

IC05 : Critical analysis of numerical data

Semester: Fall 2020

Conspiracy theories about Covid-19



Head of the UV:
BELLON Anne

Students:
DUTHEIL Rémi
LAUBY Constance
MISSAOUI Benjamin
VIERA Baptiste

I) Introduction	3
I.1) Presentation of the subject	3
I.2) Issues and news	4
I.3) State of knowledge	4
I.4) Research framework	6
II) Methods and results	7
II.1) Twitter	7
II.1.a) Working method	7
II.1.b) Aspiration and results	10
II.1.c) Analysis of results: Twitter	14
II.1.d) To go further: the mapping of the conspiracy network	17
II.2) Youtube	21
II.2.a) Working method	21
II.2.b) Aspiration and results	23
II.2.c) Analysis of the results: Youtube	24
II.3) Twitter/Youtube cross analysis	28
III) Critical analysis	29
III.1) Data Bias	29
III.2) Improvements	30
IV) Conclusion	31
IV) References	33
APPENDIX	35
Appendix 1: Extract from a JSON file .	35
Appendix 2: R codes for Youtube text analysis	38
R code of the video "Covid-19 we are saved the vaccine-viagra is coming"	38
R code of the video "When conspiracy theorists doubt the existence of Covid-19"	42
Code R of the "Hold-up" video	46
Appendix 3: GitHub link	49

I) Introduction

I.1) Presentation of the subject

Since fin 2019, the world has been hit by a pandemic caused by the SARS-CoV-2 coronavirus, also called Covid-19. In an anxiety-inducing atmosphere linked to the fear of getting sick and months of confinement, a rise in conspiracy theories can be observed in the world, and particularly in France. These theories concern different topics around Covid-19:

- **Origin of the virus:** was it created by man, is it a bacteriological weapon escaped from a Chinese laboratory?
- **Use of chloroquine** to fight the virus: is it efficient, why is its use not widespread?
- **Vaccine race:** will it be mandatory? Is it a way to control the population by injecting, along with the vaccine, other products?
- **Efficacy of wearing a mask:** is it really efficaceous? Is it a means of deprivation of liberty put in place by the government?
- **Hospital saturation:** are they really saturated with Covid-19 patients?

With the help of new technologies and new means of communication, everyone's opinion can be shared in a click. Easily accessible and offering a certain visibility, social networks such as Twitter, Facebook or Youtube appear to be the ideal place for the diffusion of conspiracy theories around the Covid-19. Within the framework of the UV **IC05: Critical analysis of digital data**, our group decided to study the propagation of conspiracy theories around the coronavirus on two networks: Youtube and Twitter. Social networks are an ideal support for freedom of expression, it seems necessary to us to study how these networks allow the diffusion of theories. By selecting several recurrent themes in the conspiracy sphere, we want to identifier the main actors in the propagation of conspiracy theories.

We will start with a preliminary study of articles to more precisely delimit our subject of study, but also to definite the first actors and relays and approach the conspiracy sphere. Then, we will make several data aspirations on Twitter and Youtube to build a network analysis. Enfin, we will conclude with a critical analysis of this data.

I.2) Issues and news

In the news for the past year, a study of the impact of the Covid-19 pandemic is of major interest. The topic of Covid-19 generates a tremendous amount of reaction from individuals. The uncertainty of finding a vaccine quickly, the frustration related to the confinement, the economic impact of the sanitary measures taken to contain this pandemic participate in the development of theories around the virus. Moreover, these theories are particularly present on social networks. The traditional media only relay very little information on these theories, or they are refuted. We find, for example, Youtube videos made by the channels *France 24*, *RTL* or *TV5MONDE Info* that denounce these theories. The freedom of expression on these digital media seems to justify the privileged use of social networks by conspiracy theorists. However, this phenomenon does not date from the year 2020. The Internet has already been flooded with conspiracy theories, since the attack of September 11, 2001 for example. Each attack leads to a wave of conspiracy theories that spread on the Web. This was the case, for example, during the attacks of November 13, 2015 in Paris, with notably "several conspiracy Youtube videos had more than 100,000 views" according to an article in *Le Figaro* (2015) [1].

I.3) State of knowledge

First of all, we have carried out a first bibliographical research on the subject of the conspiracy sphere in general, as well as on the new theories related to the Covid-19. This step allows us to select the main themes on which we will focus our research. If several media are interested in the conspiracy theories around the Covid-19, it is mainly with the aim of denouncing them, or studying the impact and the distribution of these theories in the population. An article in *Le Monde* (2020) [2] attempts, for example, to link an individual's adherence to conspiracy theories about the Covid with his or her membership in a political party. Thus, 36% of the supporters of the *Rassemblement National* would validate the existence of a "Zionist conspiracy on a global scale". However, exceptions within the media are frequent, as shown by this article in *Valeurs Actuelles* (2020) [3] which, unlike *Le Monde*, is much less critical of these conspiracy theories.

Reading more general articles around the conspiracy sphere allows us to confirm our hypothesis that social networks are favorable places for the dissemination of conspiracy theories, as this article from *Columbia Journalism Review* (2017) [4] shows. This article discusses the American far-right media outlet *Breitbart* and its impact on the 2016 US presidential election. It highlights several elements: *Breitbart* created an entire ecosystem around the American far-right sphere using mostly social networks. The media has disseminated several misleading articles on this presidential election. *Breitbart*, a pro-Trump media, is also a supporter of conspiracy theories. The article *Study: Breitbart-led right-wing media ecosystem*

altered broader media agenda (2017) [4] reveals the overwhelming use of Twitter and Facebook for the dissemination of conspiracy content. Conversely, Hillary Clinton's supporters turned much more to the traditional media (TV and newspapers). These deliberately ambiguous articles can therefore be a source of misinformation.



Figure 1: Facebook post of the Health + Magazine page.

However, this spread may not be as optimized as we thought. For example, *The Conversation* (2019) [5] explains in this article that misleading information disseminated on social networks is not the most spread. Visual content on the *Health+Magazine* Facebook page, as shown in figure 1 opposite, generates 2 to 4 times more interactions than other content. Posts based on "consensual proverbs about social relationships" get the most interactions.

If the theories related to the coronavirus are numerous and concern multiple subjects, the conspiracy theories did not appear with the pandemic of 2020. Vaccines in general are for example the target of conspiracy theories for a long time. The essay by former Quebec physician Guylaine Lanctôt, *La Mafia Médicale* released in 1996, shows that the mefiance towards vaccines and antibiotics are ancient phenomena. Covid-19 seems to be just a new springboard for these theories, a new way of renewing them to the taste of the times. We can also quote the works of the French political scientist [6] Pierre-André Taguieff who describes conspiracy theories as "*an appeal to the people against the elites and a sacralization of the people, adorned with all the virtues, supposed to be permanently deceived by the cynical and corrupt elites*". It is in this belief of being deceived, manipulated and controlled by people with power, governments or big business leaders for example, that conspiracy theories are born. This feeling of being controlled and that the truth is hidden by powerful people is particularly emphasized by conspiracy theorists, who are often censored.

After this analysis, the profil of the conspiracists seems to become clearer: they are often people who are extremely méfiant towards the government and who have no confiance in the traditional media. Moreover, as the *Le Monde* (2020) article [1] cited above shows, there seems to be a link between

the adhesion to a political party of an individual and his adhesion to the conspiracy theories on the coronavirus. The spheres of the extreme left and especially the extreme right seem to be linked to the conspiracy theories, whether in France or in the United States.

I.4) Framework of research

We will define several hypotheses and lines of research in our project. The particularly anxiety-provoking climate caused by Covid-19 may explain the interest, number and impact of conspiracy theories about this virus. Despite the numerous topics present within the conspiracy theories around the Covid-19, we decided for practical reasons to focus on the following topics:

- **5G and Covid-19:** This theory explains that Covid-19 would allow the deployment of 5G with 5G microchips implanted with the vaccine. These chips would be used to track us [7],
- **Bill Gates' role in the origin of the virus:** according to this theory, the coronavirus would serve the interests of Bill Gates [8],
- **Hydroxychloroquine:** this theory says that the treatment set up by Dr. Didier Raoult would be efficacious to fight Covid-19 but would not be diffused because using an already existing drug would not be interesting for the pharmaceutical industry [9].

