higher propensity to contribute to the dissemination of Tweets from these accounts. Dispersion analyses confirm these results:





The graph below also shows the relationship between the mean number of ReTweets for an account and the number of subscribers to the account. This analysis shows that the number of ReTweets is an increasing function of the number of subscribers, and that the growth in the number of ReTweets with the number of subscribers appears to be the same for category 2 and category 4 accounts. However, category 2 accounts have a higher number of ReTweets for the same number of subscribers than category 4 accounts. It is also interesting to observe that the number of ReTweets per subscriber is a decreasing function of the number of subscribers:



Reading: The graph on the left shows that an increasing number of subscribers is accompanied by an increasing number of ReTweets, both for category 2 and category 4 accounts. With equivalent numbers of subscribers, however, category 2 accounts have more ReTweets than category 4 accounts. The graph on the right shows that the number of ReTweets per subscriber decreases as the number of subscribers increases. With equivalent numbers of subscribers, the number of ReTweets per subscriber in category 2 accounts is higher than in category 4 accounts. The axes of the two graphs are also expressed in logarithmic form to improve their legibility, but this does not affect their reading.

This observation could be explained by the fact that an account that is followed by a large number of subscribers will attract a generalist audience whose proportion of individuals personally involved in the topics shared by the account is relatively lower. The total population of this account will therefore be less engaged and will therefore have a lower propensity to ReTweet the published content.

Summary:

The least reliable information accounts have a significantly lower number of subscribers than the majority of reliable information accounts. However, in terms of ReTweets, they are on an equal footing with reliable accounts. Subscribers to unreliable accounts have a much higher propensity to contribute to the dissemination of information on these accounts than subscribers to reliable accounts.

4.3. Analysis of the content of Tweets from accounts known to be conducive to the dissemination of fake news

The previous section highlighted the distinctive characteristics of accounts propagating fake news in terms of virality. To better understand this situation and the nature of the information shared by these accounts, the content of the most ReTweeted Tweets from these accounts (from among the last 1,000 Tweets collected[33]) was analysed by studying their formatting, the message, the presence of videos and photos, the presence of fake news[34], etc. The qualitative analysis of the content of the Tweets will be complemented by a quantitative analysis.

4.3.1. Qualitative analysis of the content of Tweets from category 2 accounts

The 10 most ReTweeted Tweets from among the last 1,000 Tweets published by each of Décodex's 29 category 2 accounts were systematically collected. 282 Tweets were collected using this methodology[35].

The study of these Tweets focused on analysing the drivers of user engagement, taken in this case to mean their propensity to share content (ReTweet). In order to do this, all 282 Tweets collected were firstly classified according to the topics addressed[36]. Secondly, a qualitative analysis of their content was carried out so as to identify common features in terms of information processing. Note that this study has been conducted on French-speaking content, and that the quoted terms have been translated so the reader can better grasp these qualitative analysis findings.

4.3.1.1. Analysis of the topics addressed

The 282 most ReTweeted Tweets for each of the 29 accounts studied were classified according to the main topic they addressed, with particular attention paid to references to false or misleading information[37]. Data collection for this part of the study alone preceded the emergence of the new coronavirus and the health crisis it triggered[38].

[33] The data was collected on 13 September 2019. The analysis therefore covers the last 1,000 Tweets published by each account at that date.

[34] In order to carry out its analysis, the *Conseil* did not itself attempt to categorise accounts, but used the unmodified categorisation from *Le Monde*'s “Décodex” tool (see section 3.2.). This system for categorising information sources according to their reliability is provided by *Le Monde* at the following address:
https://www.lemonde.fr/webservice/decodex/updates

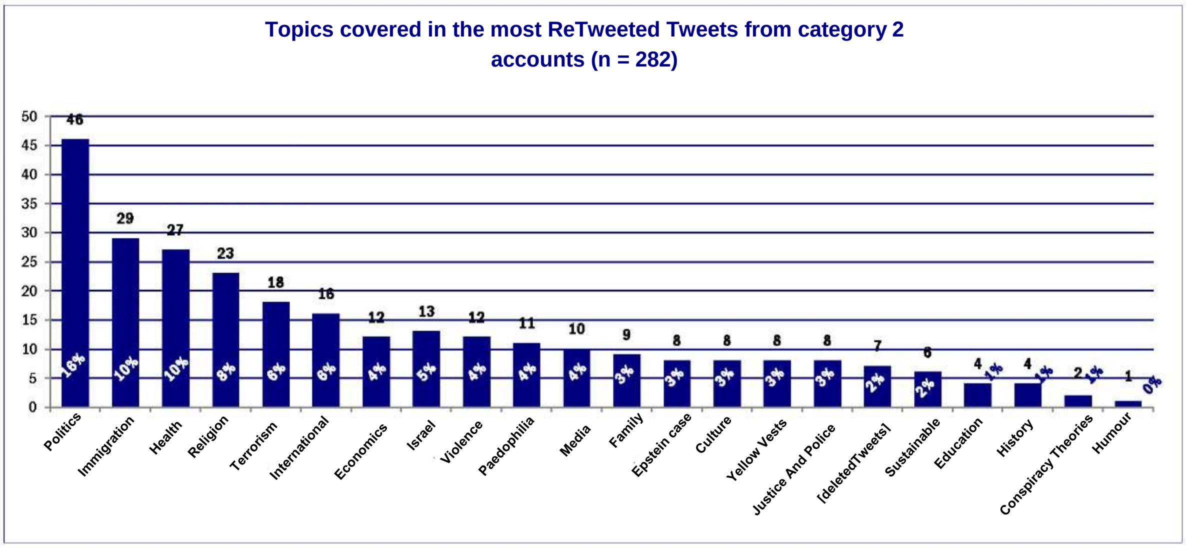
[35] As some accounts have a limited audience, it was sometimes not possible to collect these 10 ReTweeted Tweets for each account. For example, in one case, only 2 Tweets were ReTweeted and the 8 others collected did not generate any ReTweets.

[36] Each Tweet was counted only once, according to the main topic it addressed. This method may have limitations in that some Tweets may address several topics. However, the trends and proportions highlighted remain valid given the significant differences between the topics.

[37] This qualification work was performed not by this study but by a monitoring of the topics covered by journalists specialising in fact-checking.

[38] The analyses in the following section will, however, include elements relating to fake news concerning the coronavirus epidemic.

Their content sometimes combines several topics: for example, one publication addresses the potential health problems that could be suffered by people of unlawful immigration status. This was classified under the "Immigration" topic, with the "Health" topic representing a type of information processing. In addition, 7 collected Tweets had become inaccessible at the time of the study[39].



The thematic analysis will focus on **topics with the highest number of Tweets and topics which jointly cover 50% or more of the total number of studied Tweets**. **These are "Politics" (46 Tweets), "Immigration" (29), "Health" (27), "Religion" (24) and "Terrorism" (18), which total 51.1% of the corpus**.

- **Politics (46 Tweets)**

Of the 46 Tweets in this category, 34 mention political parties or personalities by name. The President of the Republic, Emmanuel Macron, is the most cited, with 9 occurrences. He is also the subject of the most ReTweeted Tweet (2,919 ReTweets), a video from an LCI programme showing a columnist criticising him. Several personalities close to the President are also the subject of posts from these accounts, such as the national delegate of Macron's *La République En Marche!* party or a former minister and member of the European Parliament, criticising their actions or sharing information about controversies concerning them in each case.

Political figures on both the left and right of the political spectrum also emerge from this corpus of Tweets, as do an environmental activist and structures linked to the far-right identitarian movement. The publications concerning them are in some cases critical, or in other cases more informative, advertising an event organised by these parties (3) or the start of a campaign for the municipal elections (1).

Finally, 6 publications refer to the political world as a whole, focusing on corruption or the existence of a dictatorship in France.

[39] This inaccessibility may be the result of several factors: the author of the account may have deleted it, making it impossible to view his or her account, either at his/her own initiative or at Twitter's request; for example, following reports from other users. Finally, when a Tweet is a response to another Tweet, if the initial Tweet is deleted for one of the reasons mentioned, the ReTweet also disappears.

- **Immigration (29 Tweets)**

The 29 Tweets dealing with the subject of immigration refer to acts committed by individuals characterised as foreigners (11), or to the implementation of anti-migration policies outside of France (9). One of these Tweets also seems to link health risks with undocumented immigration. The choice of terminology about irregular migrants is notable: the three emotionally loaded nouns "*clandestin*", "*clandos*" (both from "clandestine") and *"passeur"* (smuggler) are used in as many publications.

Within this category, the two most shared Tweets are two videos: the first shows supporters presented as Algerians *"surround[ing] the car of two young French women"* and apparently stealing their belongings while emphasising the government's silence on the matter (2,805 ReTweets). The second is that of an exchange between journalist Éric Zemmour and politician Éric Dupond-Moretti in which the two men clash on immigration (2,244 ReTweets). The third most ReTweeted publication attracted a mere 162 ReTweets.

- **Health (27 Tweets)**

One of these Tweets mirrors the Tweet conflating immigration and health risks, this time linking *"exotic immigration"* with the rise of diseases that had been thought to be extinct in France.

The most popular, however, is a publication about "*l'emprise de l'empire vaccinal sur nos enfants*" *(the vaccine empire's grip on our children)*, which has been ReTweeted 307 times. In fact, most of these Tweets address the issue of vaccines (14) and share a scepticism centred on the marketing of vaccines containing aluminium.

We also note that 2 Tweets connect food and health issues, accusing products—and particularly American products (isoglucose, Coca-Cola)—of being responsible for certain illnesses. EDF's Linky smart meter, which has been the subject of numerous communications in recent months from associations representing electro-sensitive people, appears in just one single Tweet.

- **Religion (23 Tweets)**

This topic brings together publications that approach current events from a religious perspective or use religion as a factor in explaining certain issues. We note that most references are to Islam (12 Tweets), followed by Judaism (11) and Catholicism (7). In addition, 8 Tweets jointly mention two or three of these religions.

Two Islam-related posts link the religion to the destruction of Catholic places of worship, whether criminal or not, while two others attempt to demonstrate the existence of a supposed incompatibility between Islam and European societies.

Among the publications mentioning the Jewish religion or people of Jewish origin (11 Tweets), the *Conseil représentatif des institutions juives de France* (CRIF) Jewish council is mentioned (and criticised) 5 times, while 1 Tweet shares a link on the Holocaust. Finally, 3 Tweets mention the words *"synagogue", "Kabbalist*", *"Jew-Zionist" and* even *the "tricolour star of David"* which was supposedly worn by the French President at a CRIF dinner for the purpose (according to their authors) of showing a form of benevolence on the part of the authorities towards Judaism, to the detriment of Catholicism.

- **Terrorism (18 Tweets)**

This topic focuses on terrorism of Islamist origin, particularly where linked to Daech (ISIS). Several of them criticise a form of terrorist influence on political decisions: for example, a former President of the Republic was reportedly seen *"pos[ing] with a Daech terrorist"*. 5 Tweets go further and attempt to demonstrate the involvement of public authorities in the development of Islamist terrorism. For example, the leader of the Daech terrorist organisation is allegedly a *"CIA agent"*.

Finally, 3 identical Tweets published by the same account address the topic of terrorism from the perspective of France's solidarity policy. They report information that a jihadist allegedly returned from Syria claiming for the *"payment of social security benefits"* that he had missed during his stay outside France.

4.3.2. Analysis of information processing

For this analysis, the most shared post was extracted for each account if it had less than 100 shares[40], and the two most shared posts if they had 100 or more shares each (n = 37[41]). The 37 Tweets studied were subjected to both a semantic and visual analysis, taking into account both the message of the Tweet and the various non-textual contents (links, images, videos). As a result, all the Tweets collected contain text with another type of content (links, images, videos, etc.), with the exception of one that contains only text. These collected Tweets can be analysed as pursuing **4 types of objectives: to inform (15 Tweets), to create doubt (5), to provoke criticism (11) or even enlist support (6) according to an increasing gradation, from neutrality through to a form of emotion,** such as fear or anger[42]**.**

The 15 publications that seek to inform the user operate using several drivers. First of all, they adopt a neutral, very factual tone. In some cases, verbatim is used in Tweets, identifying the authors by name. They also summon references to people whose functions or professions appear to be used as a source of legitimacy. This is the case of the *"President of the Serbian Republican of Bosnia"*, the *"Austrian Chancellor Sebastian Kurz"* and even a *"dentist"* whose actions or words are shared. Two of these publications use content linked to media or press agencies (France 3 and Associated Press). Furthermore, the quality of some of the videos or photos used suggests that, even though they were not taken by professionals, they are not amateurish and show a certain commitment to quality or a conscious choice of media.

[40] The aim here is to present the analysis of Tweets that achieve the highest numbers of shares in order to understand the forces that drive virality. Some accounts have a low popularity and are therefore little ReTweeted; however, it is still instructive to present the most shared Tweets from these accounts in order to maintain the diversity of the accounts studied. The sample of Tweets selected, however, assigns a larger share to the most shared accounts, with two Tweets for them compared to one for the least shared accounts (that is, accounts that have never had a post shared more than 100 times).

[41]42 Tweets initially met this criterion. One was deleted because it was ReTweeted an equal number of times from one single account to another, with both being the most ReTweeted, but still with less than 100 ReTweets. For this reason, the more recent of the two was chosen. Four more Tweets were not included in this corpus because they have become inactive since the time of data collection.

[42] Signs such as punctuation ("??? ", "!!! "), the use of capitalised words, insults, derogatory verbs and logical sentence constructions are among the methods used in this analysis to produce this categorisation. However, this categorisation does not claim to be authoritative in terms of analysing the intentions of the authors of such content or the truthfulness of their statements, but seeks to shed light on the obvious motives for posting to their audiences because, as already mentioned, *"most of the time, manipulators do not adopt a position in relation to the truth: they simply seek to achieve an effect"* (J. -B. Jeangène Vilmer et al, 2018).

6 Tweets can be considered as having the purpose of prompting users to ask questions. These are close to conspiracy theories in both form and substance. They use interrogative forms and establish links between subjects without providing any other form of proof than content whose origin is not directly identifiable. The most commonly used approach here is the enumeration of facts which, when taken together, supposedly provide grounds for questioning official information or assertions. The non-text content corroborating these remarks comes from sites linked to the initial Twitter account (4), from another medium for which *Le Monde*'s "Décodex" urges caution with regard to the veracity of the information it shares, or from an unidentifiable source (a truncated screen photo).

The 10 Tweets with a critical focus adopt a vocabulary and tone consistent with a judgemental position. They use expressions such as *"as if by magic"*, *"arrogant"* and *"[...] has for once been useful"*. In contrast to Tweets looking to inform, clearly identified sources of news media are used to better support these posts: for example, this video from *franceinfo*, in which the remarks made by Russian President Vladimir Putin at a joint press conference with French President Emmanuel Macron are presented as having been *"watered down"* by a partisan translator (*"a work placement trainee with an LREM party membership card"*).

Finally, the 6 Tweets that seek to enlist support use a vocabulary and expressions from a very low linguistic register ((*« #fautrigoler »* "*#yourehavingalaugh*", *« ces ordures d'islamistes » "these Islamist scum"* or *« préparons résistance » "we need resist this"*). Some use cartoons or links to petitions.

Almost all the studied posts thus tend to seek to **establish the legitimacy** of the information being shared. For example, 15 Tweets refer to content that is external to them, or at least difficult to identify as self-created (**validation by another source)** while 22 other Tweets link to content that can be directly or indirectly traced back to them, such as a site linked to the Twitter account. In the latter case, these Tweets act as **information relays** aimed at encouraging readers to view this internally produced content, which is for the most part presented in the form of articles. The methods differ in the two cases, but the objective here seems to be to add information in support of a news item or statement of position with the idea of giving it more substance, regardless of whether the final intention is to inform, to arouse doubt, to criticise or to create a form of cohesion around the same ideas.

In the case of the 15 Tweets referring to external content, 5 share the URL of sites that are not at first glance linked to the original Twitter account, but which are all listed by the "Décodex" as sites requiring caution. In addition, anoother shares content previously published on Twitter by an account considered by the Décodex either as "grey" or as unclassified, but for which a warning message is nonetheless offered[43].

[43] In its information sources search engine, the Décodex offers several labelling colours. A site labelled "red" is accompanied by the warning *"This site disseminates a significant volume of misleading information or articles. Remain alert and cross-check against other more reliable sources. If possible, go back to the source of the information"* (1 Tweet concerned here)*; an* "orange" site features the warning, *"exercise care and cross-check against other sources. If possible, go back to the source of the information" (4 Tweets*). Finally, unclassified content is "grey" and the following warning is given: "*Do not hesitate to confirm the information by cross-checking against other sources or going back to its origin. ". Le Monde. Décodex : vérification de sources d'informations, pages Facebook et chaînes YouTube (Décodex: Checking news sources, Facebook pages and YouTube channels).* URL:
https://www.lemonde.fr/verification/

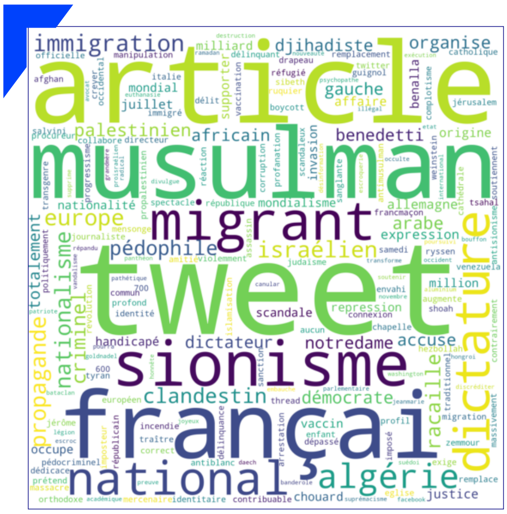
4.3.3. Quantitative analysis of the content of Tweets from category 2 accounts

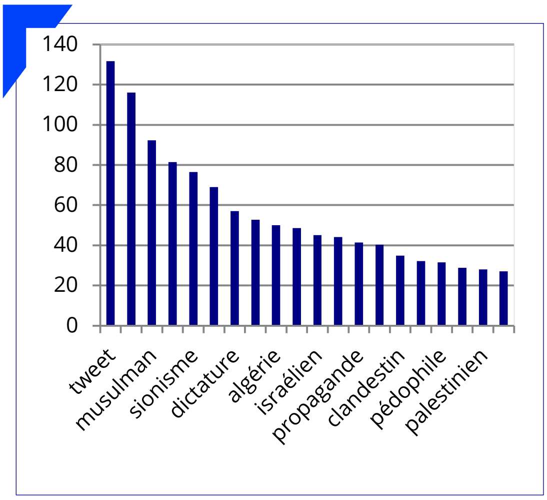
The quantitative analysis of the content of Tweets for category 2 and 4 accounts was performed on the 25% of Tweets with the highest number of ReTweets for each account (among the most recent 1,000 Tweets published); in order to draw lessons on topics likely to generate significant virality. First, the analysis consisted in removing non-significant terms (the most common words in the French language or non-alphabetic words). Similar words (but with different spellings) were also grouped[44]. The analysis then aimed to compare the frequency of occurrence of words in the two corpora and to identify terms that were over-represented in category 2 accounts as compared to category 4 accounts[45].

There are two methods of assessing the extent of over-representation. The first is to calculate the difference between the actual occurrence of the term in the corpus of category 2 accounts and the expected occurrence if the frequency was the same as for category 4 accounts. It has the advantage of making it easier to highlight over-represented terms that are frequently used by category 2 accounts. The second method is to study not the absolute difference but the ratio. For example, it highlights terms that are not used often in category 2 accounts, but which are nevertheless used significantly more often than in category 4 accounts.

- **Method of difference between actual and expected frequency**

The following word cloud represents the most over-represented words in category 2 accounts compared to category 4 accounts (based on the size of the difference between actual occurrence and expected occurrence), while the graph shows the 20 most over-represented terms[46]:





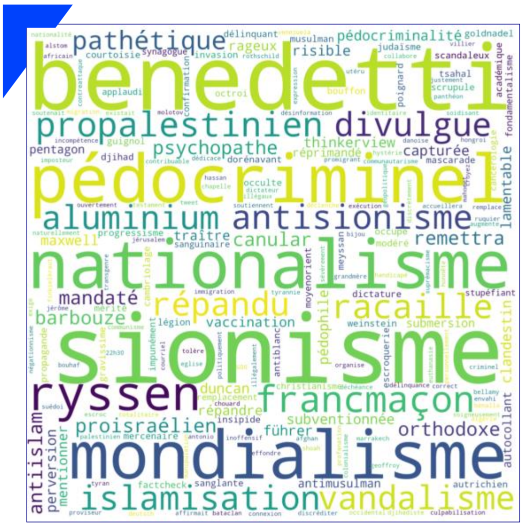
[44] The method used is Levenshtein distance, which is a mathematical distance that aims to measure the difference between two strings of characters. Many words express the same idea but are spelled in different ways (plurals, spelling mistakes, spaces, etc.). In order to group together all identical or nearly identical words, it was decided to use a Levenshtein distance of 1, i.e. to treat all words with up to one letter of difference as equivalent. The numbers of occurrences of words sharing a very strong proximity are totalled up in this way, allowing certain topics to be highlighted more accurately.