We have chosen not to deal with the anti-vaccine theories around Covid-19. Indeed, the conspiracy sphere around the vaccines in general is very vast. It does not concern only the coronavirus. We will also restrict ourselves to the conspiracy sphere in France. Finally, we will focus on data analysis from Youtube and Twitter.

Objectives of the study

- Theories are often accompanied by actors, public figures who are either, supporters of conspiracy theories or the target of theories. Within the framework of this project, we will try to identify some of them, as well as the themes treated in the conspiracy theories around the Covid-19.
- The diffusion of conspiracy theories and the reactions they provoke on social networks often take place following specific news events. We will try to observe this phenomenon.
- Enfin, we expect to observe a community, group effect in the diffusion of conspiracy theories. If conspiratorial individuals share theories around the Covid-19 on social networks, this diffusion should be mainly between conspiratorial individuals.

II) Methods and results

Our analysis is conducted using two digital media: Youtube and Twitter. For Twitter, we decided to focus on the three conspiracy theories mentioned earlier: **the link between Covid-19 and 5G, Bill Gates' place in the origin of the virus, and enfin the interest of hydroxychloroquine**. Along with the work done on Twitter, we will do a textual analysis of the comments on three YouTube videos. Each video is not about a particular theory. They deal with conspiracy theories in general around Covid-19 (use of masks, role of the confinement, origin of the virus, Bill Gates, etc...).

II.1) Twitter

II.1.a) *Method of work*

Working from Twitter, we were able to pull in several thousand tweets. The tweets were sucked in chronological order, from October 6 to December 6, 2020. The Python codes are available in the Appendix (Appendix 3).

To extract the data, we first tried to use the **Gazouilloire** tool developed by the Media Lab of Sciences Po. However, this was quickly a failure. We were confronted with a series of problems because the tool is developed in Python 2 and uses modules that are no longer up to date. This leads to many complications when running the program. Despite an attempt to convert all the Python 2 codes into Python 3, the program did not work.

In a second step, we turned to a different tool: a python module called **Tweepy**. This module allows us to retrieve tweets from keywords put in parameter (example: hydroxychloroquine, coronavirus). However, due to the youth of this module, we did not have access to all the features that the Twitter API could offer us. By using Tweepy, we could only retrieve tweets dating back to 7 days at most. This time constraint was far too strong and forced us to make weekly queries, which we found unproductive. Our goal was to retrieve tweets over a two-month period. We therefore decided to abandon the Tweepy module.

Despite these two failures, there was still a solution: using the Twitter API directly, using the command lines. This time, we did not encounter any particular problems. The Twitter API allowed us to collect tweets from up to 30 days ago without any deadline, using the **Search Tweets: 30-Days / Sandbox** tool and the **Search Tweets: Full Archive / Sandbox tool** respectively. However, we were limited in the number of queries and tweets we could run and harvest, as shown in the figures below :

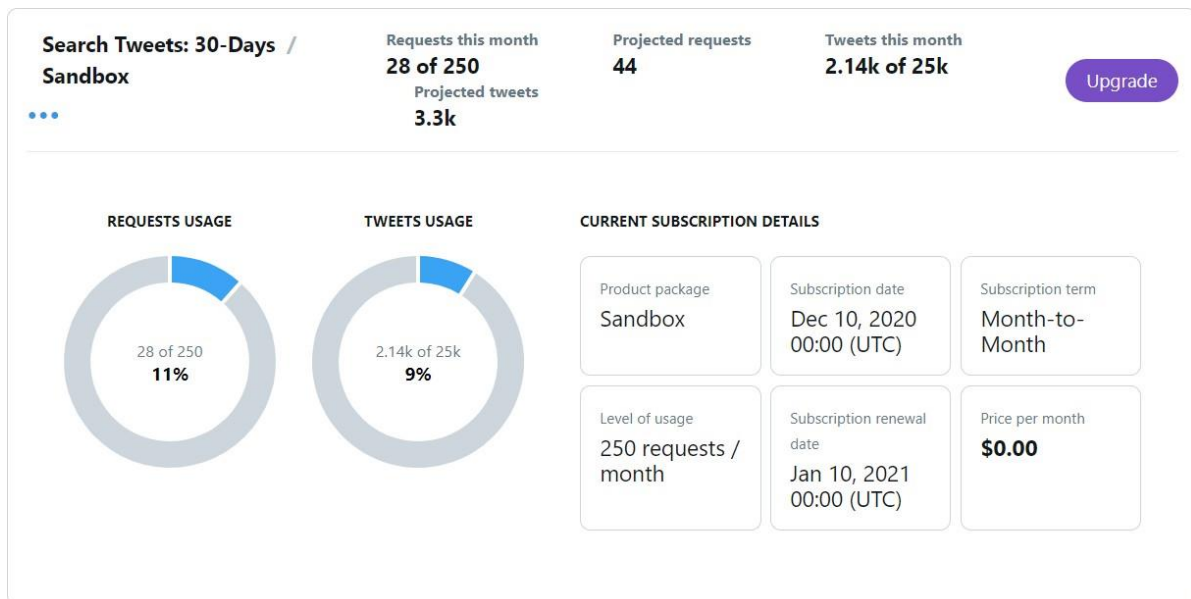


Figure 2: Limits on the number of queries allowed by the API with the Search Tweets: 30-Days / Sandbox tool.

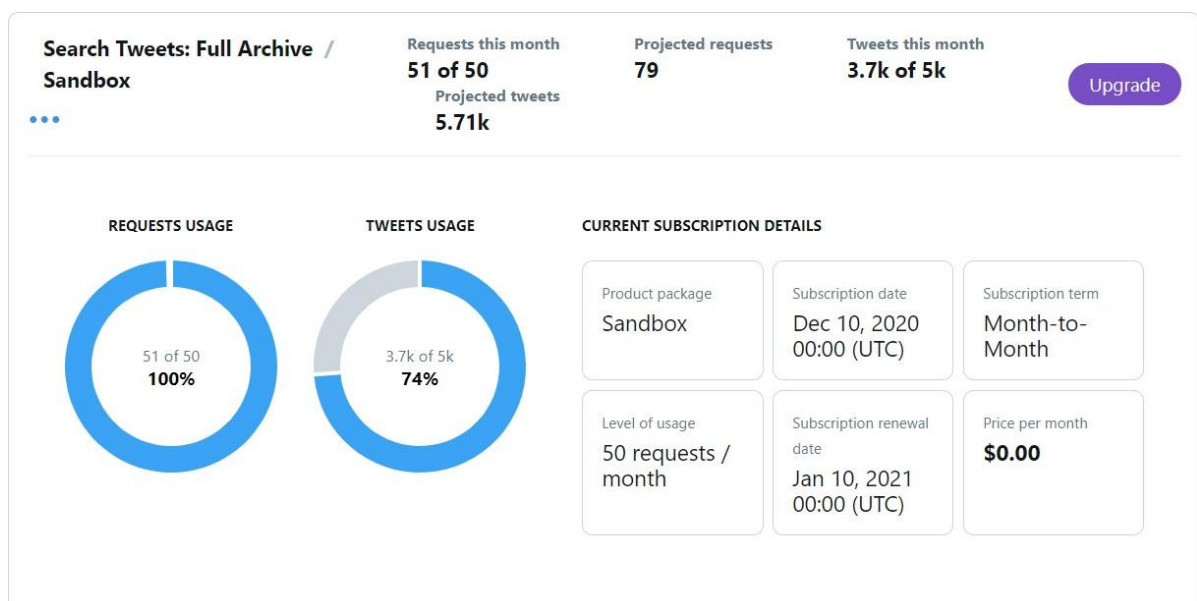


Figure 3: Limits on the number of queries allowed by the API with the Search Tweets: Full Archive / Sandbox tool.

The constraints were particularly strong for the second *Search Tweets: Full Archive / Sandbox* tool because we could only make 50 requests. Moreover, each query could retrieve a maximum of 100 tweets, which is relatively low considering the number of tweets generated in a day for example. To overcome this problem, we created 8 developer Twitter accounts to meet our needs.

The code to launch the twitter requests had the following form:


```

curl --request POST
  --url
  https://api.twitter.com/1.1/tweets/search/30day/dev30Days.js
  on
  --header 'authorization: Bearer
AAAAAAAAAAAAAAAAAAAAALB7JgEAAAAAAnmfvk1wZB6e1kRxyhFqXz6gPZrw%3
DOuEiTXM6WCeiJFe2JHGZFIJ60Vp157FJwkhjreUUWWT89E3z7H' \
  --header 'content-type: application/json' \
  --data '{
    "query": "hydroxychloroquine coronavirus lang:en",
    "maxResults": "100",
    "fromDate": "202011120000",
    "toDate": "202011121159"
  }' > 12Nov_00_12.json

```

Figure 4: Code in bash to launch twitter requests.