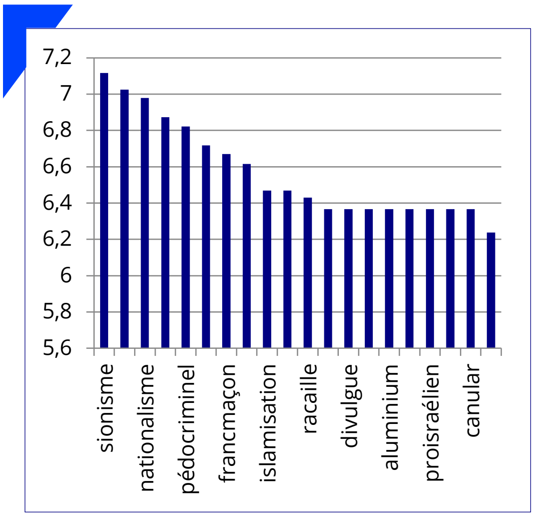
[45] The significance of over-representation was estimated using the G² log-likelihood method (Dunning, Ted. 1993. *Accurate Methods for the Statistics of Surprise and Coincidence.* Computational Linguistics, Volume 19, number 1, pp. 61-74). Only statistically significant differences of 0.01% were retained.

[46] For the purposes of creating the cloud, words with a similar spelling were assimilated. For example, the word "françai" represents both occurrences of "françai" and "français" ("French").

- **Method of the ratio of actual frequency to expected frequency**

The following word cloud represents the most over-represented words in category 2 accounts compared to category 4 accounts (according to the importance of the ratio between actual occurrence and expected occurrence), while the graph shows the 20 most over-represented terms:





The two methods of analysis are largely convergent and show that the most ReTweeted Tweets from category 2 accounts reveal a more marked use of terms related to delinquency, immigration, Israel and Palestine, paedophilia, Islam or freemasonry than do category 4 accounts. These accounts also make more use of terms such as "dictatorship", "propaganda" or "reveals" and therefore seem to position themselves as providing alternative information, concealed by official bodies.

Summary:

Category 2 accounts focus strongly on current issues and divisive topics in their posts: Tweets dealing with politics, immigration, health, religion and terrorism account for more than half of the corpus selected for the qualitative analysis. The tone used by these accounts is focused mainly on informing their subscribers and criticising the people or institutions mentioned. The use of non-textual content (images, videos, links, etc.) also seems to highlight a strategy for establishing the legitimacy of the published content, based mainly on sites linked to these accounts but also on sources from traditional media. The quantitative analysis shows the over-representation of terms related to delinquency, immigration, Israel and Palestine, paedophilia, Islam and freemasonry.

4.4. Analyses do not reveal any echo chamber phenomenon on Twitter

This section looks at the possible existence of an "echo chamber" phenomenon on Twitter. The "echo chamber" concept refers here to the risk that an individual exposed to unreliable sources of information is unlikely to be exposed to reliable sources of information. The study will examine the extent to which subscribers to category 2 (red) accounts are likely to be subscribers to category 4 (reliable) accounts. This methodology necessarily entails simplification, for several reasons. Firstly, it does not take into account the exposure of subscribers to category 2 accounts to other reliable sources of information outside Twitter. Secondly, being a subscriber to a category 4 account does not necessarily imply that the user will pay a high level of attention to the Tweets published by this account; for example, reading the articles to which the links refer in order to inform his or her opinion. Lastly, it does not take into account the fact that subscribers to category 2 accounts are also potentially exposed to Tweets from category 4 accounts even if they do not follow this type of account (in particular because they follow people who may have ReTweeted these accounts).

The echo chamber definition used here does not necessarily correspond to the definition examined in the literature, which has mostly studied the existence of ideological echo chambers (for example, by studying whether interactions between users are exclusively between members of the same political side). This work has confirmed the presence of echo chambers for some, while others have refuted them[47]. In the United States, several studies have explored the link between echo chamber phenomena and political polarisation, concluding that the former is a result of the latter and not the other way around[48].

In this study, a user who follows a category 2 (red) account and a category 4 (reliable) account with similar or close political positioning will not be considered as being locked in an echo chamber.

The analysis was conducted on a sample of 30 randomly selected users for each category 2 account. For each of these users, the list of all the accounts monitored was collected and, among this list, the accounts labelled by the "Décodex" were identified. While the analysis is based on a small sample of nearly 800 Twitter accounts, the results obtained are strong enough to suggest that the sample size is not likely to impact the conclusion.

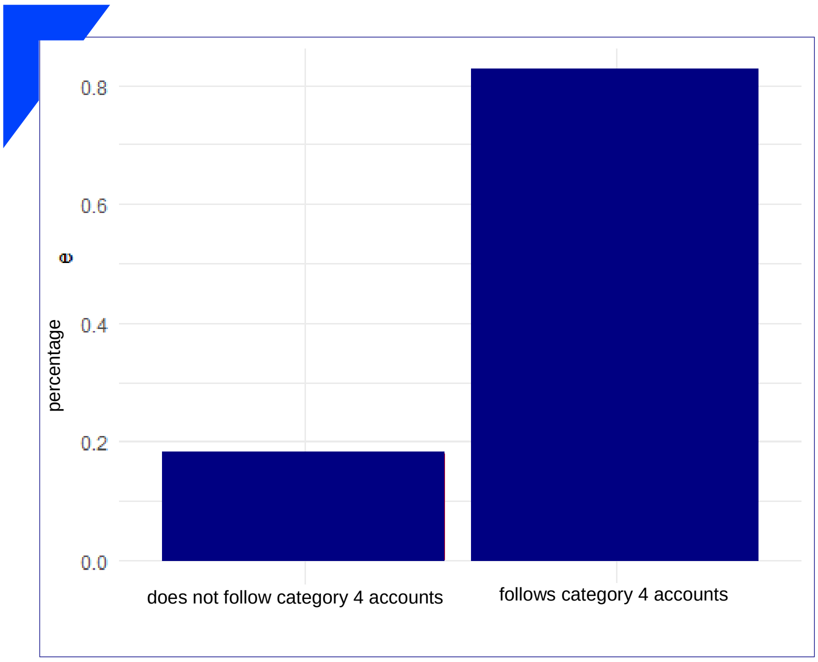
[47] Council of Europe. *Information disorder: Toward an interdisciplinary framework for research and policy making.* 09/2017." "

[48] Andrew Guess, Brendan Nyhan, Jason Reifler. *Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign*. 09/01/2018.

The median number of accounts followed by subscribers to category 2 accounts is summarised in the following table:

CATEGORY 1	CATEGORY 2	CATEGORY 3	CATEGORY 4
0	2	1	6

These subscribers follow a higher number of category 4 accounts than category 2 accounts. We can also examine the proportion of subscribers to category 2 accounts who do not follow any category 4 account:



More than 80% of subscribers to category 2 accounts, therefore, follow at least one category 4 account. And they follow more category 4 accounts than category 1, 2 and 3 accounts combined. This result corroborates the finding that the number of subscribers to category 4 accounts is much higher than the number of subscribers to other types of accounts.

These results also once again confirm the findings presented by Fletcher et al. (2018), which show that Internet users who view sites that are identified by the “Décodex” as sharing fake news also consult websites classified as credible. This observation, made using websites, is therefore also verified on Twitter.

In addition, the table below represents the 10 accounts that are the most followed[49] by the sample of category 2 subscribers studied:

ACCOUNT	CATEGORY	NUMBER OF USERS IN THE SAMPLE FOLLOWING THIS ACCOUNT
Le Monde	4	324
Mediapart	4	285
AFP	4	281
Le Figaro	4	258
BFMTV	4	217
Russia Today (French version)	3	211
Valeurs actuelles	3	210
TV Libertés	3	210
Franceinfo	4	208
Le Parisien-Aujourd'hui en France	4	204

As seen above, the median number of category 2 sites followed by the sample is 2, while the median number of category 4 sites followed is 6. The table below shows the Décodex accounts followed by the same identified subscribers who follow 2 accounts classified as category 2 and between 5 and 7 accounts classified as category 4. These subscribers are not necessarily representative of all the subscribers studied, but they do illustrate some examples of subscriber behaviour close to the median for subscribers to category 2 accounts. Each line of the table presents all the accounts categorised by the Décodex that are followed by a given follower.

[49] Among the accounts labelled by the Décodex .

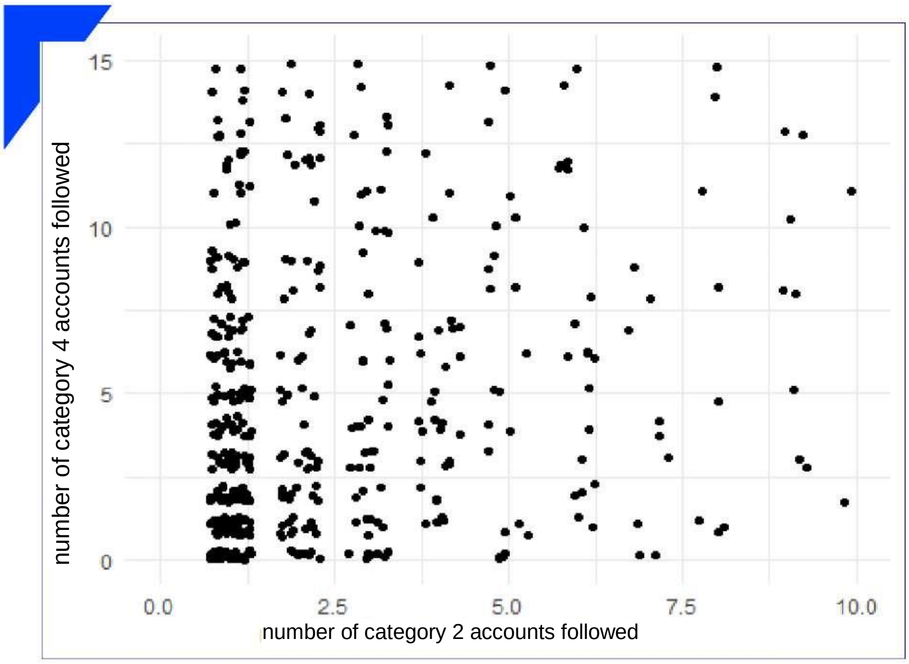
ACCOUNTS FOLLOWED BY A FEW"REPRESENTATIVE” SUBSCRIBERS TO CATEGORY 2 ACCOUNTS
[“[‘Al Kanz’]”, ‘[“Mouv'”]’, “[‘Tariq Ramadan’]”, “[‘Dieudosphere (Dieudonné)’]”, “[‘Jovanovic’]”, “[‘20 Minutes’]”, “[‘Le Monde’]”]
[“[‘The Onion’]”, “[‘The Economist’]”, “[‘BuzzFeed’]”, “[‘The New York Times’]”, “[‘BBC’]”, “[‘Tprincedelamour’]”, “[‘AE911truth’]”, “[‘France 24’]”, “[‘Counterpunch’]”]
[“[‘Panamza’]”, “[‘Egalité et Réconciliation’]”, ‘[“L'Equipe”]’, “[‘AFP’]”, ‘[“L'Obs”]’, ‘[“L'Express”]’, “[‘Libération’]”, “[‘Le Figaro’]”, “[‘Le Parisien-Aujourd’hui en France’]”]
[“[‘DailyMail’]”, “[‘Le Soir’]”, “[‘Boulevard Voltaire’]”, “[‘20 Minutes’]”, “[‘Hollande Démission’]”, “[‘Paris Match’]”, “[‘Valeurs actuelles’]”, “[‘Jeune Nation’]”, “[‘Canal+’]”, “[‘TF1’]”]
[“[‘Panamza’]”, “[‘Le Monde diplomatique’]”, “[‘Réseau International’]”, “[‘Maître Eolas’]”, “[‘Acrimed’]”, “[‘Mediapart’]”, ‘[“L'humour de droite”]’, “[‘Oumma’]”, “[‘Tariq Ramadan’]”, “[‘Al Kanz’]”]
[“[‘Panamza’]”, “[‘Valeurs actuelles’]”, ‘[“L'Obs”]’, “[‘Dieudosphere (Dieudonné)’]”, “[‘Mediapart’]”, “[‘Rue89’]”, “[‘Tariq Ramadan’]”, “[‘Libération’]”]
[“[‘Réinformation RC’]”, “[‘TV Libertés’]”, “[‘LesObservateurs.ch’]”, “[‘Russia Today (version française)’]”, “[‘Melty’]”, “[‘Jean-Marc Morandini’]”, “[‘Valeurs actuelles’]”, “[‘Libération’]”, “[‘Fdesouche’]”, “[‘Huffington Post’]”, “[‘BFMTV’]”, “[‘AFP’]”]
[“[‘Russia Today (version française)’]”, “[‘Jovanovic’]”, “[‘Agence Info Libre’]”, “[‘Etienne Chouard’]”, ‘[“L'Equipe”]’, “[‘Dieudosphere (Dieudonné)’]”, “[‘Libération’]”, “[‘Le Monde’]”, “[‘Le Figaro’]”, “[‘Rue89’]”, “[‘Maître Eolas’]”]
[“[‘Konbini’]”, “[‘Mediapart’]”, “[‘Korben’]”, “[‘Dieudosphere (Dieudonné)’]”, “[‘Egalité et Réconciliation’]”, “[‘Al Kanz’]”, “[‘Paris-Normandie’]”, “[‘France 2’]”]
[“[‘BFMTV’]”, “[‘i24 News’]”, “[‘France 2’]”, “[‘JSS News’]”, “[‘Dreuz Info’]”, “[‘Europe Israël’]”, “[‘Nice-Matin’]”, “[‘Libération’]”, “[‘Ouest-France’]”, “[‘Valeurs actuelles’]”, “[‘Le Monde’]”]

Previous analyses suggest that we should not overstate the importance of an “echo chamber” phenomenon, which would take the form of impermeability to media sources that are considered to be reliable. However, although these echo chamber phenomena do not represent the majority, they are nevertheless likely to exist on the platform. It was thus seen that just under 20% of subscribers to category 2 accounts do not follow any category 4 account.

The table below shows the subscription behaviour of some users who follow more than 3 accounts that are prone to disinformation and do not follow any news account categorised as reliable:

ACCOUNTS FOLLOWED BY A FEW CATEGORY 2 ACCOUNT SUBSCRIBERS REPRESENTING THE “ECHO CHAMBER” PHENOMENON
[“[‘Réseau International’]”, “[‘Panamza’]”, “[‘Médias Presse Info’]”, “[‘Jovanovic’]”, “[‘Valeurs actuelles’]”, “[‘LesObservateurs.ch’]”, “[‘Boris Le Lay’]”, “[‘Dieudosphere (Dieudonné)’]”, “[‘TV Libertés’]”]
[“[‘Valeurs actuelles’]”, “[‘Boulevard Voltaire’]”, “[‘Boris Le Lay’]”, “[‘Breizh Info’]”, “[‘TV Libertés’]”, “[‘Le Gorafi’]”, “[‘Jeune Nation’]”, “[‘Panamza’]”, “[‘Médias Presse Info’]”, “[‘Fdesouche’]”, “[‘Le site du Professeur Henri Joyeux’]”, “[‘Sputniknews’]”, “[‘Dieudosphere (Dieudonné)’]”, “[‘Egalité et Réconciliation’]”]
[“[‘Antipresse’]”, “[‘Réinformation RC’]”, “[‘OJIM’]”, “[‘Fdesouche’]”, “[‘Egalité et Réconciliation’]”, “[‘TV Libertés’]”, “[‘Boris Le Lay’]”, “[‘Agence Info Libre’]”, “[‘Diktacratie’]”]
[“[‘Médias Presse Info’]”, “[‘Réinformation RC’]”, “[‘Paris Vox’]”, “[‘Jeune Nation’]”, “[‘Novopress’]”, “[‘Breizh Info’]”, “[‘Russia Today (version française)’]”, “[‘TV Libertés’]”, “[‘Jovanovic’]”, “[‘Etienne Chouard’]”, “[‘Agence Info Libre’]”, “[‘Sputniknews’]”, “[‘Egalité et Réconciliation’]”, “[‘Russia Today’]”]
[“[‘Panamza’]”, “[‘Diktacratie’]”, “[‘Etienne Chouard’]”, “[‘Jovanovic’]”, “[‘Agence Info Libre’]”, “[‘Egalité et Réconciliation’]”, “[‘Dieudosphere (Dieudonné)’]”]
[“[‘Hollande Démission’]”, “[‘Russia Today (version française)’]”, “[‘Europe Israël’]”, “[‘Riposte laïque’]”, “[‘Dreuz Info’]”, “[‘Russia Today’]”, “[‘Breitbart’]”, “[‘Sputniknews’]”, “[‘TV Libertés’]”, “[‘Réseau International’]”]
[“[‘Diktacratie’]”, “[‘Egalité et Réconciliation’]”, “[‘Dieudosphere (Dieudonné)’]”, “[‘Panamza’]”, “[‘Etienne Chouard’]”, “[‘Agence Info Libre’]”, “[‘Jovanovic’]”]
[“[‘Tprincedelamour’]”, “[‘Panamza’]”, “[‘Fawkes News’]”, “[‘Novopress’]”, “[‘Boulevard Voltaire’]”, “[‘Valeurs actuelles’]”, “[‘Riposte laïque’]”, “[‘Boris Le Lay’]”, “[‘TV Libertés’]”, “[‘Sputniknews’]”, “[‘Russia Today (version française)’]”, “[‘Dieudosphere (Dieudonné)’]”, “[‘Egalité et Réconciliation’]”, “[‘Etienne Chouard’]”, “[‘Agence Info Libre’]”, “[‘Jovanovic’]”, “[‘Fdesouche’]”]
[“[‘Boris Le Lay’]”, “[‘Hollande Démission’]”, “[‘Tprincedelamour’]”, ‘[“La gauche m'a tuer”]’]
[“[‘Breizh Info’]”, “[‘Le Gorafi’]”, “[‘Réinformation RC’]”, “[‘Russia Today (version française)’]”, “[‘Boris Le Lay’]”, “[‘Les News’]”, “[‘TV Libertés’]”, “[‘Riposte laïque’]”, “[‘OJIM’]”]

Finally, the following graph represents the number of category 4 accounts followed as a function of the number of category 2 accounts followed by the user. This analysis shows the diverse range of[50] behaviour, but also shows that the most frequent case is that of a user who follows only one category 2 account and a significantly higher number of category 4 accounts:



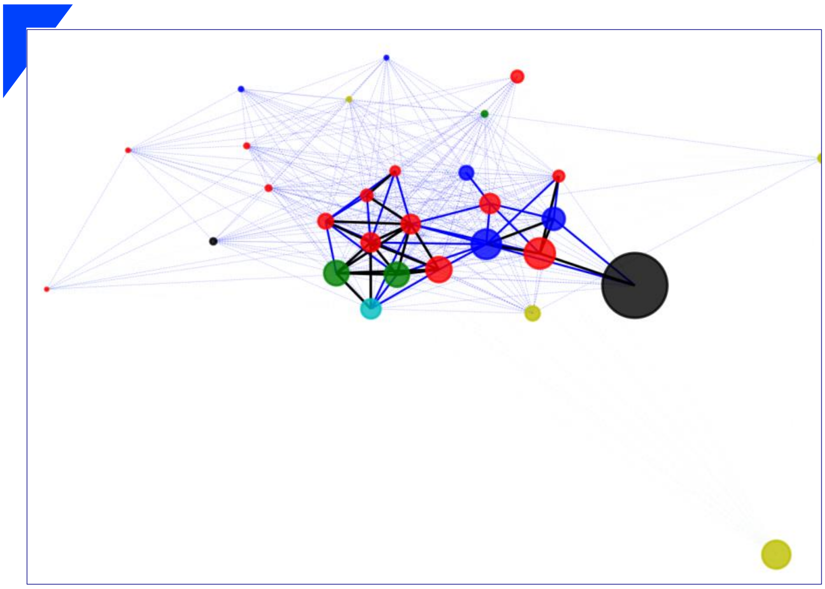
Summary:

Individuals who are exposed to sources known to share fake news, but do not at the same time also follow reliable accounts, are in the minority. On average, they even follow a majority of reliable accounts. Fewer than 20% of the subscribers to these accounts do not follow any reliable account. These individuals thus present subscription behaviour that does not present them with sources of information likely to correct fake news received elsewhere.

[50] Accounts that follow more than 15 category 4 accounts or more than 10 category 2 accounts are in a very small minority, and are not shown in order to improve graph legibility.