Code explanation:

--url: specifie which type of query we want to use, "30day" or "fullarchive".

--header : listed the account twitter developer by simply by entering the *BearerToken*.

--header: the resulting fichier will be of the form *json*.

--data:

- In the "**query**" part, we specific the keywords as follows, [name of the coronavirus theory](#), as well as the language of the tweets we want to retrieve, which is French in our case.
- In the "**maxResults**" part, we fix the maximum number of tweets we want to retrieve, knowing that we cannot go beyond 100 tweets.
- In the "**fromDate**" and "**toDate**" parts, we specific the range over which we want to retrieve the tweets following this precise format: years/months/days/hours/minutes.

II.1.b) Aspiration and results

We obtain *JSON* fichiers that contain a number of tweets (at most 100). An excerpt of a *JSON* file is available in the Appendix (Appendix 1).

Then, using various functions in python, we grouped all the *JSON* files for each of our three theories into a dataframe. We obtain *in fine* three databases corresponding to the three theories (hydroxychloroquine, 5G, Bill Gates). Each of these databases contains the tweets and their features, which are numerous as highlighted in the example in Appendix 1.

From these databases, we will make two distinct analyses. The first treatment consists in the realization of several histograms which represent the evolution of the number of tweets as a function of time for each of the theories. In addition, a pie chart represents the percentage distribution of tweets for each of the three theories treated here.

The graphs obtained are as follows:

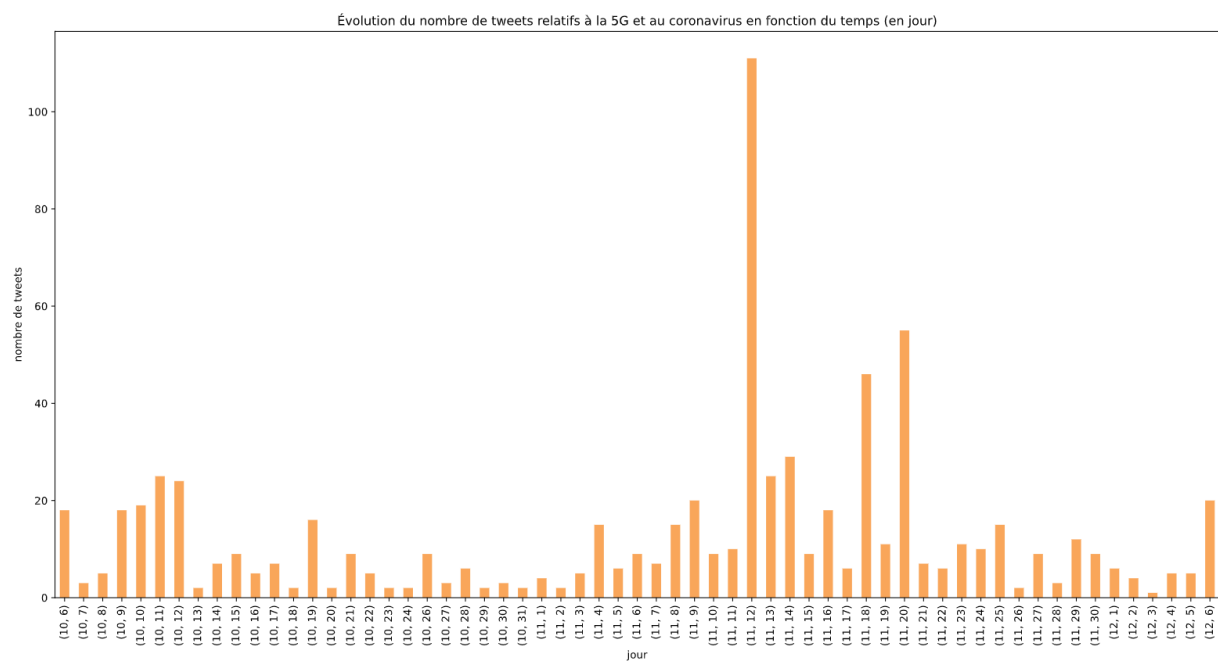


Figure 6: Evolution of the number of tweets related to 5G and coronavirus.

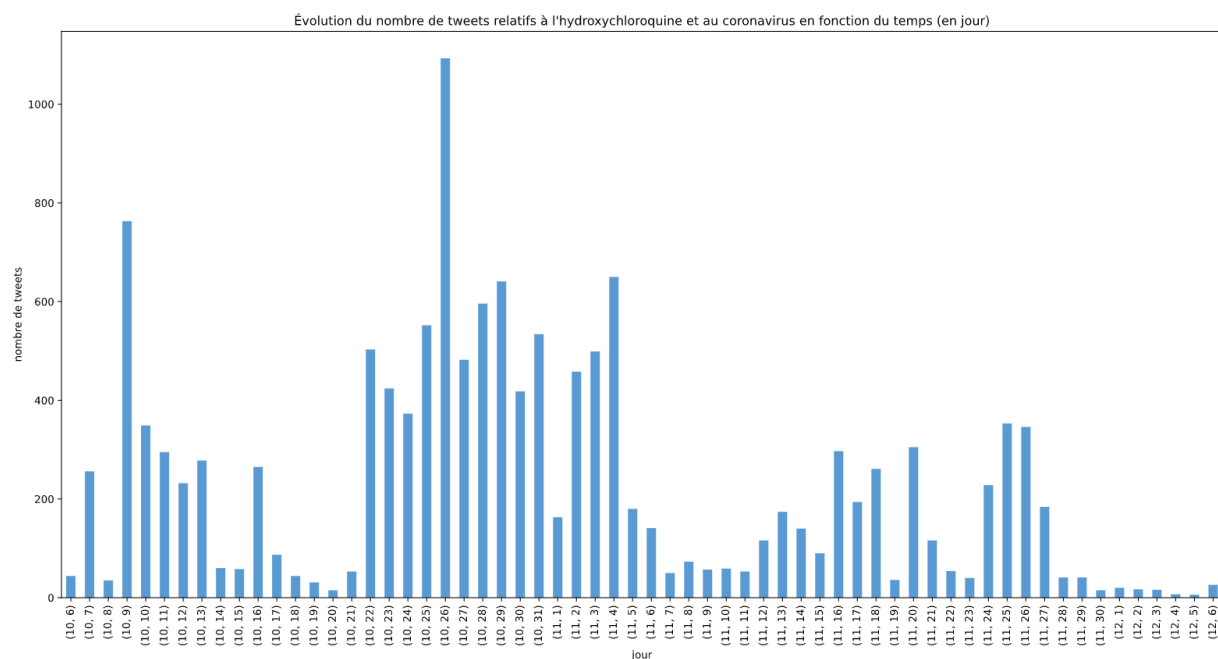


Figure 7: Evolution of the number of tweets related to hydroxychloroquine and to the coronavirus.