4.5. Analysis of the subscriber graph

The previous analysis examined the diversity of subscription behaviours among a sample of users and more specifically examined the propensity of subscribers to unreliable accounts to follow reliable accounts. This section also looks at the subscription behaviour of users by seeking to identify the existence of proximities between these different accounts. This analysis is based on a review of all subscribers to each category 2 account. It leads to a proposed analysis in the form of a graph or network showing the proximity between each category 2 account according to their proportion of common subscribers. The dots (nodes) in the graph represent category 2 accounts, and their size is proportional to the number of subscribers. The thickness of the links connecting the accounts is proportional to the number of common subscribers between the accounts. Using the descriptions employed by the Décodeurs to explain their choice of classification[51], subcategories of category 2 accounts have been created and are represented in the following graph by distinct colours[52]:



Extreme Right	Red
Anti-Semitic	Black
Health	Yellow
Conspiracy	Blue
Fake News	Green
Pro-Israel	Turquoise

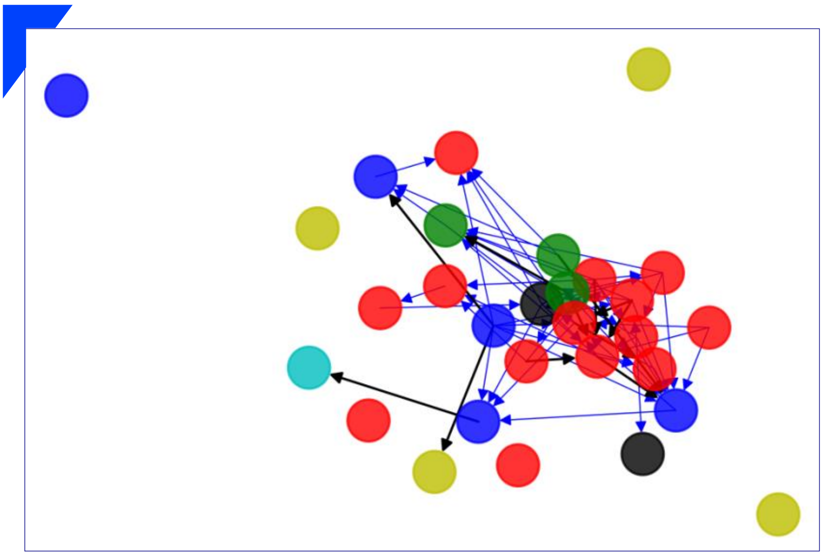
The graph shows that some accounts share a large number of common subscribers (accounts connected by thicker links). This is particularly the case for the Extreme Right themed accounts among themselves, but also between those accounts and accounts relaying fake news in general.

A second graph (using the same colour code) represents the subscription behaviour of the category 2 accounts themselves. Category 2 accounts are linked if the accounts follow one other (the arrow representing the direction of the subscription). This graph shows strong proximity

[51] *Le Monde. Décodex : vérification de sources d'informations, pages Facebook et chaînes YouTube (Décodex: Checking news sources, Facebook pages and YouTube channels).*

[52] To be totally exhaustive, six subcategories have been created: a category of accounts described as anti-Semitic or anti-Zionist; a category of right-wing extremist accounts; a category of accounts that promote "alternative" health (some of these accounts have an anti-vaccine stance); a category of pro-Israeli accounts; another category including accounts categorised as "conspiracy theorists"; and a final category relates to accounts that "relay" fake news without further clarification.

between some right-wing extremist accounts, accounts sharing fake news and conspiracy theorist accounts. Other accounts, particularly those belonging to the health topic, appear to be more autonomous:



Extreme Right	**Red**
Anti-Semitic	**Black**
Health	**Yellow**
Conspiracy	**Blue**
Fake News	**Green**
Pro-Israel	**Turquoise**

Summary:

Many category 2 accounts have a high degree of proximity: they have a significant proportion of common subscribers or they follow one another on Twitter.

4.5.1. Review of activity in fact-checkers' accounts

This section deals with a category of accounts that do not fall under the Décodex typology: fact-checking journalist accounts. This analysis does not claim to be exhaustive. It is based on the study of Twitter accounts from 20minFakeOff (*20 Minutes*), AFPFactuel (*Agence France-Presse*), Les Décodeurs (*Le Monde*), InfoIntoxF24 (*Les Observateurs de France 24*) and CheckNewsfr (*Libération*).

The statistics presented in the following table are based on information collected on 26 November 2019:

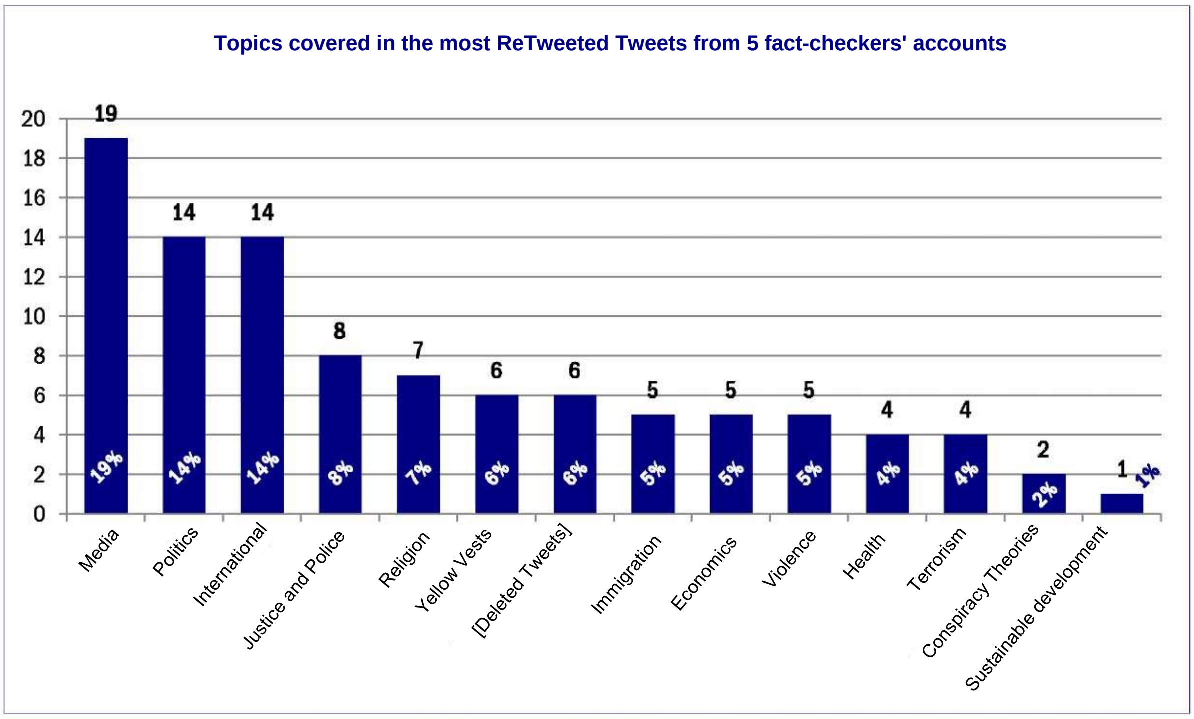
Account name	Number of Tweets analysed	Mean number of ReTweets per Tweet	Number of subscribers	Total number of Tweets published by the account	Account creation date	Mean number of ReTweets per Tweet per 1,000 subscribers
20minFakeOff	475	6.93	2,902	480	18/09/2018 at 11:58:15	2.39
AFPFactuel	2,451	78.91	91,012	2,775	20/11/2017 at 14:24:50	0.87
CheckNewsfr	3,185	7.01	40,680	7,145	26/09/2017 at 16:25:10	0.17
Décodeurs	3,083	17.16	141,819	17,070	03/11/2009 at 10:32:57	0.12
InfoIntoxF24	891	5.04	4,752	944	25/02/2019 at 11:33:55	1.06

Certain accounts such as AFPFactuel and Décodeurs—and, to a lesser extent, Checknews—are followed by a very large number of subscribers. It should be remembered that the mean number of subscribers to category 2 accounts is nearly 14,000. The mean number of ReTweets per Tweet is relatively low, except for AFPFactuel and Décodeurs accounts. In comparison, the mean number of ReTweets per Tweet for category 2 accounts is 18 whereas It is 14 for category 4 accounts. Lastly, some fact-checking accounts are able to obtain a substantial share from their subscribers. Category 2 and 4 accounts have a mean average of 1.5 ReTweets and 0.15 ReTweets per 1,000 subscribers respectively.

In order to present the activity of the fact-checkers on Twitter in a more concrete way and to better understand the nature of the Tweets that are most conducive to sharing, the content of the twenty Tweets that were the subject of the highest number of ReTweets among the selected fact-checkers is analysed below.

4.5.2. Analysis of the topics covered by the most ReTweeted Tweets by fact-checkers

This analysis was carried out on the 20 most ReTweeted Tweets of each account[53]. Using the same analysis grid as the one used for category 2 account Tweets, fact-checkers' Tweets have been classified according to 22 different topics[54].



The thematic analysis will focus on those most strongly represented in this corpus: "Media" (19 Tweets), "Politics" (14) and "International" (14) represent 47% of the most ReTweeted Tweets. By way of comparison, the topic "Immigration", the second most strongly represented among category 2 accounts, only appears in 9th place among fact-checker accounts. For the purposes of comparison, it will nevertheless be added to the thematic analysis.

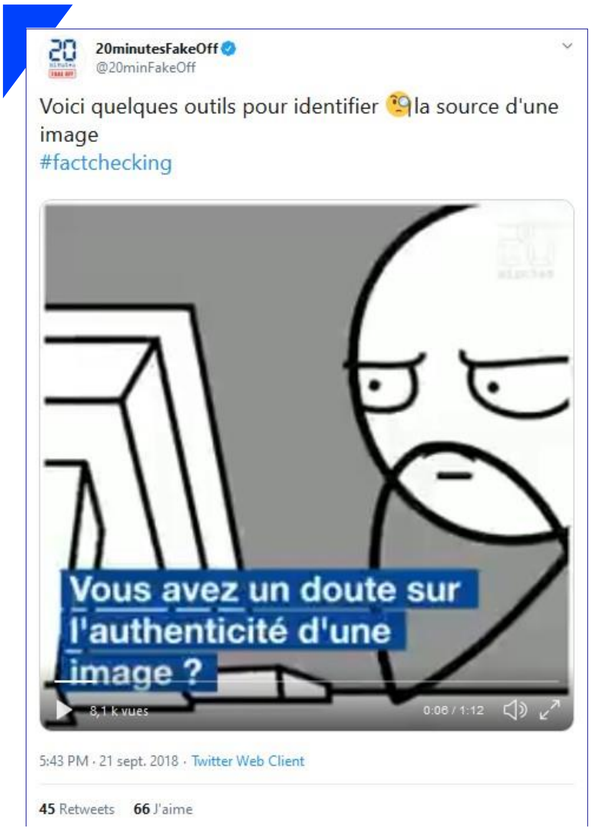
[53] Of the 100 Tweets initially collected, some were responses to publications by internet users, for which the fact-checkers promoted their verification articles. In this type of sharing, the authors of the initial Tweet have sometimes suppressed their publication, resulting in the inaccessibility of the content of the linked fact-checker. 6 Tweets could not be analysed for this reason (n = 94).

[54] Each Tweet was counted only once, according to the main topic it addressed**.** This method may have limitations in that some Tweets may address several topics. However, the trends and proportions highlighted remain valid given the significant differences observed between the topics.

- **Media (19 Tweets)**

Of the 19 Tweets in this category, 12 focus on the dissemination of tools to help the public to exercise their critical sense (vertical reference), 5 refer to fact-checking initiatives, particularly European (circular reference), and 2 Tweets focus on checking or pointing out mistakes made by other media (horizontal reference). Unlike category 2 accounts, this is not in any way a criticism of the media as a whole, but rather an attempt to help Internet users navigate an environment characterised by immediacy and density. The volume of Tweets collected on this topic thus, in addition to the subject matter, evidences the promotion of the fact-checking method used by journalists.

A large proportion of the Tweets in this category (12) therefore offer users tools to support them in their consumption of information on the Internet (vertical reference). These publications are divided between the dissemination of photo or video verification tools and case studies illustrating the factors behind misinterpreted information.

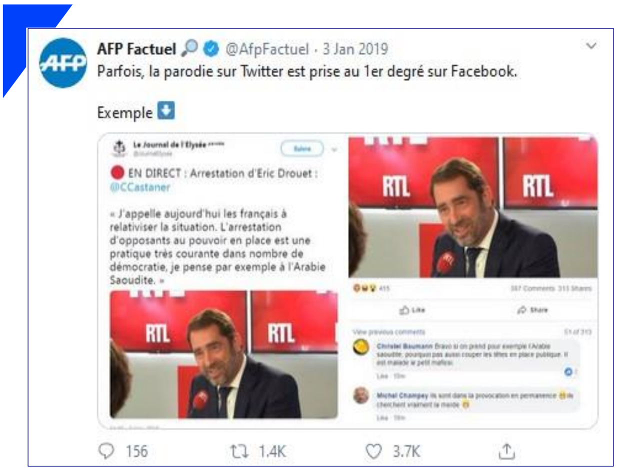


20minutesFakeOff
@20minFakeOff

Voici quelques outils pour identifier 🧐la source d'une image
#factchecking

5:43 PM · 21 sept. 2018 · Twitter Web Client

45 Retweets 66 J'aime



AFP Factuel 🔎 @AfpFactuel · 3 Jan 2019
Parfois, la parodie sur Twitter est prise au 1er degré sur Facebook.

Exemple ⬇️

Le Journal de l'Elysée

🔴 EN DIRECT : Arrestation d'Eric Drouet : @CCastaner

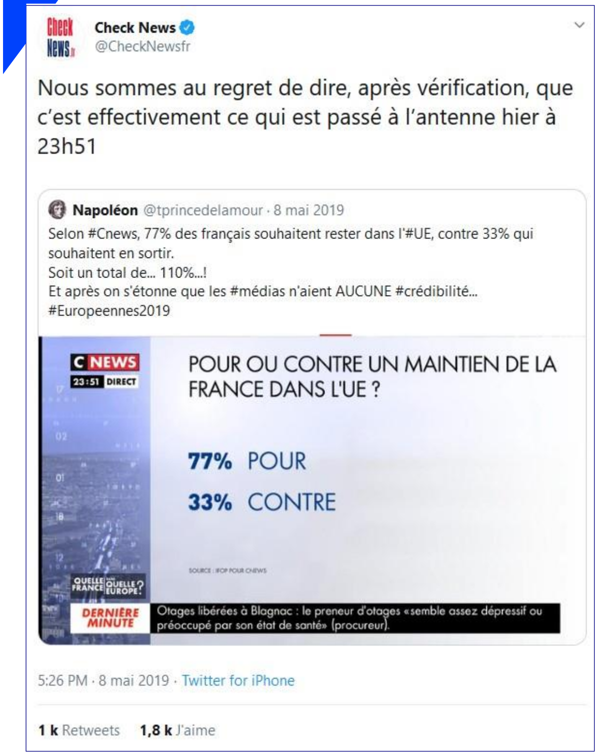
« J'appelle aujourd'hui les français à relativiser la situation. L'arrestation d'opposants au pouvoir en place est une pratique très courante dans nombre de démocratie, je pense par exemple à l'Arabie Saoudite. »

Christel Baumann Bravo si on prend pour exemple l'Arabie saoudite, pourquoi pas aussi couper les têtes en place publique. Il est malade le petit mafieux

Michel Champey Ils sont dans la provocation en permanence ils cherchent vraiment la merde

156 1.4K 3.7K

Examples of vertical reference Tweets published by fact-checkers.



Check News @CheckNewsfr

Nous sommes au regret de dire, après vérification, que c'est effectivement ce qui est passé à l'antenne hier à 23h51

Napoléon @tprincedelamour · 8 mai 2019

Selon #Cnews, 77% des français souhaitent rester dans l'#UE, contre 33% qui souhaitent en sortir.
Soit un total de... 110%...!
Et après on s'étonne que les #médias n'aient AUCUNE #crédibilité...
#Europeennes2019

C NEWS
23:51 DIRECT
POUR OU CONTRE UN MAINTIEN DE LA FRANCE DANS L'UE ?
77% POUR
33% CONTRE
SOURCE : IFOP POUR CNEWS
QUELLE FRANCE QUELLE EUROPE ?
DERNIÈRE MINUTE
Otages libérées à Blagnac : le preneur d'otages «semble assez dépressif ou préoccupé par son état de santé» (procureur).

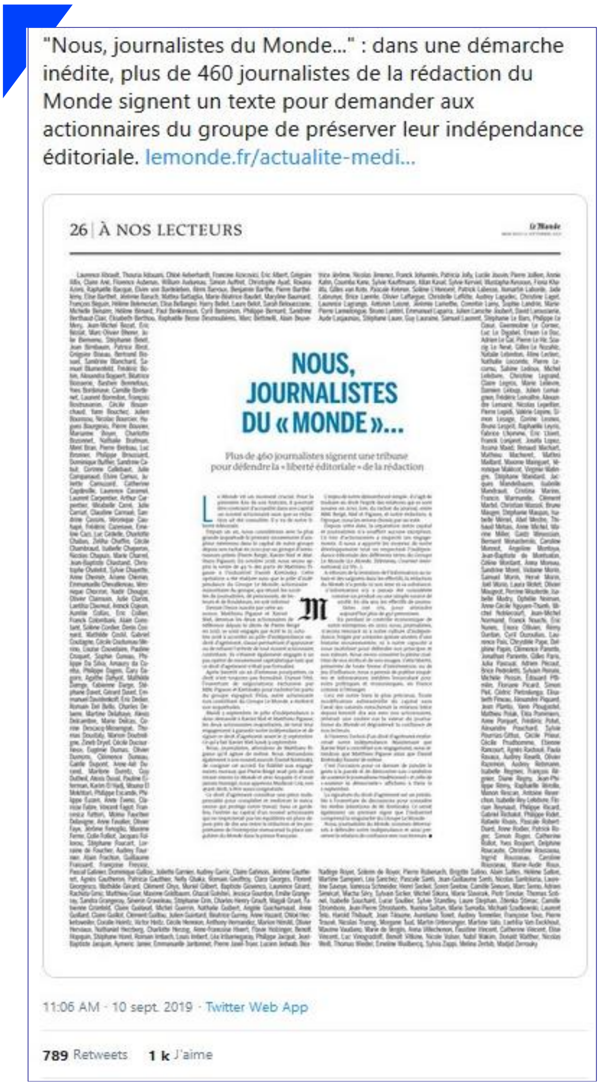
5:26 PM · 8 mai 2019 · Twitter for iPhone

1 k Retweets 1,8 k J'aime

Within the corpus of the most ReTweeted Tweets, Checknewsfr is the only one to have published a message relating to the fact-checking information published by fellow journalists (2 Tweets).

Example of a horizontal reference Tweet published by a fact-checker.

Lastly, 5 Tweets promote fact-checking initiatives or the independence of the media from which they originate (circular verification). This is notably the case for *Le Monde*'s "Décodeurs", a media outlet that has been at the heart of an issue related to the preservation of editorial independence during a change of shareholders. The most emblematic examples of communication here remain those relating to fact-checking initiatives between several editorial offices. In both cases, it is more a question of ReTweets from these organisations (3 cases out of 5) than Tweets published by the account studied.



"Nous, journalistes du Monde..." : dans une démarche inédite, plus de 460 journalistes de la rédaction du Monde signent un texte pour demander aux actionnaires du groupe de préserver leur indépendance éditoriale. lemonde.fr/actualite-medi...

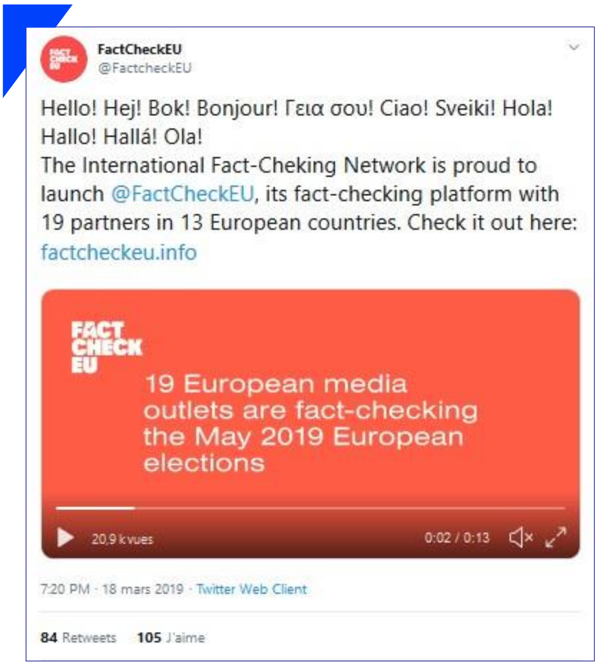
26 | À NOS LECTEURS

NOUS, JOURNALISTES DU « MONDE »...

Plus de 460 journalistes signent une tribune pour défendre la « liberté éditoriale » de la rédaction

11:06 AM · 10 sept. 2019 · Twitter Web App

789 Retweets 1 k J'aime



FactCheckEU
@FactcheckEU

Hello! Hej! Bok! Bonjour! Γεια σου! Ciao! Sveiki! Hola! Hallo! Hallá! Ola!
The International Fact-Cheking Network is proud to launch @FactCheckEU, its fact-checking platform with 19 partners in 13 European countries. Check it out here: factcheckeu.info

FACT CHECK EU
19 European media outlets are fact-checking the May 2019 European elections
20,9 k vues
0:02 / 0:13

7:20 PM · 18 mars 2019 · Twitter Web Client

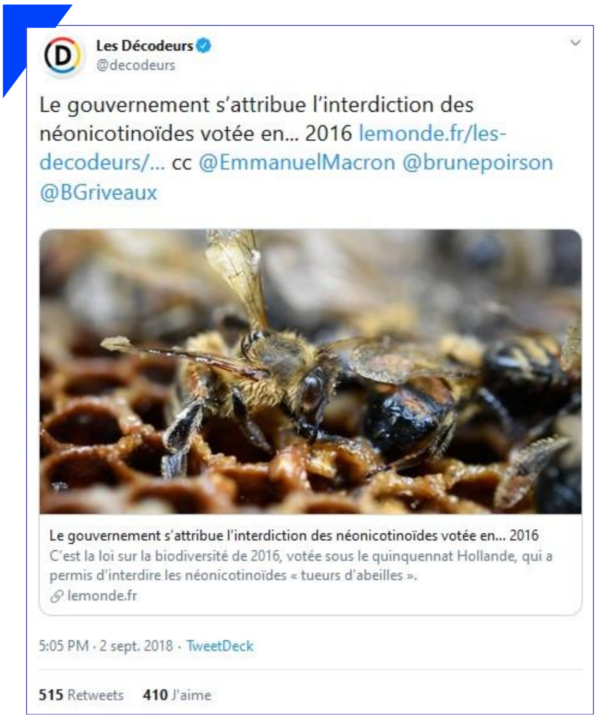
84 Retweets 105 J'aime

Examples of circular reference Tweets published by fact-checkers.

- **Politics (14 Tweets)**

Among the Tweets in this corpus, a majority deal with the head of state or government and its members (10 Tweets). The rest of the Tweets concern specific personalities (3), while only one deals with the political world as a whole.

The publications dealing with the President of the Republic or the government are also those that have a concentration of the Tweets that have generated the most reactions; these are often Tweets whose purpose is to evaluate the comments made by personalities (7 out of 10). For example, the Décodeurs analyse the ban on neonicotinoids, which the current government is said to have taken credit for—although, according to the government, the decision was made before it came to power. The other 3 Tweets, for example, deal with events in which members of the executive have participated.



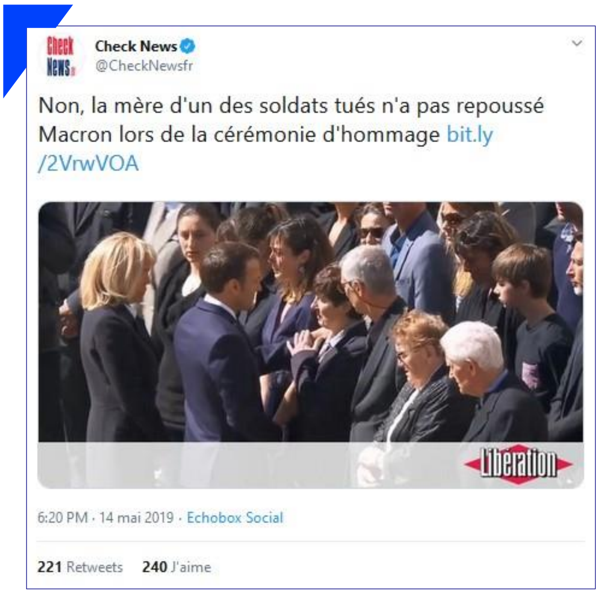
Les Décodeurs
@decodeurs

Le gouvernement s'attribue l'interdiction des néonicotinoïdes votée en... 2016 lemonde.fr/les-decodeurs/... cc @EmmanuelMacron @brunepoirson @BGriveaux

Le gouvernement s'attribue l'interdiction des néonicotinoïdes votée en... 2016
C'est la loi sur la biodiversité de 2016, votée sous le quinquennat Hollande, qui a permis d'interdire les néonicotinoïdes « tueurs d'abeilles ».
lemonde.fr

5:05 PM · 2 sept. 2018 · TweetDeck

515 Retweets 410 J'aime



Check News
@CheckNewsfr

Non, la mère d'un des soldats tués n'a pas repoussé Macron lors de la cérémonie d'hommage bit.ly/2VrwVOA

6:20 PM · 14 mai 2019 · Echobox Social

221 Retweets 240 J'aime

Examples of Tweets on the Head of State and the government published by the fact-checkers.

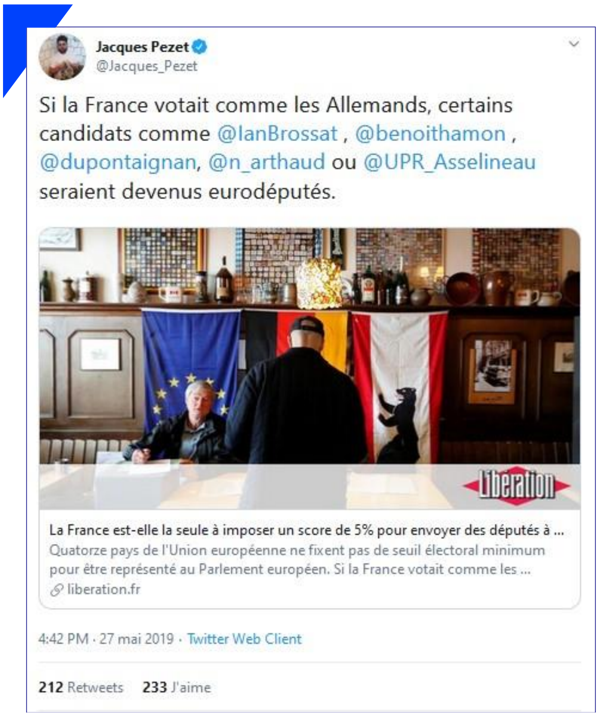
The other Tweets on this topic deal with other political figures and the political world as a whole, starting mainly with questions from citizens questioning the truthfulness of statements made by personalities. Fact-checkers refer to this using the interrogative form or expressions such as "according to this rumour". The aim here is to start from a suspicion that has been read, seen or heard by journalists and to launch a verification process based on this. The case of the photo of the Chairman of the Finance Committee of the National Assembly and former Minister Eric Woerth, whose truthfulness in climbing the Aiguille d'Argentière mountain was strongly questioned, is emblematic of this fact-checking process, which is based on doubts expressed by Internet users.

Compared to the most popular Tweets in category 2 accounts, far-right personalities are underrepresented in the corpus studied here.

- **International (14 Tweets)**

The 14 Tweets in this category mainly focus on two geographical areas: the European Union (5 Tweets) and Russia (3). The other Tweets deal with forest fires in the Amazon as well as other topics, but in an isolated way (a military operation in Mali, a news item about a French footballer during a stay in the United Arab Emirates, a hoax in Africa).

The European elections of May 2019 and an alleged ban on the Rothschild banks in Russia generate the most ReTweets. The former case is an attempt to educate. The second case relates to fact-checking. The purpose is therefore to focus on information that is tangential to national news, and not to check facts relating to a rumour with no connection to France, as some category 2 accounts have done.



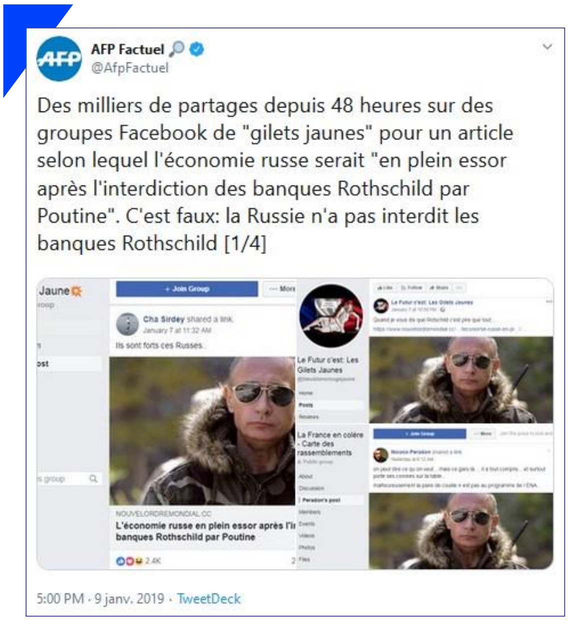
Jacques Pezet
@Jacques_Pezet

Si la France votait comme les Allemands, certains candidats comme @IanBrossat , @benoithamon , @dupontaignan, @n_arthaud ou @UPR_Asselineau seraient devenus eurodéputés.

La France est-elle la seule à imposer un score de 5% pour envoyer des députés à ...
Quatorze pays de l'Union européenne ne fixent pas de seuil électoral minimum pour être représenté au Parlement européen. Si la France votait comme les ...
liberation.fr

4:42 PM · 27 mai 2019 · Twitter Web Client

212 Retweets 233 J'aime



AFP Factuel
@AfpFactuel

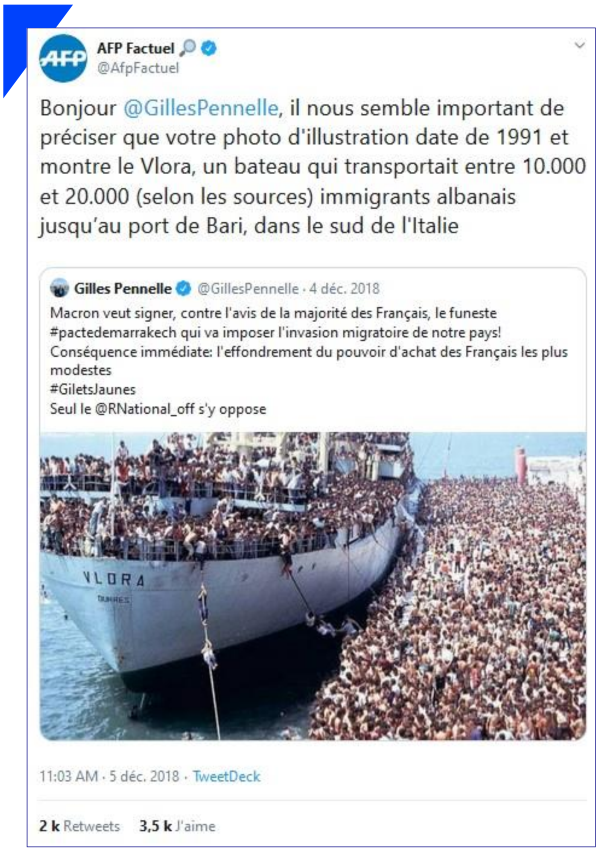
Des milliers de partages depuis 48 heures sur des groupes Facebook de "gilets jaunes" pour un article selon lequel l'économie russe serait "en plein essor après l'interdiction des banques Rothschild par Poutine". C'est faux: la Russie n'a pas interdit les banques Rothschild [1/4]

5:00 PM · 9 janv. 2019 · TweetDeck

Examples of Tweets on the European Union and Russia published by the fact-checkers.

- **Immigration (5 Tweets)**

Immigration is a topic that is very much a focus among category 2 accounts, whose most ReTweeted Tweets address this subject 29 times (10.3% of the corpus); it is also dealt with by fact-checkers. The purpose here is to verify statements, especially those made by personalities. 4 out of the 5 Tweets in this category concern comments made by RN (Rassemblement National) elected representatives which, according to the journalists, prove to be false because they are based either on images taken out of context or on figures which have proven to be wrong.



AFP Factuel 🔎 @AfpFactuel

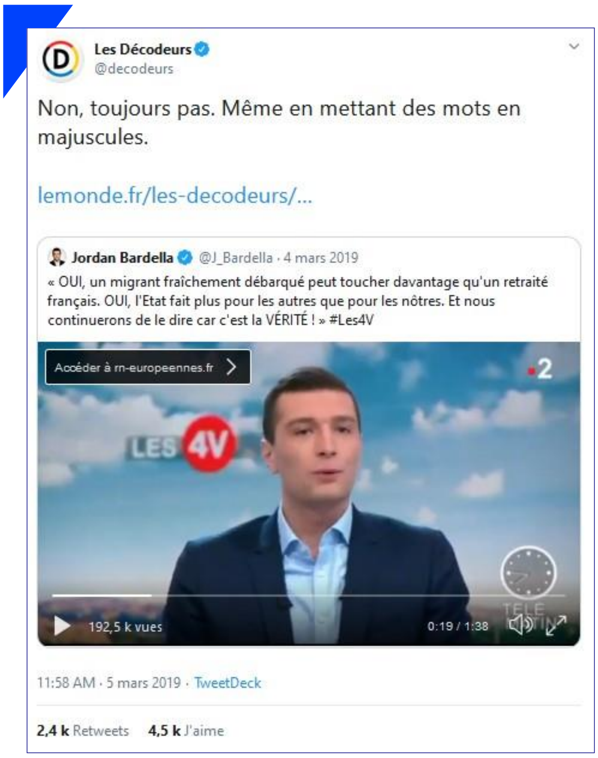
Bonjour @GillesPennelle, il nous semble important de préciser que votre photo d'illustration date de 1991 et montre le Vlora, un bateau qui transportait entre 10.000 et 20.000 (selon les sources) immigrants albanais jusqu'au port de Bari, dans le sud de l'Italie

Gilles Pennelle @GillesPennelle · 4 déc. 2018

Macron veut signer, contre l'avis de la majorité des Français, le funeste #pactedemarrakech qui va imposer l'invasion migratoire de notre pays!
Conséquence immédiate: l'effondrement du pouvoir d'achat des Français les plus modestes
#GiletsJaunes
Seul le @RNational_off s'y oppose

11:03 AM · 5 déc. 2018 · TweetDeck

2 k Retweets 3,5 k J'aime



Les Décodeurs @decodeurs

Non, toujours pas. Même en mettant des mots en majuscules.

lemonde.fr/les-decodeurs/...

Jordan Bardella @J_Bardella · 4 mars 2019

« OUI, un migrant fraîchement débarqué peut toucher davantage qu'un retraité français. OUI, l'Etat fait plus pour les autres que pour les nôtres. Et nous continuerons de le dire car c'est la VÉRITÉ ! » #Les4V

11:58 AM · 5 mars 2019 · TweetDeck

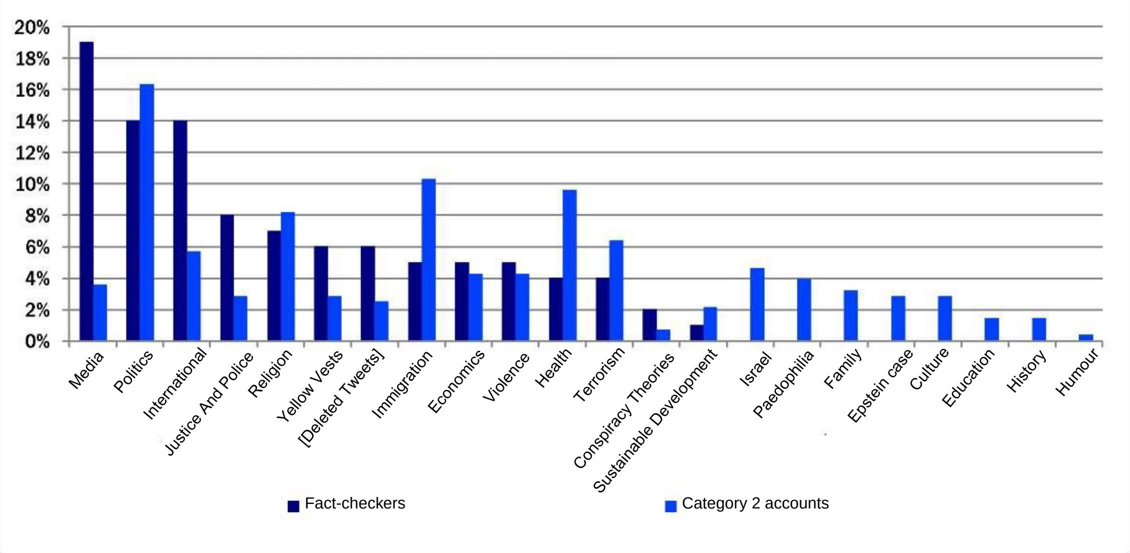
2,4 k Retweets 4,5 k J'aime

Examples of Tweets on immigration published by fact-checkers.

4.5.3. Comparison of results with those from category 2 accounts

A comparative analysis of the most widely shared Tweets among category 2 accounts and accounts from journalists specialising in fact-checking showed that **politics** (1st among category 2 accounts, 2nd among fact-checkers) and **religion** (4th vs. 5th) are also very well divided between both categories. The topics of immigration (2nd vs. 9th), health (3rd vs. 11th) and terrorism (5th vs. 12th) elicited very different forms of engagement. They have been widely recirculated among category 2 accounts, unlike the accounts of the fact-checkers.

Comparison of the proportions of Tweets by topic between category 2 accounts and fact-checkers



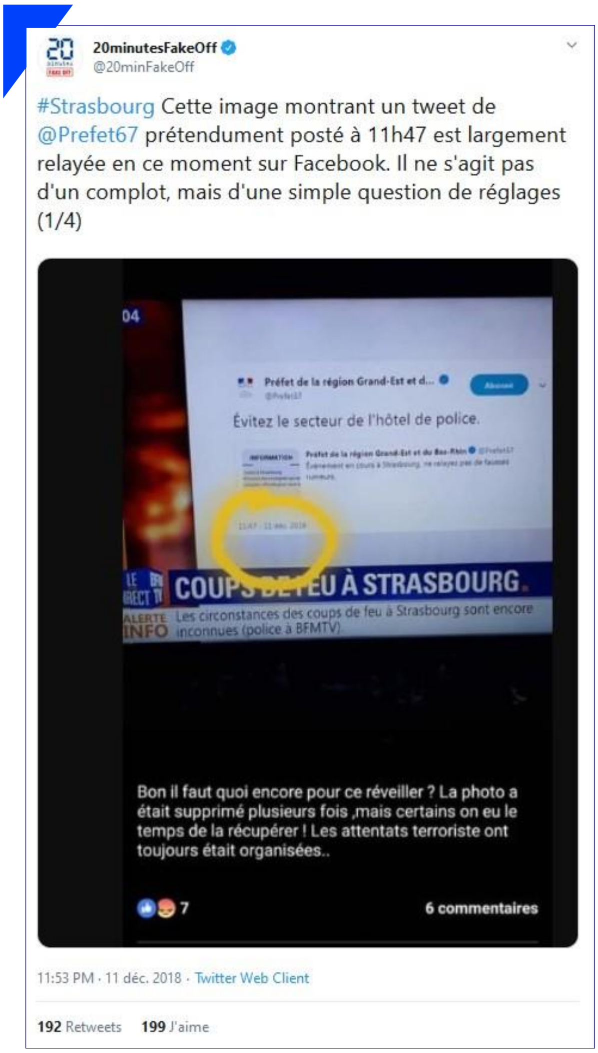
These results are consistent with the analysis carried out by a doctoral student and SciencesPo students[55]: in January 2020, they analysed the latest 50 articles of three fact-checking columns (AFP Factuel, Les Décodeurs and Checknews) and found that the topics covered were mostly political (at 30, 33 and 28% respectively) while the environment (2, 10 and 6%), religion (4, 6 and 6%) and health (6, 8 and 2%) were marginal[56].

4.5.4. Analysis of information processing

Based on the 5 most ReTweeted Tweets from each fact-checking account studied , a corpus of 25 Tweets was built up, which made it possible to highlight, **from among the areas that seem to elicit the greatest interest among Internet users, fact-checking work for content that has already been widely circulated and the use of a direct form of address**; say, by commenting directly on the content in a comment for the Tweet published by the author, rather than checking statements made by political figures, for example. These are *"media formats facilitating dialogue with the general public"* which, according to research conducted at SciencesPo, are the most effective and attractive for Internet users[57].

7 Tweets in our corpus thus refer to widely distributed content (*"this image[...] has been widely shared"* or *"122,000 shares on FB for these 7 photos"*). Here we find the previously described vertical reference phenomenon.

[55] Manon Berriche. Sciences Po. *Le fact-checking est-il vraiment efficace ? (Is fact-checking really effective?).* 24/01/2020.
[56] This result should be seen in the context of its publication date, in January 2020, prior to the widespread media coverage of Covid-19.
[57] Ibid.