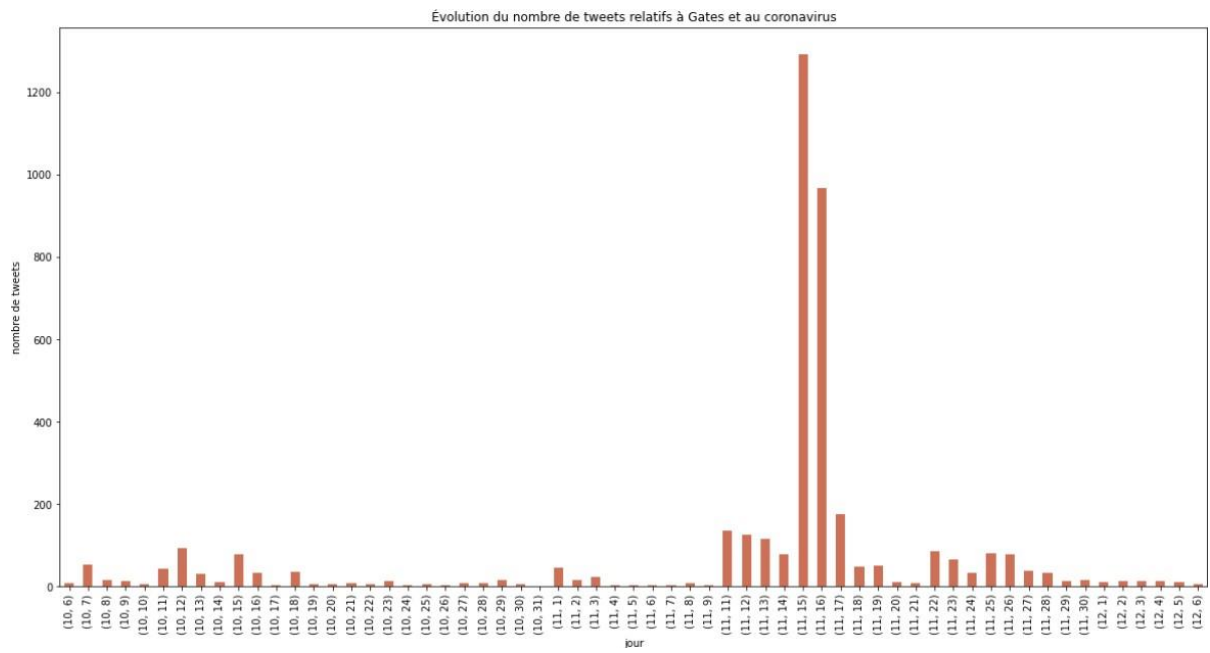


Figure 8: Evolution of the number of tweets related to Gates and the coronavirus.

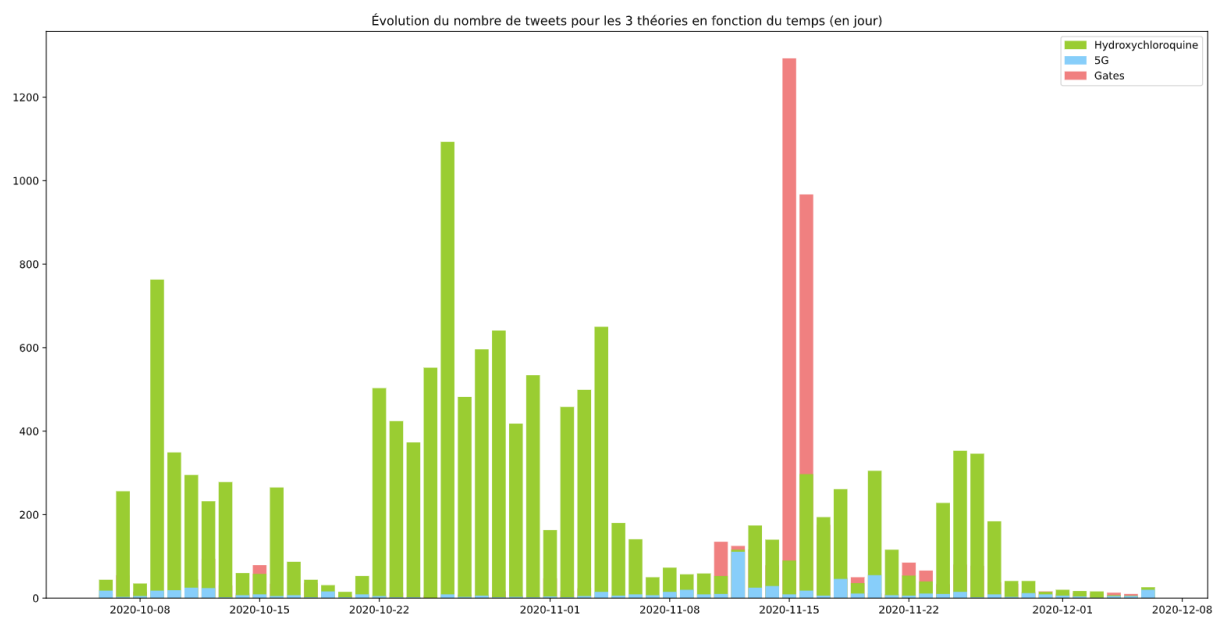


Figure 9: Evolution of the number of tweets related to 3 theories and the coronavirus.

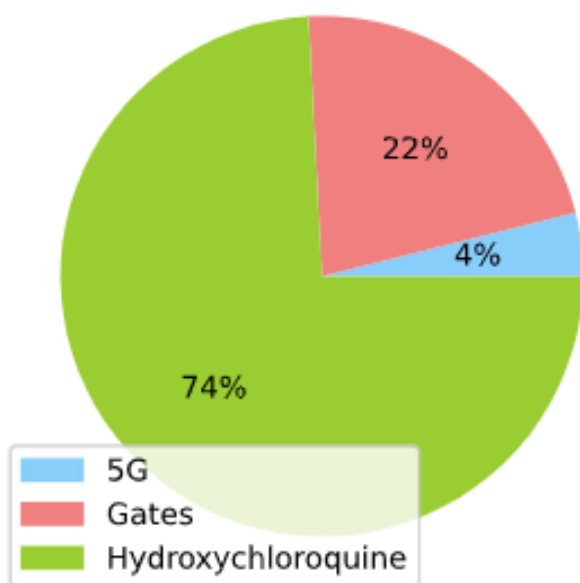


Figure 10: Proportion (%) of tweets by theory.

II.1.c) **Analysis of results: Twitter**

The first thing we can notice is that the proportion of tweets mentioning both hydroxychloroquine and coronavirus is much higher than the other two. Indeed, figure 10 shows that of the more than 20,000 tweets we sucked up, 14,000 (74%) relate to the theory

conspiracy theories about hydroxychloroquine, 5,000 (22%) about Bill Gates and 1,000 (4%) about 5G. It remains to be determined, however, whether the hydroxychloroquine theories are indeed in the majority compared to the other two, or whether our database contains more false positives than the others (non-conspiracy tweets addressing the topics of coronavirus and hydroxychloroquine, at purely scientific fins for example). We will elaborate on this point later.

The bar graphs in figures 6, 7, 8, and 9 allow us to identifier "spikes" in the number of tweets on a topic. In the case of 5G, a peak is observed on November 12, 2020 in figure 6, with 111 tweets published. This date corresponds to the day after the release of the conspiratorial independent documentary ***Hold-Up: Return to Chaos***. This documentary assures to show that the coronavirus was developed with the aim of creating "*a technological world dictatorship based on contactless, the Internet of Things and therefore 5G*", as shown in this article from *France Culture* (2020) [10]. Therefore, this documentary seems to us a good target to realize a textual analysis of the reactions and comments it provokes. We will explore this documentary in more detail in the Youtube part of this study.

Next, the chart on the conspiracy theory around Bill Gates (figure 8) also shows a peak of 1293 tweets on November 15, 2020. This spike appears to be due to retweets of a message from "@TontonFlingr":

"Billionaire Elon Musk refuses to get a coronavirus vaccine and calls Bill Gates a 'moron'," @TontonFlingr.

This tweet is actually the title of an article from the *PLANETES360* (2020) website [11]. It is accompanied by the link to the latter. It is difficult to say why this tweet generated so many reactions. However, the twitter account *@TontonFlingr* happens to be the most retweeted account in our entire database about the theory around Bill Gates. In fact, 31% of the retweets in our database came from the *@TontonFlingr* account, as shown in figure 11 below.