20minutesFakeOff
@20minFakeOff

#Strasbourg Cette image montrant un tweet de @Prefet67 prétendument posté à 11h47 est largement relayée en ce moment sur Facebook. Il ne s'agit pas d'un complot, mais d'une simple question de réglages (1/4)

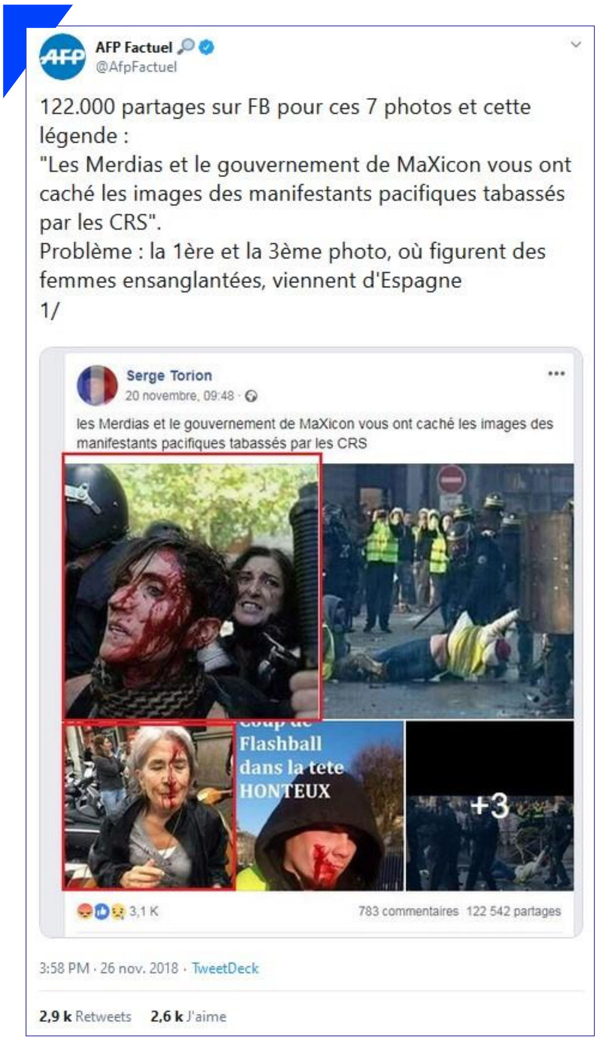
04
Préfet de la région Grand-Est et d...
Évitez le secteur de l'hôtel de police.
COUPS DE FEU À STRASBOURG
LE DIRECT TV
ALERTE INFO
Les circonstances des coups de feu à Strasbourg sont encore inconnues (police à BFMTV)

Bon il faut quoi encore pour ce réveiller ? La photo a était supprimé plusieurs fois ,mais certains on eu le temps de la récupérer ! Les attentats terroriste ont toujours était organisées..

7 6 commentaires

11:53 PM · 11 déc. 2018 · Twitter Web Client

192 Retweets 199 J'aime



AFP Factuel
@AfpFactuel

122.000 partages sur FB pour ces 7 photos et cette légende :
"Les Merdias et le gouvernement de MaXicon vous ont caché les images des manifestants pacifiques tabassés par les CRS".
Problème : la 1ère et la 3ème photo, où figurent des femmes ensanglantées, viennent d'Espagne
1/

Serge Torion
20 novembre, 09:48

les Merdias et le gouvernement de MaXicon vous ont caché les images des manifestants pacifiques tabassés par les CRS

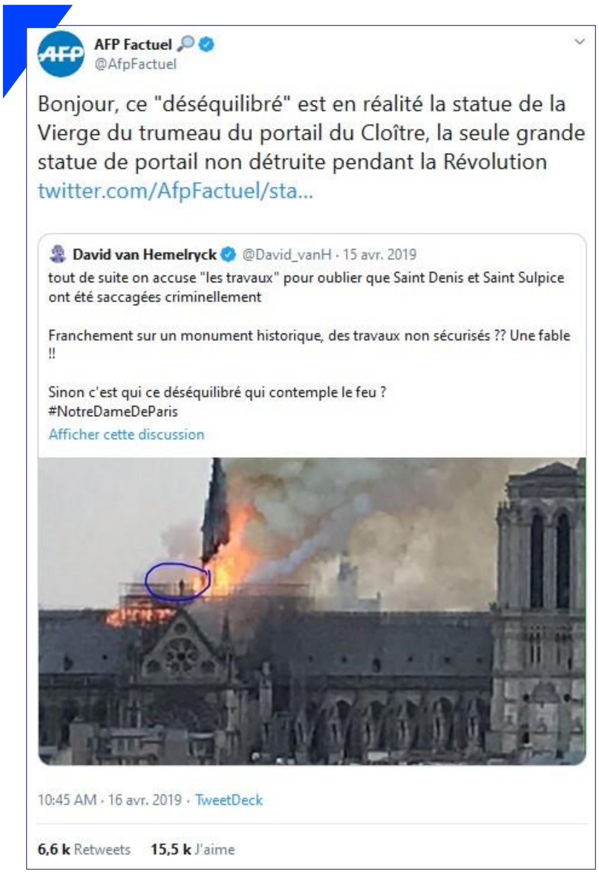
3,1 K 783 commentaires 122 542 partages

3:58 PM · 26 nov. 2018 · TweetDeck

2,9 k Retweets 2,6 k J'aime

Examples of Tweets on the fact-checking of content that is widely circulated by fact-checkers.

The direct form of address (3 Tweets) is mainly used by AFP Factuel—which includes its fact-checking analysis into a dialogue with Internet users who have circulated information that turns out to be false—and by CheckNewsfr. This type of post works not by posting a Tweet but by replying to an initial message which has, for example, circulated information that is then categorised as erroneous (thread or conversation). Firstly, this makes it easier to reach the person who is the source of the fake news—in particular, because he or she receives a notification—and secondly, it creates visibility among the target audience and enables the fact-checking effort to be circulated in a more targeted way to people who are sensitive to fake news.



AFP Factuel
@AfpFactuel

Bonjour, ce "déséquilibré" est en réalité la statue de la Vierge du trumeau du portail du Cloître, la seule grande statue de portail non détruite pendant la Révolution twitter.com/AfpFactuel/sta...

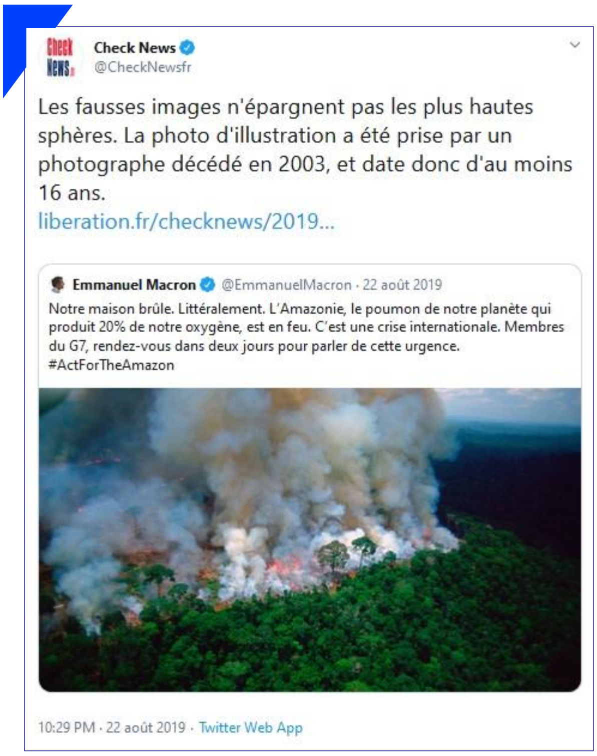
David van Hemelryck @David_vanH · 15 avr. 2019
tout de suite on accuse "les travaux" pour oublier que Saint Denis et Saint Sulpice ont été saccagées criminellement

Franchement sur un monument historique, des travaux non sécurisés ?? Une fable !!

Sinon c'est qui ce déséquilibré qui contemple le feu ?
#NotreDameDeParis
Afficher cette discussion

10:45 AM · 16 avr. 2019 · TweetDeck

6,6 k Retweets 15,5 k J'aime



Check News
@CheckNewsfr

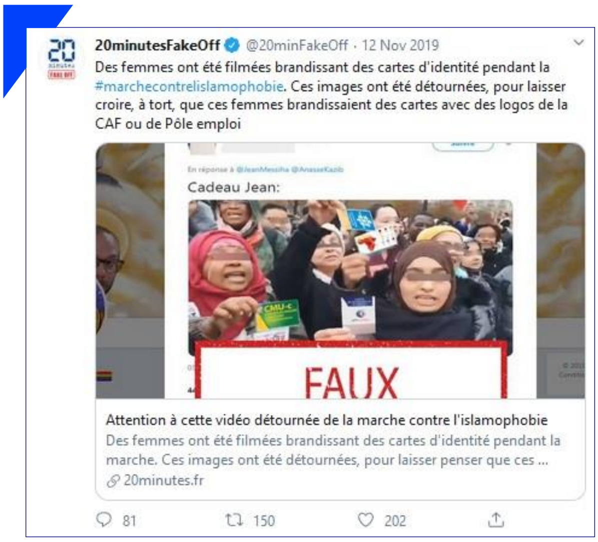
Les fausses images n'épargnent pas les plus hautes sphères. La photo d'illustration a été prise par un photographe décédé en 2003, et date donc d'au moins 16 ans.
liberation.fr/checknews/2019...

Emmanuel Macron @EmmanuelMacron · 22 août 2019
Notre maison brûle. Littéralement. L'Amazonie, le poumon de notre planète qui produit 20% de notre oxygène, est en feu. C'est une crise internationale. Membres du G7, rendez-vous dans deux jours pour parler de cette urgence.
#ActForTheAmazon

10:29 PM · 22 août 2019 · Twitter Web App

Examples of Tweets using the direct form of address posted by fact-checkers.

Finally, in terms of the type of content used, the use of non-text material such as photos (9), quotes from verified Tweets (6) and animation or .gif files (1) to support the verification purpose is notable. This is particularly the case with Tweets that focus on image verification, to show the decontextualisation and even manipulation that can be hidden behind them.



20minutesFakeOff @20minFakeOff · 12 Nov 2019

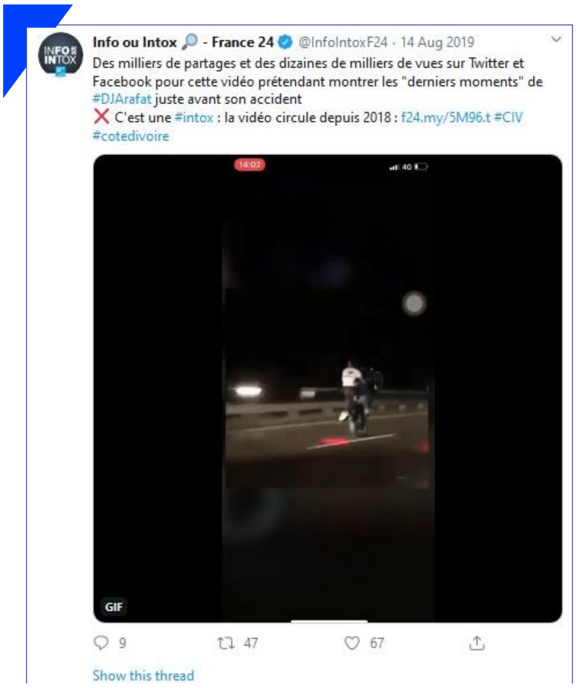
Des femmes ont été filmées brandissant des cartes d'identité pendant la #marchecontrelislamophobie. Ces images ont été détournées, pour laisser croire, à tort, que ces femmes brandissaient des cartes avec des logos de la CAF ou de Pôle emploi

Attention à cette vidéo détournée de la marche contre l'islamophobie

Des femmes ont été filmées brandissant des cartes d'identité pendant la marche. Ces images ont été détournées, pour laisser penser que ces ...

20minutes.fr

81 150 202



Info ou Intox 🔎 - France 24 @InfoIntoxF24 · 14 Aug 2019

Des milliers de partages et des dizaines de milliers de vues sur Twitter et Facebook pour cette vidéo prétendant montrer les "derniers moments" de #DJArafat juste avant son accident

❌ C'est une #intox : la vidéo circule depuis 2018 : f24.my/5M96.t #CIV #cotedivoire

9 47 67

Show this thread

Examples of Tweets using video content published by fact-checkers.

Summary:

On average, the accounts of journalism units specialising in fact-checking have more followers than category 2 accounts, but their publications generate fewer interactions than the latter accounts. In terms of the topics most frequently dealt with, only politics and religion are common to these two types of accounts. Fact-checkers take more interest in topics related to their profession, such as media news or content related to media and information literacy.

5. Analysis of Tweets linked to fake news: chronology of propagation and correction

This section addresses the problem of fake news from a complementary angle to the previous section by identifying the Tweets related to a given fake news item and by attempting to identify users who propagate the fake news and users who will, conversely, correct it. The accounts categorised by the “Décodex” examined in the previous section do not constitute an exhaustive list of accounts likely to share or correct fake news on Twitter. The methodology used in this section refers to a number of case studies and will therefore provide a better understanding of the phenomena of propagation of fake news.

5.1. Description of selected fake news and data collection methodology

The fake news studied in this section was identified on fact-checking sites. For each item of fake news selected, keywords were chosen to help identify the related Tweets (they were tested to ensure that they were the most appropriate). The Twitter API was then used to collect all Tweets containing these keywords. The Twitter API in its free version (which is the one used in this study) only allows the collection of Tweets posted over the previous 7 days (approximately). This time limit is a significant constraint, as it implies that fake news must be identified less than 7 days after its appearance on the social media platform if it is to be possible to collect the entire history of the propagation of this fake news. For some of the selected fake news, the Tweets relating to the selected keywords were also collected several times, separated in time and thus enabling the study of fake news dissemination for more than 7 days.

The fake news selected is not exhaustive, nor is it necessarily representative of all fake news. The result of focusing only on fake news that has been refuted by fact-checkers could be that particularly viral fake news is selected. The analyses will therefore form case studies that will make it possible to assess examples of the propagation of fake news on Twitter without necessarily making it possible to draw general conclusions from the set of results obtained.

The following table summarises the fake news discussed in this section. Not all items represent deliberate manipulation that could be categorised as disinformation, but they all meet the only criterion used by the *Conseil*; that is, that they have been verified by a fact-checking site.

TOPIC	SELECTEDKEYWORDS	DESCRIPTION OF THE FALSE INFORMATION AND LINK TO A FACT-CHECKER'S ARTICLE
Greta Thunberg's activity on Facebook	Greta and Facebook	Greta Thunberg's Facebook posts are supposedly written by her father and a Swedish environmental activist/Indian activist[58].
Coronavirus patent	Patent and Coronavirus or Coronavirus and patented	The new coronavirus was allegedly patented in 2003 in the United States, proving that the virus was created intentionally[59].
Parcels /coronavirus	Parcels and Coronavirus	Coronavirus can supposedly be transmitted via parcels from China[60].
Coronavirus in Argenteuil	Argenteuil and Coronavirus	In January, people suffering from coronavirus were hospitalised in Argenteuil, Paris[61].
Coronavirus in Perpignan	Perpignan and Coronavirus	In January, cases of coronavirus were reportedly detected in Perpignan[62].
Scabies in hospital in town of Nevers	Epidemic and Nevers or Scabies and Nevers	The hospital in Nevers was allegedly placed in quarantine following a scabies epidemic caused by immigrant patients.
Links between Bill Gates and the coronavirus	Gates and coronavirus	The Bill Gates Foundation has been linked to several fake news stories: the Foundation allegedly predicted the coronavirus outbreak and funded the group holding the patent on the new coronavirus.[63]
Sign board in Limoges	Limoges and sign board or Limoges and sign board or Limoges and women	A sign board in the city of Limoges supposedly asks women not to walk alone in the street and to avoid dark alleys[64].
Netflix	Netflix and sharing	Netflix is allegedly planning to end account sharing[65].
Water pollution in Rouen	Water and Rouen	Following the fire at the Lubrizol factory, water is reportedly no longer drinkable in Rouen.[66]
Statements by	Blanquer and	Minister Jean-Michel Blanquer is reported to have claimed that

[58] *Liberation. Un bug sur Facebook a-t-il révélé que les messages de Greta Thunberg étaient écrits par son père ? (Did a bug on Facebook reveal that Greta Thunberg's messages were written by her father?)* 16/01/2020.

[59] *Le Monde. Le coronavirus qui sévit en Chine n'a pas été « créé en 2003 aux USA » (The coronavirus ravaging China was not "created in 2003 in the USA*"). 27/01/2020.

[60] *franceinfo:. Coronavirus : un colis en provenance de Chine peut-il transmettre la maladie ? (Coronavirus: can a parcel from China transmit the disease?)* 28/01/2020.

[61] *AFP. Nouveau coronavirus : attention aux fausses captures d'écran sur les réseaux sociaux (New coronavirus: beware of false screenshots on social media).* 28/01/2020.

[62] *20 Minutes. Coronavirus: Non, aucun cas n'a été détecté à Perpignan, indique la prefecture (No, no cases were detected in Perpignan, says the prefecture)*. 26/01/2020.

[63] *Le Monde. Coronavirus: Bill Gates ciblé par des rumeurs et infox complotistes (Bill Gates targeted by rumours and conspiracy theorists)*. 05/02/2020.

[64] *Le Monde. « Mesdames, évitez de rentrer seule le soir... » : le message factice de la ville de Limoges ("Ladies, don't go home alone in the evening...": the dummy message from the city of Limoges).* 12/11/2019.

[65] *franceinfo:. Netflix : la rumeur sur la suppression du partage de compte fait son grand retour (Netflix: Rumour about the removal of account sharing is back*). 23/10/2019.

[66] *Le Monde. Après l'incendie de l'usine Lubrizol à Rouen, des intox en série (After the fire at the Lubrizol factory in Rouen, a series of hoaxes)*. 30/09/2019.

Mr Blanquer	99.9%.	99.9% of teachers support the reform of the baccalaureate qualification[67].
Nurses' salaries	nurse and happy meal	A home care nurse is allegedly paid the equivalent of one Happy Meal per patient[68].
TPMP	TPMP and cancelled or TPMP and CSA	A source inside the CSA reportedly announced the cancellation of the "Touche pas à mon poste" (TPMP) TV programme[69].

The following table shows the number of Tweets collected for each selected fake news item and the period covered by the collection. For example, the search for Tweets concerning the transmission of coronavirus by parcel delivery (keywords "Parcel and coronavirus") returned 12,506 Tweets. A significant proportion of these Tweets are ReTweets, so the number of different Tweets is lower, at 1,438.

TOPIC	NUMBER OF TWEETS COLLECTED	NUMBER OF DIFFERENT TWEETS COLLECTED	CAPTURE TIME (IN DAYS)70
Activity of Greta Thunberg on Facebook	6,093	894	25
Coronavirus patent	669	300	18
Coronavirus parcels	12,506	1,438	20
Coronavirus in Argenteuil	45	31	10
Coronavirus in Perpignan	50	23	4
Scabies in Nevers hospital	745	60	11
Gates and Coronavirus	819	406	18
Netflix	835	339	8
Sign board - Limoges	23,983	291	9
Water pollution in Rouen	101,734	4,470	51
Statements by Mr Blanquer	3,100	630	15
Nurses' salaries	5,720	26	30
TPMP	156	29	24

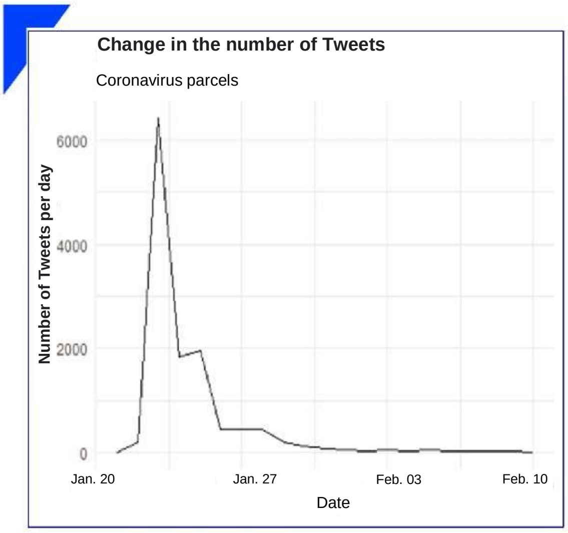
[67] *Liberation. Blanquer a-t-il vraiment dit que «99,9% des enseignants soutiennent la réforme du bac» ? (Did Blanquer actually say that "99.9% of teachers support the baccalaureate reform"?)* 23/01/2020.

[68] *franceinfo:. Une infirmière à domicile est-elle payée moins par patient que le prix d'un "Happy Meal" ? (Is a home care nurse paid less than the price of a "Happy Meal" per patient?)* 18/10/2019.

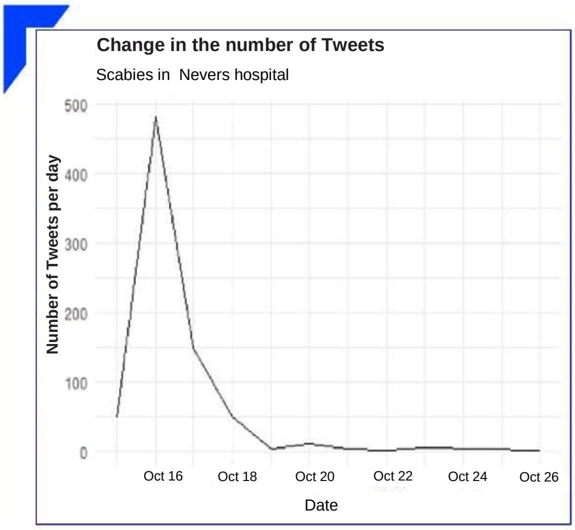
[69] *20 Minutes. « Touche pas à mon poste » déprogrammée ? ("Touche pas à mon poste" cancelled?) Denial by Cyril Hanouna.* 26/11/2019.