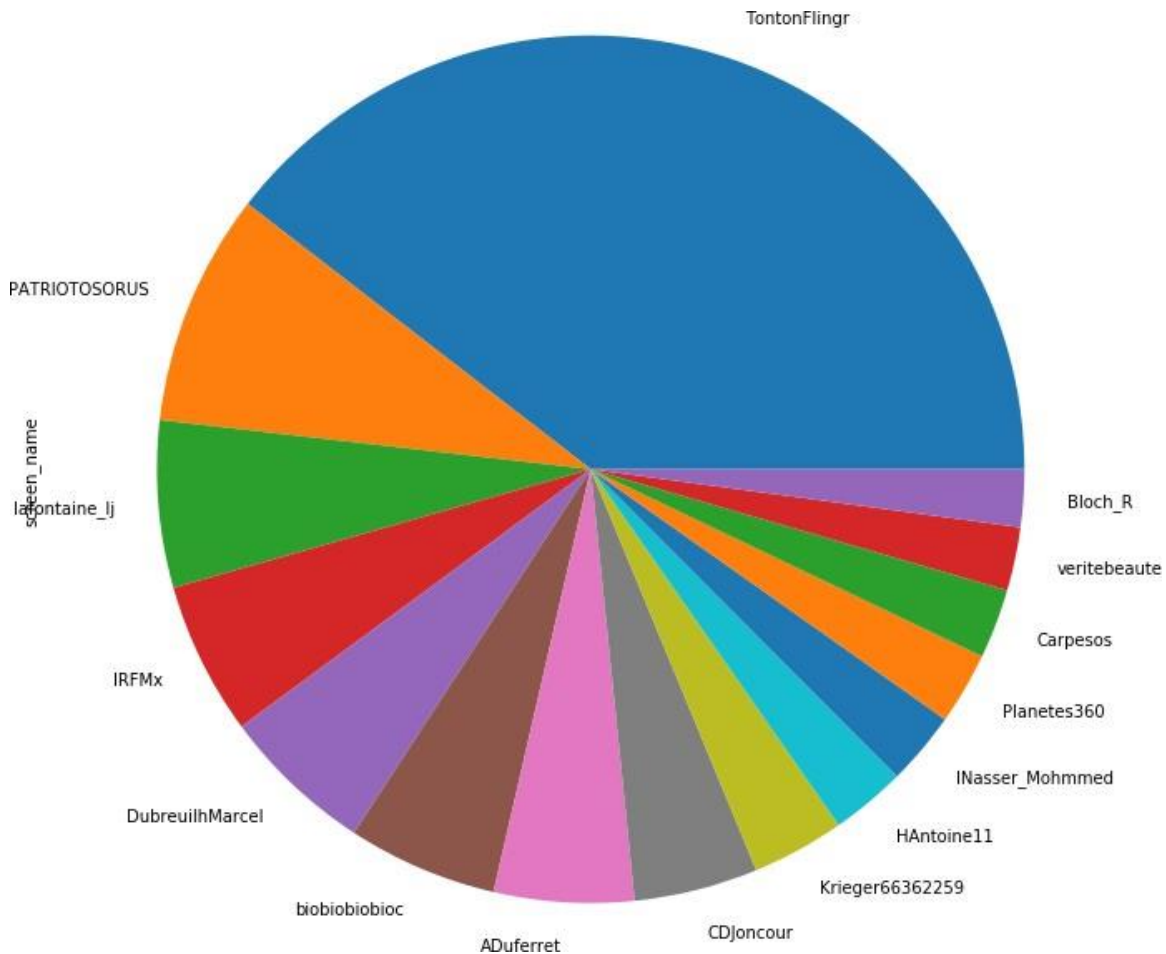


Figure 11: Distribution of the 15 most retweeted users among the tweets around the conspiracy theory about Bill GATES.

The account *@TontonFlingr* was subsequently deleted. It was held by an individual named Pascal Laborie, who was followed by over 4,000 people before it was deleted. The fact that an individual holds such an influent account shows that conspiracy theories can quickly spread using social networks such as Twitter due to the high visibility to a large audience. This audience, however, seems, in the case of conspiracy theories, to be composed only of conspiracists. These conspiracy theories are therefore spread between individuals of the same group, within a community and with a possible leader. The opinions of the latter are directly accepted and massively propagated by those who follow him, without questioning. Moreover, no official media account is present among the most retweeted accounts. Only user accounts are present, which confirms well this idea of a **conspiracy community**.

Finally, the banding diagram in figure 7 shows that the conspiracy theory around hydroxychloroquine is overall more active and discussed compared to our other two theories. The number of daily tweets about this theory is larger compared to the other two theories, as shown in figure 9. A peak in the number of tweets is present on 26

October 2020. This day corresponds to the reaction of Didier Raoult, professor of microbiology, on Twitter after the ANSM (Agence nationale de sécurité du médicament et des produits de santé) decided to stop the extensive use of hydroxychloroquine in testing (ANSM, 2020) [12]. Didier Raoult himself seems to be very little retweeted by the conspiracists. In our data aspiration, only 7 users retweeted him mentioning the words "coronavirus" and "hydroxychloroquine", between October 6 and December 6, 2020. The hydroxychloroquine theory also has a rather influent user, whose handle is *@biobiobiobio*. He currently has over 15,000 followers.

Enfin, the base of tweets on the latter theory includes some official media twitters accounts. Nevertheless, these accounts do not cause as much reaction compared to the other twitters accounts. Indeed, none of the 13 most retweeted accounts for hydroxychloroquine is not the account of an official media outlet. This shows the ease with which conspiracy theories can spread over information transmitted by official media.

To conclude, these initial analyses reveal that conspiracy theories on Twitter are propagated mostly among conspiracy theorists, in fairly closed groups such as communities, and not through comments under the posts of certain major media outlets, for example. These users profitent certain news events related to the theory they believe in, but largely distort reality to fit their vision. This false information then tends to spread much more easily, especially in this current uncertain context.

II.1.d) To go further: the mapping of the network complotiste

The databases developed so far have allowed us to identifier some important actors in the propagation of conspiracy theories such as *@TontonFlingr* for the theory about Bill Gates, and *@biobiobiobio* for the one about hydroxychloroquine. Each tweet contains information about its author, but also about the author he retweets if necessary. Thus, by extracting the tweets that are indeed retweets, and for which Twitter provides us with a location clue, we can afficher the conspiracy network as a graph on a map, shown in figure 12 below.

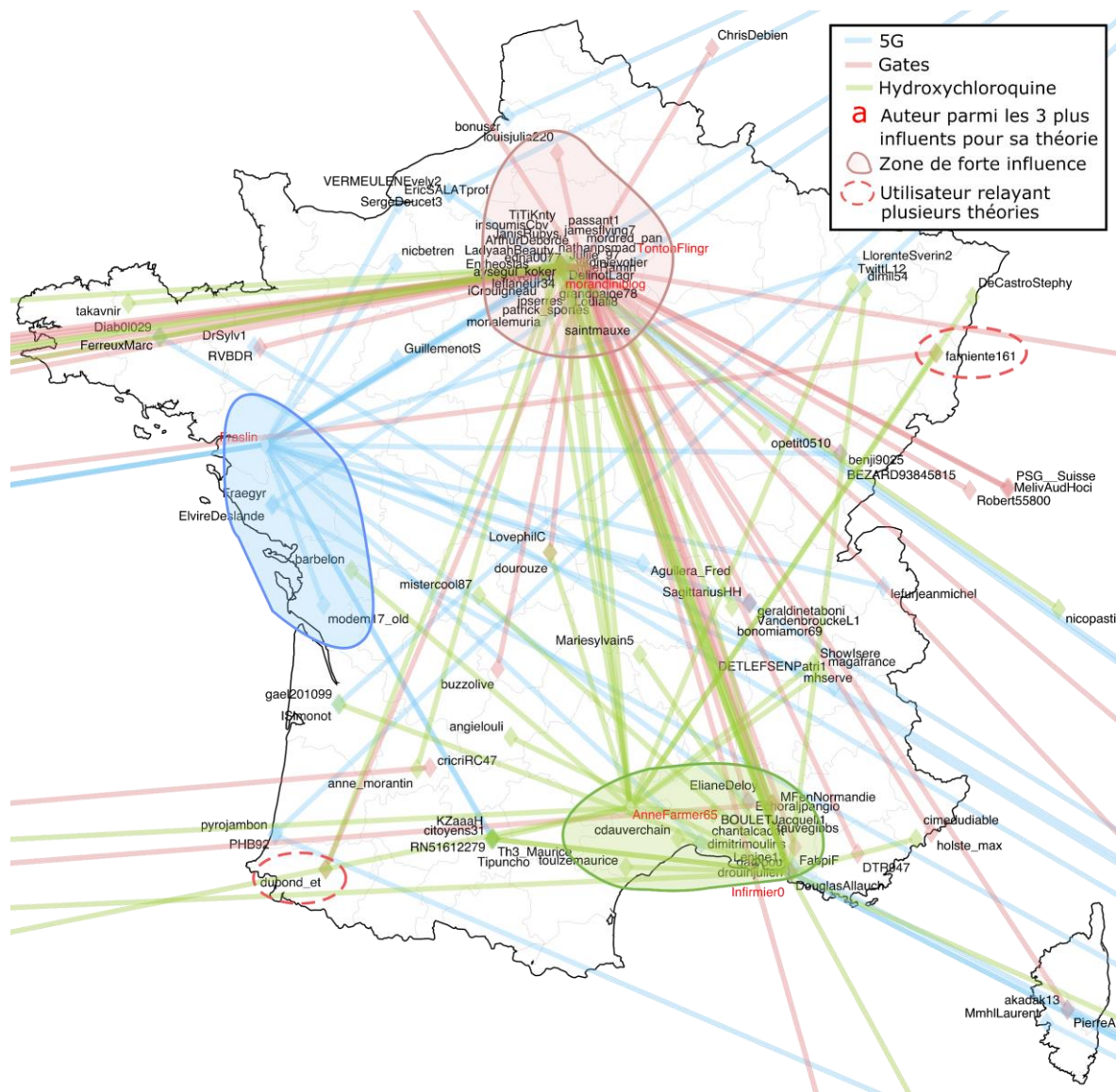


Figure 12: Network of different conspiracy theories in France.