[70] The collection time was not equivalent for all topics. This column indicates the number of days for which Tweets were collected for a given topic.

All of the fake news studied has the characteristic of presenting a high concentration of Tweets over a very short period. The following graphs show an example of the chronology of the Tweets for two of the fake news items studied: the transmission of the coronavirus by parcel and the quarantine of the hospital in Nevers due to a scabies epidemic (graphs for the other fake news items studied are shown in the appendix):


Change in the number of Tweets

Coronavirus parcels


Change in the number of Tweets

Scabies in Nevers hospital

This propagation profile is more or less the same for all the fake news items studied. It should be noted, however, that the collection, in this case, covered a limited period, and that in some cases the circulation of fake news may have resumed beyond the collection period. For example, the fake news concerning the existence of a patent on the new coronavirus assumed different forms at different times[71].

An optimistic hypothesis to explain the form taken by the Tweets timeline could be that the fake news started to spread and was then refuted by fact-checkers, whose Tweets were subsequently widely ReTweeted, making the fake news disappear. The following sections will show whether or not this hypothesis is verified.

5.2. Methodology for qualifying collected Tweets

The collection phase was followed by a phase of qualifying the content of the Tweets collected. Because of the sheer number of these Tweets, it was not possible to qualify them in full. Only the 50 most ReTweeted messages for each topic were studied and qualified.

In all cases, this approach, restricted to the most ReTweeted messages, made it possible to process the majority of the Tweets collected (each collected Tweet being counted as a Tweet, even if it was a ReTweet). This observation immediately highlights the phenomenon of the

[71] *20 Minutes. Coronavirus: Non, le Covid-19 n'a pas été breveté par l'Institut Pasteur en 2004 (No, Covid-19 was not patented by the Pasteur Institute in 2004).* 18/03/2020.

concentration of the conversation on Twitter around certain Tweets that are widely ReTweeted. For example, it was noted that the search for Tweets relating to the transmission of coronavirus via parcels had resulted in the collection of 12,506 Tweets (original or ReTweeted). By labelling only the 50 most ReTweeted Tweets on this subject, it was possible to label 86% of the collected Tweets.

The following table provides a summary of the percentage of qualified Tweets for each of the fake news items studied:

TOPICS	PERCENTAGE OF QUALIFIED TWEETS
Greta Thunberg's activity on Facebook	82%
Coronavirus patent	62%
Coronavirus parcels	86%
Coronavirus in Argenteuil	100%
Coronavirus in Perpignan	100%
Scabies in Nevers hospital	100%
Gates and Coronavirus	55%
Netflix	64%
Sign board - Limoges	99%
Water pollution in Rouen	89%
Statements by Mr Blanquer	77%
Nurses' salaries	100%
TPMP	100%

More specifically, the categorisation consisted of classifying the selected Tweets into category "1" for those contributing to the spread of false information as identified by fact-checkers, and category "2" for humorous Tweets, category "3" for Tweets that contain information verified by fact-checkers and category "4" when the message is not related to the fake news analysed and was collected in error because it contained the selected keywords (these latter Tweets were then deleted from the qualified database).

This exercise sought, as far as possible, to rely on fact-checkers' articles and to avoid subjective classification. While in most cases, the nature of the Tweets made it easy to classify them, this operation was sometimes more delicate, especially in the case of Tweets adopting a humorous tone but sharing false information[72].

[72] Several fairly popular Tweets were categorised as contributing to the spread of fake news in that they approached them with humour and without questioning them. This was the case for a reused sequence from the animated television series

5.3. Analysis of the qualified Tweets database

The table below summarises the number of qualified Tweets for each subject (as a reminder, each ReTweet counts as a Tweet) as well as the distribution of these Tweets according to their categorisation:

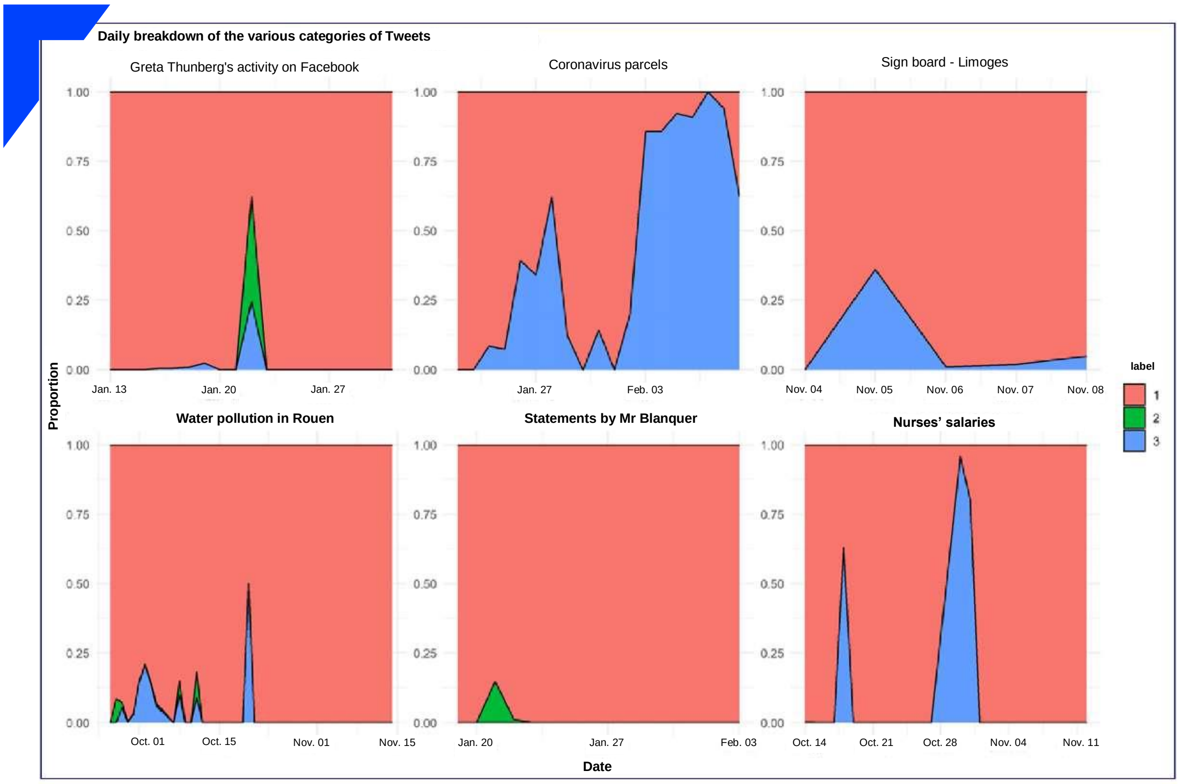
Topic	Number of qualified Tweets	Percentage of Tweets qualified as "1".	Percentage of Tweets qualified as "2".	Percentage of Tweets qualified as "3".
Greta Thunberg's activity on Facebook	4,967	99.19%	0.34%	0.46%
Coronavirus patent	416	31.73%	0.00%	68.27%
Coronavirus parcels	10,783	93.46 %	0.00%	6.54%
Coronavirus in Argenteuil	45	42.22%	6.67%	51.11%
Coronavirus in Perpignan	50	48.00%	0.00%	52.00%
Scabies in Nevers hospital	745	86.04%	2.28%	11.68%
Gates and Coronavirus	453	24.72%	0.00%	75.28%
Netflix	538	15.06%	0.00%	84.94%
Sign board - Limoges	23,708	98.62%	0.00%	1.38%
Water pollution in Rouen	90,048	95.24%	1.16%	3.60
Statements by Mr Blanquer	2,393	99.16%	0.84%	0.00%
Nurses' salaries	5,720	98.72%	0.02%	1.26%
TPMP	156	4.49 %	0.00 %	95.51 %

Irrespective of the importance of the information itself, it appears that the topics are divided more or less equally between those for which false information appears in the majority and those for which, conversely, verified information dominates. However, in the case of topics with a higher volume (more than 1,000 Tweets), fake news is in all cases in the majority, and in most cases even an ultra-majority. Tweets sharing verified information are unlikely to achieve very high levels of virality, while those with more significant virality are those who have achieved this virality through the widespread sharing of fake news.

While analysing the volume of Tweets according to their nature (sharing of false or proven information) provides interesting insights, it may also be useful to study the chronology of the sharing of different categories of Tweets. As pointed out above, one (optimistic) hypothesis would

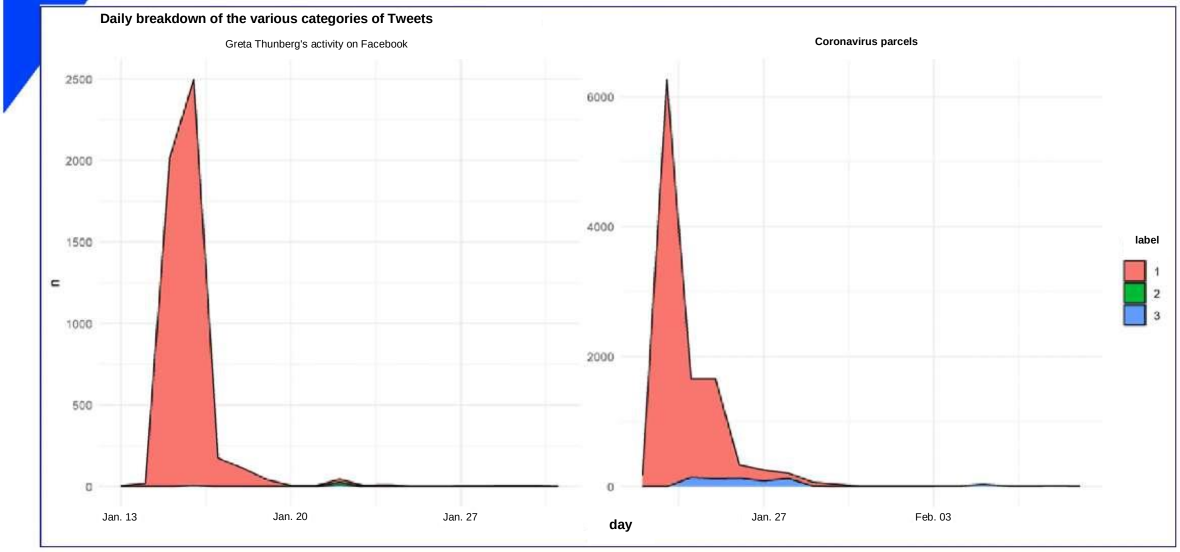
The Simpsons which "predicted" that parcels coming from China could be vectors of the virus, repeating a fake news story that circulated for a time on social media, yet without correcting it.

be that this chronology could be as follows. When a topic appears on Twitter as fake news, Tweets correct this false information and the presence of the fake news decreases, with Tweets containing the verified information becoming the majority. However, the chronologies observed, which vary from one subject to another, do not correspond to this hypothesis. The graphs below show the evolution of the distribution of the labels over time for the topics that have been shared the most, i.e. more than 1,000 Tweets (the graphs corresponding to the other topics are in the appendix):



It thus appears that it is only in the case of the topic of coronavirus transmission by parcel post that verified information finally becomes dominant. In other cases, the verified information experiences a brief growth peak before giving way again to fake news.

These graphs also depict a probably still over-optimistic view of the chronology of the spread of fake news: they highlight the change in the distribution of the categories of Tweets, disregarding the change in their volume in absolute value. The following graphs show the change in the number of Tweets in each category (and not the evolution of the distribution) for the examples of coronavirus transmission by parcel and Greta Thunberg's Facebook account. The other graphs are shown in the appendix.



With regard, for example, to the transmission of coronavirus by parcel post, in cases where genuine information ultimately assumes majority status on this subject, it is only when the subject has almost ceased to be discussed on Twitter. Similarly, in the case of Greta Thunberg's Facebook account, the spike in the proportion of verified information corresponds to a period when this topic was only receiving marginal coverage on Twitter.

Summary:

All the fake news studied shows a high concentration of Tweets over a very short time. Moreover, they are equally divided between topics where the majority of information shared is false and topics where, conversely, verified information is in the majority. However, for the topics with the highest volumes of Tweets, fake news is in the majority in all cases. A study of the chronology of the spread of fake news shows that, contrary to what might have been expected or hoped for, "genuine" news information does not displace fake news. For topics where fake news is in the majority, discussion on these topics quickly fades away on Twitter before the verified information can ever achieve a majority position.

5.4. Analysis of the potential audience for each category of Tweet

Previous analyses have made it possible to assess the volume of Tweets relating to false or proven information. However, these analyses are insufficient to assess the visibility of the different categories of Tweets as they do not take into account the number of subscribers to each account who have shared true or false information. The following table shows, for each topic analysed, the sum of the number of subscribers to each account who have posted a Tweet on this topic in the study's labelled database. This number of subscribers is then distributed among the accounts that have issued a Tweet categorised as 1 (fake news), 2 (humour) and 3 (verified information). It should be noted that this analysis is subject to a high margin of error: if an account with a large audience publishes verified information on the topic studied but this Tweet does not receive many ReTweets, it will not be selected for our labelling and will therefore not be counted.

Topic	Sum of the number of subscribers to accounts that have shared a Tweet on the subject[73]	Percentage of the number of subscribers exposed to a Tweet categorised as 1	Percentage of the number of subscribers exposed to a Tweet categorised as 2	Percentage of the number of subscribers exposed to a Tweet categorised as 3
Greta Thunberg's activity on Facebook	6,541,160	98.78 %	0.86 %	0.35%
Coronavirus patent	6,431,347	4.41 %	0.00 %	95.59%
Coronavirus parcels	6,396,890	82.99%	0.00%	17.01%
Coronavirus in Argenteuil	27,914	29.03%	4.95%	66.02%
Coronavirus in Perpignan	3,959,948	0.65%	0.00%	99.35%
Scabies in Nevers hospital	5,672,059	11.95%	0.55%	87.50%
Gates and Coronavirus	761,422	38.83%	0.00%	61.17%
Netflix	1,028,193	46.37%	0.00%	53.63%
Sign board - Limoges	11,244,122	97.34%	0.00%	2.66%
Water pollution in Rouen	56,654,501	80.88%	1.26%	17.86%
Statements by Mr Blanquer	2,618,692	99.82%	0.18%	0.00%
Nurses' salaries	4,821,826	67.90%	0.00%	32.10%
TPMP	264,733	0.35%	0.00%	99.65%

One would expect the audience share of reliable information to be higher than that of fake news since reliable information accounts that are likely to publish Tweets sharing proven information frequently have high audiences. However, although for certain topics the audience share of sites sharing verified information is actually higher than the volume share of these Tweets, some fake news remains the majority in terms of visibility (Greta Thunberg's Facebook account, transmission of coronavirus by parcels, etc.).

[73] Note that for each topic, the same subscriber can be counted several times if he or she subscribes to different accounts that have shared a Tweet on that topic.

Summary:

An analysis in terms of the potential audience for false and verified information reveals a greater share of verified information compared to an analysis based solely on the volume of Tweets. However, for some topics, false information is still more visible than verified information.

5.5. Analysis of correction times

The following table shows the time elapsed between the first fake news item recorded in the labelled database and the first verified information on the same topic. Once again, this analysis, which is based on the labelled database, only considers the most ReTweeted Tweets, and information that is not widely shared is not taken into account, which can lead to an overestimation of the correction time.

TOPIC	DATE OF FIRST FAKE NEWS	DATE OF FIRST VERIFIED INFORMATION	TIME ELAPSED BETWEEN FAKE AND GENUINE INFORMATION
Greta Thunberg's activity on Facebook	13/01/2020 at 22:03:31	16/01/2020 at 14:50:45	64.79 hours
Coronavirus patent	22/01/2020 at 05:48:34	24/01/2020 at 16:42:18	58.90 hours
Coronavirus parcels	22/01/2020 at 10:19:28	24/01/2020 at 14:35:00	52.26 hours
Coronavirus in Argenteuil	26/01/2020 at 13:27:35	26/01/2020 at 23:09:44	9.70 hours
Coronavirus in Perpignan	26/01/2020 at 17:50:44	26/01/2020 at 13:46:18	-4.07 hours
Scabies in Nevers hospital	15/10/2019 at 21:53:39	16/10/2019 at 15:44:13	17.84 hours
Gates and Coronavirus	24/01/2020 at 19:59:38	29/01/2020 at 19:20:54	119.35 hours
Netflix	21/10/2019 at 08:32:33	21/10/2019 at 10:07:26	1.58 hours
Sign board - Limoges	04/11/2019 at 22:53:48	05/11/2019 at 09:17:21	10.39 hours
Water pollution in Rouen	26/09/2019 at 06:34:20	28/09/2019 at 05:26:15	46.87 hours
Statements by Mr Blanquer	19/01/2020 at 15:00:26	-[74]	-
Nurses' salaries	14/10/2019 at 09:05:10	15/10/2019 at 00:35:06	15.50 hours
TPMP	25/11/2019 at 14:49:35	25/11/2019 at 15:54:33	1.08 hours

Fake news appears on the Twitter network in a fairly natural way, usually before later being refuted (except for the case of Coronavirus in Perpignan[75]). In some cases, the rebuttal appears

[74] Lack of Tweet correcting false information in the database.

very quickly, as is the case with the cancellation of the TPMP broadcast. In other cases, the periods appear to be very significant, especially taking into account the sometimes very limited lifespan of discussions on a given topic. These findings, and the findings presented above, illustrate the difficulty of the work of fact-checkers. If they wish to influence the spread of fake news on Twitter, they have to react extremely quickly since fake news will most often be shared over a very short period, while at the same time they are subject to the traditional constraints of journalistic work of fact-checking sources, which necessarily takes time, and their work may be based on identifying fake news that has already achieved certain popularity online.

From this point of view, the partnerships implemented by certain social media with[76] fact-checkers to display corrections soon after the detection of the fake news, in order to limit its viral spread, will be all the more likely to produce a significant effect if these corrections are added very quickly after the publication of the Tweet.

5.6. Analysis of the most ReTweeted messages

To better understand the nature of Tweets propagating or debunking fake news, a qualitative analysis was carried out, for each topic, on the two most shared Tweets containing verified information and the two most shared Tweets containing unverified information[77], whether or not they came from media accounts. The result is a corpus of 48 Tweets, equally divided between incorrect and corrective messages (24 Tweets each time).

Among these 24 Tweets, the **main erroneous messages** are based on fake news of various types: they may be either unproven or incomplete facts, or taken out of context in a way that changes their meaning. They **express direct criticism (16), or a sense of panic (3) or humour (2).**[78] In one single case, described at the end of this section, the false information was published by media with "verified" status on Twitter (2).

[75] It seems that in this case, the fake news first appeared on the Facebook network, was later debunked by the local press and this rebuttal was shared by the press headline on Twitter: https: //www.lindependant.fr/2020/01/26/coronavirus-non-aucun-cas-na-ete-detecte-a-perpignan,8688548.php. The false information then appeared on Twitter after the verified information.

[76] For example, in the case of Facebook: https://www.facebook.com/business/help/182222309230722

[77] The corpus should consist of 52 Tweets (4 per topic with 13 topics under study) labelled half as false (1) and half as comprising verifications or shared journalistic fact-checking (3) among the most ReTweeted, but for two topics studied, no Tweets labelled 1 and 3 respectively were ReTweeted enough to be included in our analysis (n = 48).

[78] A Tweet simply repeats an article that was later categorised as false by the fact-checking units of *Libération* and *franceinfo*, with no text other than the title and link. It is therefore not commented on in this analysis. This therefore gives us n= 24.

Direct accusations are in the majority in the corpus studied. Targets include members of the government (*"#Blanquer thinks everything's OK because 99.9% of the teachers support him"*), elected representatives ("*#Nevers hospital is in quarantine following a scabies epidemic due to the hospitalisation of #migrants Omerta from the #LREM mayor of [...]"*) and personalities (*"A bug on Facebook reveals that the messages posted by Greta Thunberg are written by her father... Only idiots could think that this school dropout could write her own messages... she should sit her baccalaureate exams before giving lessons to the whole world"*). More general references (*"they"*, *"yeah, right"*) seem to accuse the authorities of lying, without naming them: this is, for example, the case of the Tweets relating to the fire at the Lubrizol factory in Rouen (*"But the good news is, they told us the water is drinkable in Rouen"*). Others do it in a roundabout way using, for example, the expression *“coincidence?*. These accusations are mostly based on unchecked testimonies, in the form of pictures or text, or on facts that have been described as inaccurate or false by journalists.

The global health crisis linked to the Covid-19 epidemic seems to be particularly conducive to the sharing of Tweets expressing a sense of panic. These take two main forms: either they are very emotional with the use of capital letters combined with interrogative forms and negative emoticons (*"Did the Gates Foundation actually finance 'the group that holds the patent' for the coronavirus? I'm afraid it's either * THE TRUTH *”*), or they're compiling facts and figures to support their point (*"U.S. Patent for the coronavirus patent (Patent No. 10,130,701 issued 20 November 2018)”*) even though these had been described as false by fact-checking journalists.

In addition, two Tweets were labelled as contributing to the spread of fake news in that they approached them with humour and without questioning them. Of the four categories in this analysis, these are the ones with the highest mean average share (29,511 ReTweets in total, including 21,600 for the most shared of the two). This was the case for a reused sequence from the animated television series *The Simpsons* which "*predicted*" that parcels coming from China could be vectors of the virus, repeating a fake news story that circulated for a time on social media, yet without correcting it.

Finally, the only two Tweets published by the media that spread false information deal with the same topic: Netflix's alleged announcement that they were stepping up the fight against account sharing. The words of the service's product manager were echoed in one of the Tweets in support of this assertion, although it is mentioned that this was *“implied”*. Links to the sites of these media accompany these Tweets. It seems here that these two media reused information without cross-checking it, leading to denials published by colleagues ("*nobody said that*", Tweeted a colleague from *Numerama*). Between the two of them, there are only 45 Tweets.

Concerning the **corrective Tweets, half of them come from so-called "certified" accounts**[79] (12 from 11 media sources, including units specialising in fact-checking information and 1 institutional) **and the other half from uncertified accounts**. Some of these are accounts that describe themselves as media accounts (4), while others appear to be more personal accounts (8).

The first ones alternate between repeating phrases from the false information in interrogative form before including a link to their fact-checking article ("*Is Nevers hospital locked down because of a scabies epidemic carried by migrants?* ", *Franceinfo*, 17 October 2019). Others are closer to the specific formats of Twitter, for example by being more discreet and giving detailed information directly in their Tweets (*"□ The message displayed on the digital sign on rue Jean-Jaurès is a dummy message that was posted yesterday from 1pm to 4pm and will be repeated at the same time today for the purposes of a feature film currently being shot in the city centre of Limoges* □", City of Limoges, 5 November 2019) or by creating a thread made up of several Tweets - thread (*"[Breaking] Explosions, dead birds, air quality and drinking water... Legitimate concerns about the consequences of the fire at the Lubrizol factory in Rouen have been the breeding ground for many distortions*. *Here are our tips for checking certain statements ⬇", AFP Factuel*, 1st October 2019). These Tweets also include a large number of ReTweets (1,653 in all, 138 per Tweet on average), half of them (810 ReTweets) from AFP Factuel's post alone, which has just been cited.

With regard to the corrections made by uncertified accounts, a qualitative analysis of the corrective Tweets they have published highlights certain common features. Several cite so-called historical media or authoritative sources (*"The CSA has just denied it, calling it a hoax, as has @Cyrilhanouna just now live on air.* "), or deconstruct the origins of the false information (*"You copied this from the satirical newspaper @Nordpresse (cousin of #Gorafi)"*). The Tweets which seem to have come from personal accounts seek either to produce summaries concerning fake news circulating on a given subject (*"Some Internet theories on the Coronavirus: [...]")* or, in one case, to criticise the correction made by a verified media outlet (*"France Info, which reposted my Tweet and labelled it pseudo+photo without warning me [...]"*). These Tweets focus on fewer ReTweets (1,179 in all, 98 per Tweet on average), the most widely distributed being from an account appearing as a media outlet (*Brèves de presse*, 610 ReTweets).

On the whole, the analysis of the most shared Tweets in terms of disinformation highlights the very important role played by the traditional media in the fight against disinformation. In the majority of cases, it is more specifically the units specialising in fact-checking that are able to obtain the greatest visibility. However, this success is only partial, since the Tweets disseminated by these fact-checkers rarely manage to reach the degree of virality of Tweets sharing fake news.

[79] Twitter affixes a badge next to the name of some accounts consisting of a white checkmark on a blue background to "provide a guarantee to users that an account of public interest is authentic"; for example, in the field of politics, media and public authorities. However, Twitter states that this badge "in no way implies a recommendation" from the social networking service. Twitter, A propos des comptes certifiés, Centre d'assistance (*About certified accounts, Help Desk*, page consulted on 30 April 2020). URL: https://help.twitter.com/fr/managing-your-account/about-twitter-verified-accounts

Summary:

An analysis of some fake news items reveals several interesting characteristics. On the one hand, such content, when not produced by fact-checkers, is often difficult to qualify: the status of humour thus raises questions in how Tweets are interpreted. Such fake news is also mainly a vehicle for criticising the authorities or expressing a sense of panic; for example, on sensitive health issues. It is based on information that has not been fact-checked or has been described as false by journalists, who in turn make corrections using practices specific to social media. The low level of engagement (ReTweet, comments) for these publications in relation to the accounts likely to disseminate fake news seems to indicate that this method is relatively successful. It should nevertheless be noted that in several cases, the corrections were made directly by Internet users, based on the elements of these media articles, showing that this work is reaching its intended target.

6. Conclusion

This study showed that the least reliable news accounts on Twitter have significantly fewer subscribers than the majority of reliable news accounts. However, in terms of "ReTweets", accounts that are known to share fake news are on an equal footing with reliable accounts. Subscribers to unreliable accounts have a much higher propensity than subscribers to reliable accounts to contribute to the dissemination of information shared via these accounts, with 10 to 20 times more ReTweets per subscriber, depending on the indicator used.

The analysis also showed that the unreliable accounts focus strongly on current issues and divisive topics. Tweets dealing with politics, immigration, health, religion and terrorism represent more than half of the corpus studied. The quantitative analysis shows an over-representation of terms related to delinquency, immigration, Israel and Palestine, paedophilia, Islam and freemasonry. These accounts make extensive use of non-text content (images, videos, links, etc.) and frequently link to the websites linked to the accounts, but also to sources from traditional media.

Individuals who follow only unreliable accounts are in the minority on Twitter: under 20% of subscribers to these accounts do not follow any reliable account. A graph-based analysis also shows that many unreliable accounts have a significant proportion of common subscribers or follow each other on Twitter.

On average, the accounts of journalism units specialising in fact-checking have more followers than accounts categorised as unreliable, but their publications generate fewer interactions than the latter accounts. In terms of the topics generating the most engagement, only politics and religion are common to these two types of accounts.

Finally, the study presents an analysis of Tweets relating to fake news identified as such by journalists. All the fake news in the study shows a high concentration of Tweets over a very short period. Moreover, the fake news in the study is evenly divided between topics for which most of the information shared is fake news, and topics for which, conversely, most of the information shared is verified. However, for the topics with the highest number of Tweets, fake news is in all cases in the majority. A study of the chronology of the spread of fake news shows that "genuine" news information does not displace fake news. For topics where fake news is in the majority, discussion on these topics quickly fades away on Twitter before the verified information can ever achieve a majority position.

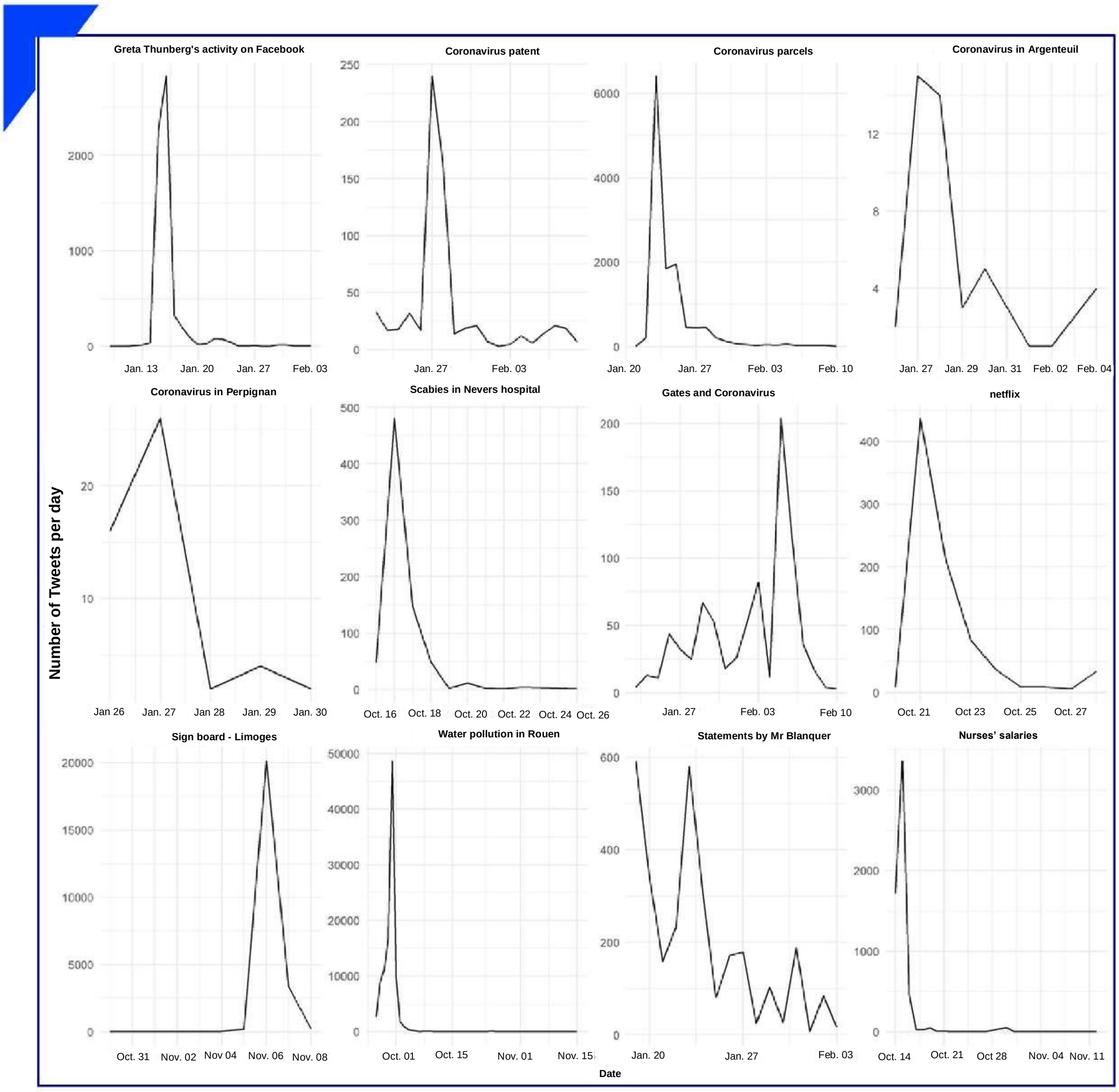
By analysing the most viral Tweets in the corpus under study, we can highlight several interesting features. Fake news is mainly a vehicle for criticising the authorities or expressing a sense of panic; for example, on sensitive health issues. It is based on information that has not been fact-checked or has been described as false by journalists, who in turn make corrections using practices specific to social media. However, these corrective measures have a low level of engagement (ReTweets, comments) compared to accounts that are likely to disseminate fake news.

While this study has focused on the virality of fake news on a social media platform, the context of its publication is a strong reminder that social media can facilitate the sharing of other forms of harmful content, such as those inciting hatred and violence. Understanding how different forms of content can be disseminated on the platforms will therefore remain a central issue for the *Conseil*.

7. Annexes

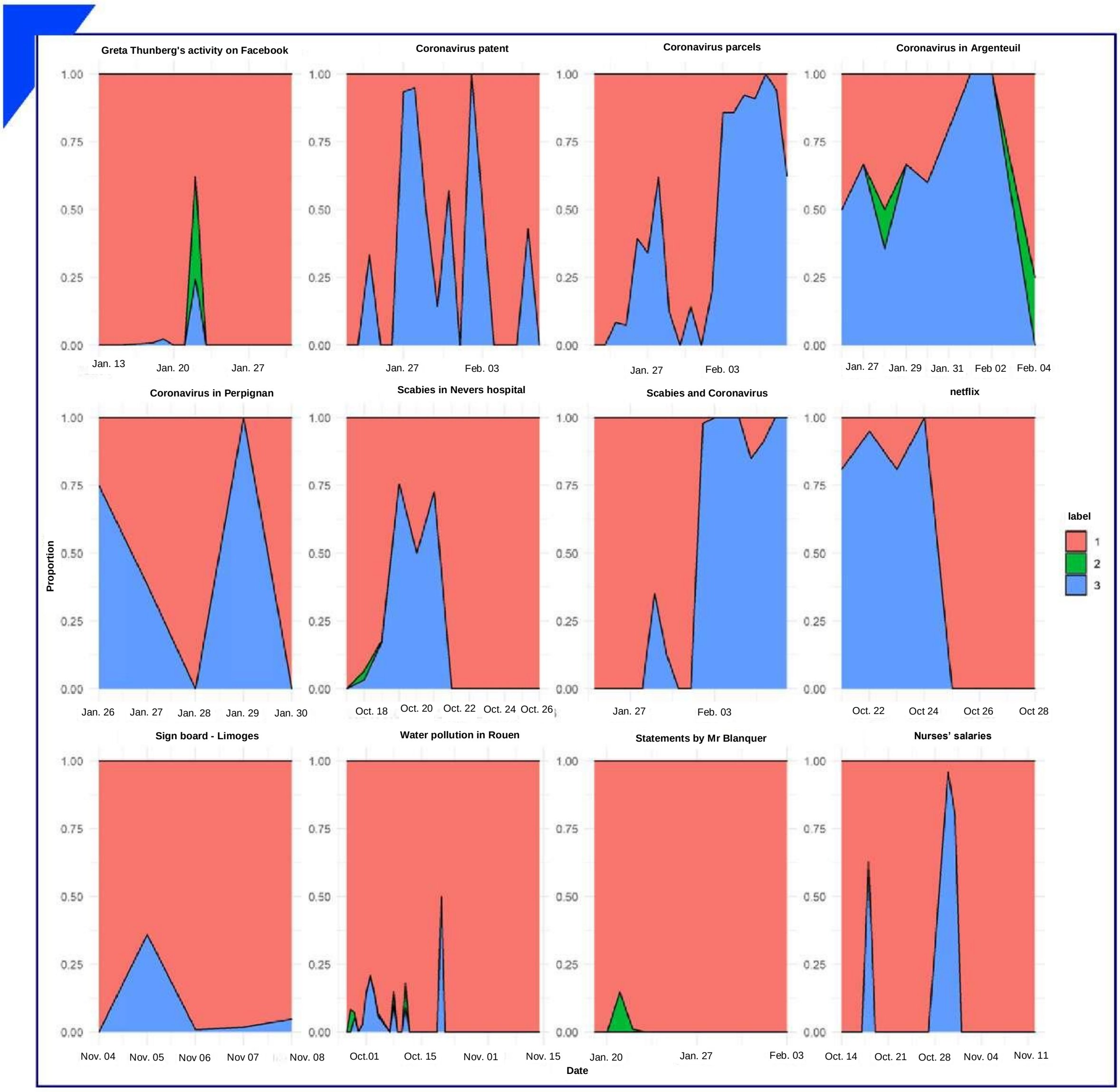
7.1. Chronology of the circulation of Tweets

Change in the number of Tweets



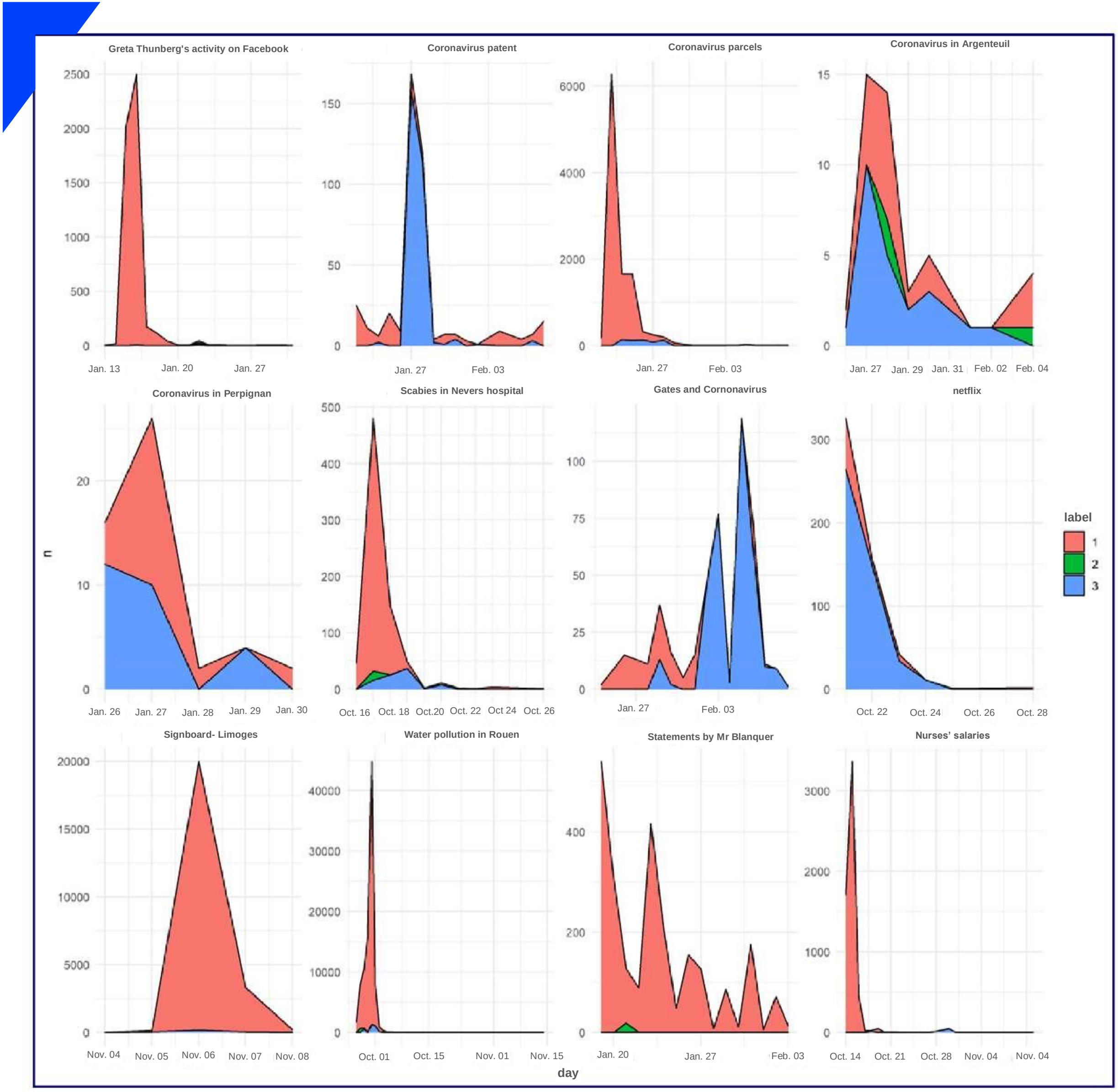
7.2. Change in the distribution of Tweets by category

Daily breakdown of the different categories of Tweets



7.3. Change in the volume of Tweets by category

Daily breakdown of the different categories of Tweets



8. Bibliography

Legal sources

Law no. 2018-1202 of 22 December 2018 relating to the fight against information manipulation. Official Journal of the French Republic, 23/12/2018. URL: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000037847559&categorieLien=id. .

Recommendation no. 2019- 03 of 15 May 2019 of the Conseil supérieur de l'audiovisuel to online platform operators as part of the duty to co-operate in the fight against the dissemination of fake news. Official Journal of the French Republic, 17/5/2019. URL: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000038480745&categorieLien=id.

Decree no. 2019-297 of 10 April 2019 on the information obligations of online platform operators promoting information content related to a debate of general interest. Official Journal of the French Republic, 11/4/2019. URL: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000038359165&categorieLien=id. ..

The Digital Services Act package. (European Commission), URL: https: //ec.europa.eu/digital-single-market/en/digital-services-act-package. Accessed on: 11/06/2020.

Scientific literature

Hunt Allcott, Matthew Gentzkow. *Social media and fake news in the 2016 election*. Journal of Economic Perspectives 2: 1-28 2017.

Dimitrios Bountouridis et al. *Annotating credibility: Identifying and mitigating bias in credibility datasets*. 25/07/2019.

Peter Burger, Soeradj Kanhai, Alexander Pleijter, Suzan Verberne. *The Reach of Commercially Motivated Junk News on Facebook*. PLOS ONE 14(8): e0220446. 01/08/2019.

Cardon, Dominique, Ooghe-Tabanou, Benjamin, Plique, Guillaume, Cointet, Jean-Philippe. *Les nouveaux circuits de l'information numérique (The new digital information channels)*. SciencesPo Médialab. 11/06/2019. URL: https://medialab.sciencespo.fr/en/news/les-nouveaux-circuits-de-linformation-numerique. Accessed on: 28/10/2019.

Ted Dunning. *Accurate Methods for the Statistics of Surprise and Coincidence*. Computational Linguistics 19(1): 61-74 1993.

Emilio Ferrara. *Disinformation and social bot operations in the run up to the 2017 French presidential election*. First Monday 22(8). 31/07/2017. URL: http://journals.uic.edu/ojs/index.php/fm/article/view/8005. Accessed on: 24/10/2019.

Richard Fletcher, Rasmus Kleis Nielsen. *Are people incidentally exposed to news on social media? A comparative analysis*. New Media &amp; Society 20(7). 17/08/2017. URL: https://journals.sagepub.com/doi/abs/10.1177/1461444817724170. Accessed on: 04/06/2020.