Afin order to improve readability while conveying as much information as possible, we first decided to keep only the arcs related to one of the three most influential authors for each theory. We also kept only 50 arcs per theory (respecting of course the proportions, i.e. if authors A, B and C are the most important for 5G, and they have in been retweeted 500, 2000 and 2500 times, we will therefore afficher respectively 5, 20 and 25 arcs for each of them).

The 3 most influential authors for each theory (those whose links we have affiché) can be summarized by the following table:

Theory	Author	Number of arcs	Location
5G	Fraslin (16,600 subscribers)	21	Nantes, France
	RaoultDidierOff (1 736 subscribers)	18	Tel Aviv, Israel
	BobLeCentriste (16 300 subscribers)	9	Koszalin, Poland
Gates	TontonFlingr (4 137 subscribers)	35	Paris, France
	PATRIOTOSORUS (8,522 subscribers)	7	Bryce, Utah, USA
	lafontaine_lj (16 subscribers)	5	Quebec, Canada
Hydroxychloroquine	Infirmier0 (16,900 subscribers)	19	Marseille, France
	AnneFarmer65 (15 200 subscribers)	15	Le Caylar, France
	morandiniblog (662 400 subscribers)	12	Paris, France

Table 1: Geographic location, numbers of arcs, and authors of retweets in figure 12 by conspiracy theory around Covid-19.

Interpretation

This map raises several interesting points. Each colored line represents the fact that one user retweeted another (indicating the terms "coronavirus" and "5G"/"Gates"/"Hydroxychloroquine" in their retweet). If the line is blue, it connects two users supposedly spreading theories about 5G; the red line connects two users relaying theories about Bill and Melinda Gates; and the green line connects users spreading theories about hydroxychloroquine. At the end of each arc is the Twitter handle of the relevant user, and this text is colored red if it is one of the most influential (one of the most retweeted) users for its theory. In addition, we added colored areas around the locations of the most influential users for each theory.

Analyzing this map, the first thing we can notice are the trends. Two of the three most retweeted users for hydroxychloroquine (@AnneFarmer65 and

@Infirmier0) are located in the southeast, not far from Marseille, while *@TontonFlingr* and other prominent accounts

for theories about Bill Gates, are rather located in Paris. These areas make particular sense: Didier Raoult being based at the IHU in Marseille, it seems normal to notice that individuals living not far from this area feel more concerned by the theory around hydroxychloroquine. We also notice that many users are located in the United States for the theories around Bill Gates (several red arcs go out from France to the West, towards the USA).

Enfin, the area of influence in the west also makes sense for the theory around 5G. Nantes is one of the only cities that should be covered by the 5G network in this mostly rural area. 5G is quite a controversial topic in this city, as this article from *metropole.nantes.fr* shows [13] written on December 8, 2020.

Enfin, we notice that despite the few arcs affected on this map (only 50 for each theory, or about 0.8% of the arcs in the original databases), we can still observe two users, circled by a dotted red circle, who retweeted and thus relayed several different theories: *@farniente161* and *@dupond_et*. If this phenomenon can already be observed with this extremely small number of nodes and arcs, there is no doubt that it is even more visible on a larger scale. If a twitter account shares a conspiracy theory, it must surely share other conspiracy theories on different topics. This observation contributes again to our hypothesis that there is a "conspiracy community", a set of users in which conspiracy theories are particularly popular and relayed.

To conclude this first part, the analysis reveals that the main relays of conspiracy theories around Covid-19 are not traditional online media, but a certain number of particularly followed users, who instrumentalize certain facts and spread false information. Moreover, this sharing of false information seems to take place mostly and only within the conspiracy community. Conspiracists share little information outside of the conspiracist sphere. We have also learned that the capacity of a theory to spread in a geographical area is all the greater if the individuals in this area feel concerned, worried or ill-informed about the subject. This can explain the phenomenon of **geographical area** for a specific subject of a conspiracy theory. This phenomenon is, for example, present in figure 12, with a green zone representing a trend of conspiracy theories about hydroxychloroquine near Marseille, the city where Didier Raoult, a major actor in the debate around the role of this drug in the fight against coronavirus, is located.

However, we still have to explain the mechanisms that allow these theories to spread so much. For this, we perform a textual analysis on Youtube.

II.2) Youtube

On Youtube, the channels and videos talking about the conspiracy theories of the coronavirus are mainly those of traditional media such as *France Culture*, *Le Parisien* or *Le Monde*. These channels and videos denounce these theories and do not adhere to them. In view of their position against conspiracy theories, we can think that the extraction of data from these media could lead to a selection bias in our data. We can therefore assume that the comments under such videos will be primarily **against** conspiracy theories. However, proponents of conspiracy theories are often active in the comments of such videos, even if they are only a minority. To address this possible bias, we will also draw data from videos or youtube channels adhering to the theories. The *WhataBeautifulWorld Youtube Channel* and the *Riposte laïque channel* contain numerous videos promoting theories about Covid-19, notably about wearing a mask and the origin of the virus. It is important to note that these channels are difficultly accessible from the youtube search bar. They do not have a good referencing, which explains why they are not highlighted by the platform. This difficulty of access to this type of content on Youtube may be due to a moderation phenomenon on the part of the platform. A study on this topic would be particularly interesting, especially in the context of the fight of social networks against Fake-news and conspiracy theories. However, our project will not deal with this aspect.

II.2.a) **Method of work**

We analyzed the comments under three videos. We have two conspiracy videos and one non-conspiracy video:

- **"Hold-up"** [14]: reportage put for free on the *WhataBeautifulWorld Youtube Channel*¹. This French documentary released on November 11, 2020 was produced by Pierre Barnérias, Nicolas Réoutsky and Christophe Cossé. It has had a great impact in conspiracy circles because it questions the government's decisions (wearing masks, usefulness of confinement, hydroxychloroquine). It also raises the theory of a global conspiracy behind Covid-19 [10]. The video totals 524 comments and the *WhataBeautifulWorld Youtube Channel* has 2,860 subscribers.
- **"Covid-19 we are saved the vaccine-viagra is coming"** [15]: video totaling 10 198 views and 364 comments. It is posted on the youtube channel *Riposte Laïque*, which has about 42 400 subscribers. *Riposte Laïque* is above all a far-right website created in 2007, whose creator, Pierre Cassen, arrived on Youtube in 2007. On his channel, he shares his conspiracy and Islamophobic ideas.

¹ The video was removed before the fin of our project.

- **"When conspiracy theorists doubt the existence of Covid-19"** [16]: video with 31,848 views and 416 comments, from the *FRANCE 24* media youtube channel, with 3.08 million subscribers. This short report denounces the conspiracy theories around Covid-19, including theories affirming that the virus is harmless, or even that it does not exist.

The first two videos come from openly conspiracy-minded youtube channels. We therefore assume that the comments are written by conspiracy-minded individuals. On the contrary, the last video is posted by a news channel. We therefore hypothesize that the comments will be written by individuals who are not conspirators. These three videos are therefore separated into two distinct groups: conspiracy videos and non-conspiracy videos. These hypotheses will be disproved or validated with the help of the textual analysis of these comments.

We use *RStudio* to perform a textual analysis from the comments of the three selected Youtube videos. We experimented with the analysis method in the framework of the TD n°8 realized on Thursday, November 19th. *Rstudio* allows us to process the raw data retrieved by *Youtube Data Tools*. A processing and filtering step allows us to extract the most recurrent words and expressions in the comments, but also to remove from this list parasitic words such as "the", "the" or "in". By comparing the results obtained with the three videos, the textual analysis allows us to identifier common themes. In addition, we can also identifier whether a video mentions a particular topic or personalities. This is how links can appear between these theories and a political party or a public figure, for example.