Lucas Graves. *A smarter conversation about how (and why) fact-checking matters*. Nieman Lab. 12/2019 URL: https://www.niemanlab.org/2019/12/a-smarter-conversation-about-how-and-why-fact-checking-matters/. Accessed on: 04/06/2020.

Andrew Guess, Jonathan Nagler, Joshua A. Tucker. *Less than you think: Prevalence and predictors of fake news dissemination on Facebook*. Science Advances 5(1). 09/01/2019.

Andrew Guess, Brendan Nyhan, Benjamin Lyons, Jason Reifler. *Avoiding the Echo Chamber about Echo Chambers: Why Selective Exposure to like-Minded Political News Is Less Prevalent than You Think*. Knight Foundation White Paper. 2018.

Andrew Guess, Brendan Nyhan, Jason Reifler. *Selective exposure to fake news: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign*. 09/01/2018. URL: https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf. .

Kristoffer Holt, Tine Ustad Figenshou, Lena Frischlich. *Key dimensions of alternative news media*. Digital Journalism 7(7): 860-69. 2019.

Edda Humprecht. *How Do They Debunk "Fake News"? A Cross-National Comparison of Transparency in Fact Checks*. Digital Journalism. 14/10/2019. URL: https://www.tandfonline.com/doi/abs/10.1080/21670811.2019.1691031. Accessed on: 04/06/2020.

David M.J. Lazer et al. *The science of fake news*. Science 359(380): 1094-96. 2018.

Briony Swire, Ullrich K.H. Ecker, Stephan Lewandowsky. *The role of familiarity in correcting inaccurate information*. Journal of ExperimentalPsychology: Learning, Memory and Cognition 43(12): 1948-61. 12/2017

Soroush Vosoughi, Deb Roy, Sinan Aral. *The spread of true and false news online*. Science 359: 1146-51. 09/03/2018.

Reports

A Multi-Dimensional Approach to Disinformation, Report of the Independent High Level Group on Fake News and Online Disinformation. 3/2018

Digital news report 2019. 2019.

Richard Fletcher, Alessio Cornia, Lucas Graves, Rasmus Kleis Nielsen. *Measuring the reach of "fake news" and online disinformation in Europe*. Factsheet, 2/2018. URL: https://www.press.is/static/files/frettamyndir/reuterfake.pdf. .

Divina Frau-Meigs. *Societal costs of "fake news" in the Digital Single Market*. 12/2018 URL: https://sanef.org.za/wp-content/uploads/2019/02/European-Parliament-Societal-costs-of-fake-news-December-2018.pdf. .

Lucas Galan, Jordan Osserman, Tim Parker, Matt Taylor. *How young people consume news and the implications for mainstream media*. 8/2019 URL: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-08/FlamingoxREUTERS-Report-Full-KG-V28.pdf. Accessed on: 24/10/2019.

J.-B. Jeangène Vilmer, A. Escorcia, M. Guillaume, J. Herrera. *Les Manipulations de l'information : un défi pour nos démocraties* (*Information Manipulations : a challenge for our democracies)*. 8/2018 URL: https://www.diplomatie.gouv.fr/IMG/pdf/les_manipulations_de_l_information_2__cle04b2b6.pdf. Accessed on: 12/06/2020.

Joshua A. Tucker et al. *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*. 3/2018 URL: http://eprints.lse.ac.uk/87402/1/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf. .

Claire Wardle, Hossein Derakhshan. *Information disorder: Toward an interdisciplinary framework for research and policy making*. 27/09/2017. URL: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

Encyclopaedia article

Fact checking (France section). Wikipedia. URL: https://fr.wikipedia.org/wiki/V%C3%A9rification_des_faits#France. Accessed on: 28/10/2019.

Press

20 Minutes. Coronavirus: Non, aucun cas n'a été détecté à Perpignan, indique la préfecture (No, no cases were detected in Perpignan, says the prefecture). 26/01/2020. URL: https://www.20minutes.fr/sante/2703999-20200126-coronavirus-non-aucun-cas-detecte-perpignan-indique-prefecture. Accessed on: 23/07/2020.

20 Minutes. Coronavirus: Non, le Covid-19 n'a pas été breveté par l'Institut Pasteur en 2004 (No, Covid-19 was not patented by the Pasteur Institute in 2004). 18/03/2020. URL: https://player.acast.com/5ab28781627e02027cf860fe/episodes/5e68ec5b7936a86027281fb6?ref=facebook. Accessed on: 23/07/2020.

20 Minutes. " Touche pas à mon poste " cancelled? Denial by Cyril Hanouna. 26/11/2019. URL: https://www.20minutes.fr/arts-stars/television/2660459-20191126-touche-poste-deprogrammee-cyril-hanouna-dement. Accessed on: 23/07/2020.

AFP Factual. *Australie : les causes des incendies, mine de désinformation sur les réseaux sociaux (Australia: the causes of the fires, a mine of disinformation on social networks)*. 13/01/2020.

URL: https://factuel.afp.com/australie-les-causes-des-incendies-mine-de-desinformation-sur-les-reseaux-sociaux. Accessed on: 17/01/2020.

AFP Factual. *Nouveau coronavirus : épidémie mondiale de fausses informations (Newcoronavirus: Global epidemic of fake news)*. 28/01/2020. URL: https://factuel.afp.com/nouveau-coronavirus-epidemie-mondiale-de-fausses-informations. Accessed on: 31/01/2020.

Axios. *Big Tech — and China — Struggle with Coronavirus Misinformation*. 28/01/2020. URL: https://www.axios.com/coronavirus-misinformation-facebook-twitter-google-china-246a0325-b4ea-4465-92ae-5f364a7e965c.html. Accessed on: 03/02/2020.

Bloomberg. *Sandberg Says Facebook 'Filter Bubble' Problem Is Misunderstood*. 22/10/2019. URL: https://www.bloomberg.com/news/articles/2019-10-22/sandberg-says-facebook-filter-bubble-problem-is-misunderstood. Accessed on: 24/10/2019.

BuzzFeed. *Facebook Won't Do Anything To Stop Climate Deniers Capitalising On Australia's Bushfire Crisis*. 22/01/2020. URL: https://www.buzzfeed.com/hannahryan/facebook-australia-bushfires-climate-change-deniers-facebook. Accessed on: 27/01/2020.

franceinfo. *#VraiOuFake Un colis en provenance de Chine peut-il transmettre le coronavirus ? (#RealOrFake Can a package from China carry*coronavirus?) 28/01/2020. URL: https://www.francetvinfo.fr/sante/maladie/coronavirus/coronavirus-un-colis-en-provenance-de-chine-peut-il-transmettre-la-maladie_3801681.html. Accessed on: 23/07/2020.

franceinfo. *Coronavirus : quels sont les comptes Twitter « super propagateurs » de fausses informations en France (Coronavirus: which Twitter accounts are "super spreaders" of fake news in France?)* 04/06/2020. URL: https://www.francetvinfo.fr/sante/maladie/coronavirus/coronavirus-quels-sont-les-comptes-twitter-super-propagateurs-de-fausses-informations-en-france_3982961.html. Accessed on: 05/06/2020.

franceinfo. *Netflix : la rumeur sur la suppression du partage de compte fait son grand (Netflix: Rumour about the removal of account sharing is back*. 23/10/2019. URL: https://www.francetvinfo.fr/culture/series/netflix/netflix-la-rumeur-sur-la-suppression-du-partage-de-compte-fait-son-grand-retour_3671539.html. Accessed on: 23/07/2020.

franceinfo. *Une infirmière à domicile est-elle payée moins par patient que le prix d'un "Happy Meal" ? (Is a home care nurse paid less than the price of a "Happy Meal" per patient?)* 18/10/2019. URL: https://www.francetvinfo.fr/sante/politique-de-sante/une-infirmiere-a-domicile-est-elle-payee-moins-par-patient-que-le-prix-dun-happy-meal_3664829.html. Accessed on: 23/07/2020.

L'ADN. *Nous vivons une véritable épidémie de fake news et il n'y a pas vraiment de remède (We are experiencing a real epidemic of fake news and there is no actual cure)*. 29/11/2019. URL: https://www.ladn.eu/media-mutants/reseaux-sociaux/fakes-news-continuent-progression-reseaux-sociaux/. Accessed on: 04/06/2020.

L'ADN. *Un outil pour comprendre l'orientation politique des médias (A tool for understanding the political orientation of the media)*. 21/01/2020. URL: https://www.ladn.eu/media-mutants/reseaux-sociaux/comment-echapper-bulles-filtre-reseaux-sociaux/. Accessed on: 22/01/2020.

L'Indépendant. *Coronavirus: no, no cases have been detected in Perpignan*. 26/01/2020. URL: https://www.lindependant.fr/2020/01/26/coronavirus-non-aucun-cas-na-ete-detecte-a-perpignan,8688548.php. Accessed on: 23/07/2020.

Le Figaro. *Facebook a franchi la barre des «37 millions d'utilisateurs» en France en 2019 (Facebook crossed the "37 million users" mark in France in 2019)*. 10/02/2020. URL: https://www.lefigaro.fr/secteur/high-tech/facebook-a-franchi-la-barre-des-37-millions-d-utilisateurs-en-france-en-2019-20200210. Accessed on: 23/07/2020.

Le Monde. "*« Mesdames, évitez de rentrer seule le soir… » : le message factice de la ville de Limoges ("Ladies, don't go home alone in the evening…": the dummy message from the city of Limoges)*. 12/11/2019. URL: https://www.lemonde.fr/les-decodeurs/article/2019/11/12/mesdames-evitez-de-rentrer-seule-le-soir-le-message-factice-de-la-ville-de-limoges_6018869_4355770.html. Accessed on: 23/07/2020.

Le Monde. *Après l'incendie de l'usine Lubrizol à Rouen, des intox en série (After the fire at the Lubrizol factory in Rouen, a series of hoaxes)*. 30/09/2019. URL: https://www.lemonde.fr/les-decodeurs/article/2019/09/30/oiseaux-morts-eau-non-potable-communique-bidon-intox-en-serie-apres-l-incendie-de-rouen_6013661_4355770.html. Accessed on: 23/07/2020.

Le Monde. *Coronavirus: Bill Gates targeted by rumours and conspiracy theorists*. 05/02/2020. URL: https://www.lemonde.fr/les-decodeurs/article/2020/02/05/coronavirus-bill-gates-cible-par-des-rumeurs-et-infox-complotistes_6028482_4355770.html. Accessed on: 23/07/2020.

Le Monde. *Décodex : vérification de sources d'informations, pages Facebook et chaînes YouTube(Décodex: Checking news sources, Facebook pages and YouTube channels*. URL: https://www.lemonde.fr/verification/. Accessed on: 23/07/2020.

Le Monde. *L'annuaire des sources du Décodex : mode d'emploi (Décodex directory of sources:user manual)*. 23/1/2017.a. URL: https://www.lemonde.fr/les-decodeurs/article/2017/01/23/l-annuaire-des-sources-du-decodex-mode-d-emploi_5067719_4355770.html. Accessed on: 23/07/2020.

Le Monde. *Le coronavirus qui sévit en Chine n'a pas été « créé en 2003 aux USA » (The coronavirus ravaging China was not "created in 2003 in the USA"*). 27/01/2020. URL: https://www.lemonde.fr/les-decodeurs/article/2020/01/27/le-coronavirus-qui-sevit-en-chine-n-a-pas-ete-cree-en-2003-aux-usa_6027390_4355770.html. Accessed on: 23/07/2020.

Le Monde. *Le Décodex, un outil de vérification de l'information (The Décodex: a fact-checking tool)*. 23/1/2017.b. URL: https://www.lemonde.fr/les-decodeurs/article/2017/01/23/le-decodex-un-premier-premier-pas-vers-la-verification-de-masse-de-l-information_5067709_4355770.html. Accessed on: 12/06/2020.

Le Monde. *Incendie à Rouen : intox en série (Fire in Rouen:a series of hoaxes)*. 30/09/2019. URL: https://www.lemonde.fr/les-decodeurs/article/2019/09/30/oiseaux-morts-eau-non-potable-communique-bidon-intox-en-serie-apres-l-incendie-de-rouen_6013661_4355770.html. Accessed on: 23/07/2020.

Le Monde. *Les fausses informations circulent de moins en moins sur Facebook (The circulation of fake news is declining on Facebook)*. 17/10/2018. URL: https://www.lemonde.fr/les-decodeurs/article/2018/10/17/les-fausses-informations-perdent-du-terrain-sur-facebook_5370461_4355770.html. Accessed on: 04/06/2020.

Les Echos. *Coronavirus : un rapport épingle la persistance des « fake news » sur Twitter (Coronavirus: a report highlights the persistence of " fake news " on Twitter)*. 20/05/2020. URL: https://www.lesechos.fr/tech-medias/hightech/coronavirus-un-rapport-epingle-la-persistance-des-fake-news-sur-twitter-1204358. Accessed on: 20/05/2020.

Libération. *About - Checknews*. URL: https://www.liberation.fr/checknews/about/. Accessed on: 28/10/2019.

Libération. *Did Blanquer actually say that "99.9% of teachers support the reform of baccalaureate qualification"?* 23/01/2020. URL: https://www.liberation.fr/checknews/2020/01/23/blanquer-a-t-il-vraiment-dit-que-999-des-enseignants-soutiennent-la-reforme-du-bac_1774760. Accessed on: 23/07/2020.

Libération. *Brèves de presse, Alertes infos, Les News... : who is behind these news accounts on Twitter?* 16/01/2020. URL: https://www.liberation.fr/checknews/2020/01/16/breves-de-presse-alerte-infos-les-news-qui-est-derriere-ces-comptes-d-info-sur-twitter_1770396. Accessed on: 04/06/2020.

Libération. *Un bug sur Facebook a-t-il révélé que les messages de Greta Thunberg étaient écrits par son père ? (Did a bug on Facebook reveal that Greta Thunberg's messages were written by her father?)* 16/01/2020. URL: https://www.liberation.fr/checknews/2020/01/16/est-il-vrai-qu-un-bug-sur-facebook-a-revele-que-les-messages-de-greta-thunberg-etaient-ecrits-par-so_1773407. Accessed on: 23/07/2020.

Numerama. *Conflicts_FR : qui est derrière le compte Twitter qui partage des informations sur le coronavirus sans les vérifier (Conflicts_FR: Who is behind the Twitter account that shares information on coronavirus without checking it?)* 16/03/2020. URL: https://www.numerama.com/politique/611424-conflits_fr-qui-est-derriere-le-compte-twitter-qui-partage-des-informations-sur-le-coronavirus-sans-les-verifier.html. Accessed on: 24/03/2020.

Siècle Digital. *L'histoire de la propagande chinoise sur Twitter pendant l'épidémie du Covid-19 (The history of Chinese propaganda on Twitter during the Covid-19 outbreak)*. 31/03/2020. URL: https://siecledigital.fr/2020/03/31/lhistoire-de-la-propagande-chinoise-sur-twitter-pendant-lepidemie-du-covid-19/. Accessed on: 06/04/2020.

Slate. *Twitter's New Order*. 05/03/2017. URL: https://www.slate.com/articles/technology/cover_story/2017/03/twitter_s_timeline_algorithm_and_its_effect_on_us_explained.html?via=gdpr-consent. Accessed on: 29/04/2020.

The Conversation. *Rating News Sources Can Help Limit the Spread of Misinformation*. 02/12/2019. URL: http://theconversation.com/rating-news-sources-can-help-limit-the-spread-of-misinformation-126083. Accessed on: 04/06/2020.

The Verge. *Google's ads just look like search results now*. 23/01/2020. URL: https://www.theverge.com/tldr/2020/1/23/21078343/google-ad-desktop-design-change-favicon-icon-ftc-guidelines. Accessed on: 27/01/2020.

Wired. *How One Particular Coronavirus Myth Went Viral*. 19/03/2020. URL: https://www.wired.com/story/opinion-how-one-particular-coronavirus-myth-went-viral/. Accessed on: 24/03/2020.

Wired. *YouTube Gaming's Most-Watched Videos Dominated by Scams and Cheats*. 10/02/2020. URL: https://www.wired.com/story/youtube-gaming-scams-cheats-livestreams/. Accessed on: 25/02/2020.

Other sources

About. Twitter. URL: https://about.twitter.com/fr.html. Accessed on: 23/07/2020.

About verified accounts. Twitter. URL: https://help.twitter.com/fr/managing-your-account/about-twitter-verified-accounts. Accessed on: 23/07/2020.

Ce que les éditeurs doivent savoir à propos de la vérification des informations par des tiers sur Facebook | Pages d'aide de Facebook Business (What Publishers Need to Know About Third-Party Fact Checking on Facebook | Facebook Business Help Pages). Facebook. URL: https://fr-fr.facebook.com/business/help/182222309230722. Accessed on: 23/07/2020.

Données personnelles : le CSA mène une étude sur le phénomène de propagation des fausses informations sur les réseaux sociaux (Personal data: The CSA is studying the phenomenon of the spread of fake news on social media). Conseil supérieur de l'audiovisuel. 13/09/2019. URL: https://www.csa.fr/Informer/Toutes-les-actualites/Actualites/Donnees-personnelles-le-CSA-mene-une-etude-sur-le-phenomene-de-propagation-des-fausses-informations-sur-les-reseaux-sociaux. Accessed on: 23/07/2020.

Direction générale de l'enseignement scolaire (French General Directorate of School Education). *Infographie sur les outils de vérification des faits (Infographics on fact-checking tools) (March 2020)*. Twitter. 27/03/2020. URL: https://twitter.com/eduscol_emi/status/1243472130441789440. Accessed on: 06/04/2020.

IFCN Covid-19 Misinformation. Poynter. URL: https://www.poynter.org/ifcn-covid-19-misinformation/. Accessed on: 06/04/2020.

La carte des théories du complot sur le coronavirus (The map of conspiracy theories on the coronavirus). Conspiracy Watch. 23/03/2020. URL: https://www.conspiracywatch.info/la-carte-des-theories-du-complot-sur-le-coronavirus.html. Accessed on: 06/04/2020.

Misinformation Research - Airtable Universe. Airtable. URL: https://airtable.com/universe/expPeddCpX0wOeNNE/misinformation-research. Accessed on: 04/06/2020.

NewsGuard : l'outil de confiance sur Internet (NewsGuard: the trusted tool on the Internet). NewsGuard. 28/10/2019. URL: https://www.newsguardtech.com/fr/. Accessed on: 28/10/2019.

Signalement : campagne électorale (Notification: election campaign). CNIL. URL: https://www.cnil.fr/fr/webform/signalement-campagne-electorale. Accessed on: 19/02/2020.

Superspreaders. NewsGuard. 23/04/2020. URL: https://www.newsguardtech.com/superspreaders/. .

The Truth behind Filter Bubbles: Bursting Some Myths. Reuters Institute for the Study of Journalism. URL: https://reutersinstitute.politics.ox.ac.uk/risj-review/truth-behind-filter-bubbles-bursting-some-myths. Accessed on: 17/02/2020.

Dr Tedros Adhanom Ghebreyesus. *Munich Security Conference, speech by the Director General*. 15/02/2020. URL: https://www.who.int/fr/dg/speeches/detail/munich-security-conference. .

Manon Berriche. *Le fact-checking est-il vraiment efficace ? (Is fact-checking really effective?)*. SciencesPo. 22/01/2020. URL: https://www.sciencespo.fr/actualites/actualit%C3%A9s/le-fact-checking-est-il-vraiment-efficace/4539. Accessed on: 24/01/2020.

Laurent Bigot. *Le fact checking à l'épreuve des fake news (Fact-checking tested by fake news)*. La Revue des Médias. 17/10/2019. URL: http://larevuedesmedias.ina.fr/le-fact-checking-lepreuve-des-fake-news. Accessed on: 04/06/2020.

Conniefan.com. *Classifying fake news*. 22/03/2017. URL: https://www.conniefan.com/2017/03/classifying-fake-news/. Accessed on: 24/10/2019.

Rory Smith. *How to Investigate Health Misinformation (and Anything Else) Using Twitter's API*. First Draft. 06/03/2020. URL: https://firstdraftnews.org:443/latest/how-to-investigate-health-misinformation-and-anything-else-using-twitters-api/?utm_source=Daily+Lab+email+list&utm_campaign=86bc6e883c-dailylabemail3&utm_medium=email&utm_term=0_d68264fd5e-86bc6e883c-396240277. Accessed on: 10/03/2020.

9. Acknowledgements

The *Conseil supérieur de l'audiovisuel* would like to thank all the people it met during the preparation of this study and who, through their constructive feedback, helped to improve it. However, interactions do not imply validation or approval of the work by these people.

Other parties who took the time to receive the services of the *Conseil supérieur de l'audiovisuel*:

- *AFP Factuel*, represented by Grégoire Lemarchand, AFP's digital investigation editor.
- Romain Badouard, Assistant Professor in information and communication sciences at Paris-II, researcher at the CARISM laboratory.
- Les Décodeurs, *Le Monde*: Maxime Ferrer, Assma Maad, Jonathan Pariente and Adrien Sénécat.
- Twitter France, represented by Audrey Herblin-Stoop, Public Policy Director, France and Russia.