The comments are extracted using the *Youtube Data Tools* tool, presented in TD in IC05. We used the *Video Info and Comments Module* option to obtain CSV files grouping the comments. Then, using the *Rstudio* application, we performed a textual analysis afin highlighting the most used words and expressions in the comment space. The R codes used are presented in the Appendix (Appendix 2).



Figure 15: Textual analysis of the video "When conspiracists doubt the existence of Covid-19".

II.2.c) **Analysis of the results: Youtube**

As a first step, it is important to differentiate the individuals who commented under the three videos. We assume that the comments on the videos "hold-up" (figure 13) and "Covid-19: we are saved the viagra vaccine is coming" (figure 14) were written by supporters of conspiracy theories. This hypothesis is based on the highly conspiratorial nature of the youtube channels with these videos. Conversely, we consider the comments under the video "When conspiracists doubt the existence of Covid-19" (figure 15), from the *FRANCE 24* media, were written by individuals who do not adhere to these conspiracy theories. We therefore analyzed two conspiracy videos and one video against these conspiracy theories.

Many commonalities appear in these three textual analyses. First, the lexical field of **fear** is present in all three figures. Words such as "fear," "death," "evil," or "die" clearly show the fear of the population in the face of the threat posed by Covid-19. The anxiety-inducing atmosphere of today's society is clearly evident in the comments of the videos.

The **vaccine** theme is also present in all three figures, with the words "vaccine" and "vaccinate." Here, vaccine is used to talk about a specific theme of conspiracy theories around the coronavirus. However, it can

also refer to the conspiracy around vaccines in general. This shows that the conspiracy sphere around Covid and the anti-vaccine conspiracies are linked. Also, **"government"** is a common word in all three videos. The latter is particularly criticized for its management of the sanitary crisis, or because of the sanitary measures taken, impacting enormously the economy of the country but also the individual freedom.

Next, themes common to only two out of three videos are revealed using our textual analysis. The documentary "hold-up" is present in figures 13 and 14. This highly controversial and criticized 2020 documentary has had a resounding impact in the conspiracy sphere. It is cited numerous times.

The acknowledgements with the words "thank you" and "bravo" in figures 13 and 14 seem to validate our hypothesis that it is indeed conspiracists who have commented on these same videos. They thank the publication of this documentary. A desire to "share" and to know the "truth" shows that the conspiracy individuals are calling for the dissemination of the "hold-up" video link afin to propagate the so-called truth that this documentary would bring.

Individuals' mefiance of media reporting is marked by the terms "lies," "false," and "truth." The possibility of a "conspiracy" is mentioned in figures 13 and 15.

In addition, several actors appear to be identified using this textual analysis. Three public figures are mentioned in figures 13 and 14: "Macron", "Trump" and "Bill Gates". The French president is criticized for his choices and his management of the crisis. This is accentuated by the general criticism of the French government. Donald Trump seems to be a major actor in the diffusion of conspiracy theories. Joe Biden is also mentioned in figure 14. It is interesting to note that the video "Covid-19: we're saved the viagra vaccine is coming" was posted on the platform before the fin of the US presidential elections. We know that conspiracists are often sympathetic to the worldview promoted by Donald Trump. The "QAnon" movement (*LeMonde*, 2020) [17] testifies for example to this conspiracy and pro-Trump association. We can therefore assume that the comments mentioning Trump are cheering for him, while his opponent Joe Biden is criticized or even mocked. Bill Gates is also mentioned because of the theory surrounding his involvement/role in the Covid-19 pandemic.

Recurring themes related to conspiracy theories around covid-19 thus seem to emerge from this analysis. The government is particularly singled out in its decision-making and management of the 2020 crisis. The presence of "masks" in figures 13 and 15 shows that their use and usefulness is highly controversial and is another major theme within the conspiracies around the coronavirus.

If the textual analysis of the Youtube comments shows us common points between these videos, some differences appear however. The *France 24* video is the only one for which we hypothesize that the comments were written by individuals who are not supporters of the conspiracy theories around Covid-19. The virus-related theory appears to be figure 15 related to many other conspiracy theories. The terms "moon," "ISS," "banks," or "[earth] flat" do not seem to be part of the lexical field around Covid-19. Many conspiracy theories are evoked here. They are very popular, such as the flat earth theory, or the one affirming that humanity never set foot on the moon. While these various theories are present in figure 15, commentators openly mock them. The figure 16 shows the top commentators, ridiculing the conspiracies.



Figure 16: Excerpt from the comments of the video "When conspiracy theorists doubt the existence of Covid-19".

The conspiracies evoked are associated with "bullshit" as shown in figure 15. Here, the individuals are mostly non-adherents. This criticism of conspiracy theories seems to correspond to the profile of a subscriber to the *France 24* channel, which is basically a television news channel. This accentuates the fact that it is not the same individuals who consume content from the *Riposte Laïque* channel as from a news channel.

In contrast to the *France 24* video, the lexical field of the documentary "Hold-up" and of the video of the *Riposte Laïque* channel suggests that the commentators are in favor of conspiracy theories about covid-19. For the documentary "Hold-up", the textual analysis is marked by acknowledgements, with words such as "thank you" and "bravo". The commentators appear to be congratulating the documentary makers. This confirms our hypothesis that these comments are written by conspiracy supporters. These thank the documentary for telling the "truth" about Covid-19.

The difference in comments between these videos can be explained by the political opinion of the commentators, but also their social status. Under the *France 24* video, people seem to be indignant about these theories. On the contrary, under the two other videos, individuals adhere to and wish to share the "truths" promoted by the conspiracy theories. The audience of these videos is therefore very different.

Enfin, the third video uses a very provocative register, like many videos from the *Riposte Laïque* youtube channel. The title of the video is meant to be shocking: the vaccine is associated with "viagra". The comments are favorable to conspiracy theories and criticize the vaccine. Donald Trump, a public figure very active on social networks and on the subject of the coronavirus, and Bill Gates, at the heart of a conspiracy theory, are mentioned. The individuals seem to support Pierre Cassen, the owner of the Youtube channel. The public of *Riposte Laïque* seems to be conspiratorial and adheres to the ideas of Pierre Cassen.

II.3) Cross analysis Twitter/Youtube

The analyses of Twitter and Youtube can be put in parallel. The interpretation of the results shows us first that social networks seem to be indeed a favorable ground for the diffusion of conspiracy theories. The freedom of expression and the possibility of anonymity explain this phenomenon. However, these theories seem to spread mainly within the conspiracy community. The analysis of conspiracy theories on Twitter shows that conspiracy tweets mainly reach a public of conspiracists initiated to these theories. Conspiracy theories therefore reach very few other users of social networks.

Then, conspiracy theorists profit of current events or more or less controversial government measures to propagate their theories and interact on social networks. The peaks on tweets correspond to specific dates when events have provoked a lot of reactions from Internet users, as with the release of the documentary *Hold-up* for example. The publication of Youtube conspiracy videos also coincides with current events.

Moreover, our analyses allow us to identify several actors of conspiracy theories. In particular, they highlight the involvement of the extreme right-wing sphere. On Youtube, the most popular conspiracy channels often claim to be far-right. This is the case of the channel *Riposte Laïque*, which we mentioned earlier. The videos of the channel talk about the favorite subjects of the extreme right: immigration, Islam [18], nationalism or anti-parliamentarianism [19]. A majority of the people who approve these conspiracy videos also seem to adhere to the extreme right-wing ideas of *Riposte Laïque*.

Generally speaking, we can affirm that political figures are actors in these theories. This is the case of Donald Trump, for example. His name can be found in the textual analyses of each Youtube video. The latter has repeatedly defended Doctor Raoult's treatment with hydroxychloroquine [20], which is in line with the conspiracy theory on this drug. Similarly, it is no coincidence that someone as influential as Bill Gates is at the heart of one of the theories. Indeed, for conspiracy people, the richest and influentes personalities in this world are all connected in some way to conspiracy theories or secret organizations [21].

Infin, we understand that social networks are both actors and relays of conspiracy theories. They are relays because today, it is on social networks that conspiracy theories are mainly and most effecttively propagated. But the networks are also actors in this dissemination, because their algorithms have a role to play in whether or not conspiracy theories are put forward. We have seen, for example, that on Youtube, it was much easier to find videos from traditional media criticizing conspiracy theories using the Youtube search bar than conspiracy videos.

III) Analysis critical

III.1) Bias of data

In our project, we sucked in a rather small amount of data in the Youtube part. We have to hypothesize that our data has a **selection bias**. Introduced in UV IC05 course #6, titled "Data Bias", selection bias can be definiated as a systematic error related to the selection of the subjects studied. This observation may be particularly present in the Youtube part of our project. Indeed, as we have presented, the majority of videos on the platform dealing with conspiracy theories about covid-19 come from media outlets that do not adhere to these theories. On the contrary, the videos criticize and dismantle these theories by showing their limitations or by presenting evidence to refute them. We can think that the consumers of these Youtube channels agree with the message conveyed by these videos, namely that these theories are false or ridiculous. The comments may therefore not be representative of the population in favor of these theories, but on the contrary be mostly made up of people criticizing conspiracy theories. It is also difficult to find videos on Youtube of conspiracists even by typing in keywords or doing an advanced search. We can assume that Youtube's algorithms are designed to not properly reference videos of conspiracists and to make them fall "into oblivion."

Moreover, the selection bias can also be observed with the data extracted from Twitter. Indeed, the aspiration works by keywords, and returns all the tweets corresponding to these keywords. Our database on hydroxychloroquine, for example, represents the tweets whose text includes "coronavirus" and "hydroxychloroquine". Our goal is to target conspiracy tweets related to hydroxychloroquine, but of course, conspiracists are not the only ones to use these two words together. By browsing through our database, we can see that the authors of some tweets are indeed not conspiracy theorists at all, but "serious" media outlets, which are simply reporting the news:

Figure 17: Extract from the authors of the hydroxychloroquine database.

```
2015nemtsov
2022_pour
20320Vero
20Minutes
20minutesMars
22Patounet
23Meurant
24h_algerie
24sante_fr
29aatea
2K21Dalyells
2LaRocheJaquell
```

So, while these traditional media are (seemingly) in the vast minority, their presence could qualify our results. The best way to rule them all out would be to go through all the bases by hand, but that would mean going through 20,000 tweets one by one.

III.2) Improvements

The amount of videos and tweets aspirated could be increased in a new study afin order to have results that are more representative of the overall population. While data mining on Youtube is only dependent on the time and technical skills of the team in charge of the project, there is a particular difficulty in using Twitter as a medium. The site's API requires permission from Twitter for the research team to use the platform's digital data. Without this permission, data mining is nearly impossible.

Second, although our team had access to the Twitter API, we faced several technical limitations. Indeed, the number of tweets that can be sucked in is prefini, and this limit cannot be exceeded.

A second important point to emphasize is the temporality of the tweets collected. The site's API allows us to acquire tweets from less than a month ago without restrictions on their number. While this can allow for temporal analysis, as the news unfolds, we cannot analyze data from the beginning of the pandemic, for example. Twitter allows for older aspirations, but the queries are extremely limited. We needed

8 different developer accounts to cover a period of only two months. This global pandemic has now been going on for over a year, and conspiracy theories began to emerge early on. Verifiable and fiable information about Covid-19 was scarce, making it easier to spread rumors and false information. Even so, our knowledge of this virus and our view of the crisis has largely evolved over the filths. It is therefore possible that sucking up older tweets (from the first confinement, for example) will highlight completely different results, with new actors and new communities.

IV) Conclusion

Social networks appear at first as the Eldorado for the massive sharing of information, conspiracy theories or not. The freedom of expression and the promise of being able to reach a maximum number of people thanks to the visibility of publications on the networks seem to be the right ingredients for the sharing of conspiracy theories around Covid-19. The way Twitter works is particularly favorable to this wide diffusion: a news event can generate reactions from Internet users that will be shared instantly.

Our study was able to identifier public figures such as Emmanuel Macron, Bill Gates, and Donald Trump at the center of these conspiracy theories around Covid-19. Major players in the spread are also identified on Twitter. A small number of accounts are centralizing reactions and retweets on the social network about the coronavirus.

In addition, conspiracy theories about covid-19 are raised alongside many other news stories, but also other conspiracy theories. The government and the management of the health crisis are often pointed at, and covid-19 brings to the forefront the general conspiracy around vaccines for example.

Enfin, a community effect is observed on Twitter. Conspiracy theory tweets are disseminated within a specific circle of individuals. A geographical phenomenon is also observed. Reactions about a particular topic will tend to be concentrated in a geographic area where public figures relay information about conspiracy theories. This is observed with the reactions on the subject of hydroxychloroquine around Marseille, the city where Didier Raoult, a major actor on this particular theme, lives.

However, a fight against Fake news and, at the same time, against conspiracy theories, is visible on these platforms [9]. On Youtube, we know that a non-monetized video will have a lower referencing than a monetized content. This variation in referencing can help videos to be promoted, but also reverse this process. This seems for example

be the case for the videos of the channel *Riposte Laïque*. It has 42,500 subscribers and a total of nearly 30 million views. It could therefore theoretically have access to monetization, but this is not the case. The monetization of a channel is obtained after vérification to the compliance of the platform's rules by Youtube evaluators [22]. In particular, the content, the subject matter and the language are analyzed. The subjects of Pierre Cassen's videos do not seem to give him access to this monetization. His referencing is therefore bad. The algorithm of Youtube does not put these videos forward, and this contributes to the non-dissemination of theories since only the subscribers of the channel are warned of the publication of a new video of his channel.

Infin, new methods of fighting Fake news have emerged on the Internet. Recently, the labelling of Fake news on Twitter [23] is an example of such measures. A study on the fight of social networks against the dissemination of false information, such as conspiracy theories, therefore seems to be necessary.

IV) References

- [1] "After the November 13 attacks, the fleau du complotisme," LeFigaro.fr, published January 01, 2015. Available at: <https://www.lefigaro.fr/actualite-france/2015/12/01/01016-20151201ARTFIG00046-apres-les-attentats-du-13-novembre-le-fleau-du-complotisme.php>
- [2] "How the coronavirus became a breeding ground for conspiracy theories conspiracy ". LeMonde.fr, published on 28 March 2020. Available at : https://www.lemonde.fr/politique/article/2020/03/28/le-coronavirus-terrain-fertile-pour-le-conspirationnisme_6034764_823448.html
- [3] "In Wuhan, the conspiracy tree hides a forest of evidence," Valeursactuelles.fr, published April 21, 2020. Available at: <https://www.valeursactuelles.com/societe/wu-han-l-arbre-du-complot-cache-une-foret-de-preuves-118394>
- [4] "Study: Breitbart-Led Right-Wing Media Ecosystem Altered Broader Media Agenda," Columbia Journalism Review, published March 3, 2017. Available at: <https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>
- [5] "Fake News and Health: Internet Users Not So Gullible," published June 16, 2019. Available at <https://theconversation.com/fake-news-et-sante-des-internautes-pas-si-credibles-118786>
- [6] "Taguieff unpacks conspiracy theories," LExpress.fr, published May 12 2018. Available at https://www.lexpress.fr/actualite/societe/taguieff-les-complots-repondent-a-un-be-soin-d-ordre_2007876.html
- [7] "BigPharma, 5G chips... The five most prevalent Covid-19 conspiracy theories," LAVDN.com, published October 6, 2020. Available at: <https://www.lavoixdunord.fr/875498/article/2020-10-06/big-pharma-puces-5g-les-cinq-theories-complotistes-on-the-covid-19-the-most>.
- [8] : "Covid-19: Why Bill Gates has become the target of so many conspiracy theories," lejdd.fr, published December 4, 2020. Available at: <https://www.lejdd.fr/Societe/covid-19-pourquoi-bill-gates-est-devenu-la-cible-de-tant-de-theories-du-complot-4010118>
- [9] : "Coronavirus: how Professor Didier Raoult became a central figure of conspiracy theories," LeMonde.fr, published March 28, 2020. Available at https://www.lemonde.fr/pixels/article/2020/03/28/comment-didier-raoult-est-devenu-la-nouvelle-egerie-des-complotistes_6034761_4408996.html

[10] FranceCulture : "In what way is 'Hold-up' a conspiracy documentary? FranceCulture.fr, published on November 18, 2020. Available at: <https://www.franceculture.fr/emissions/les-idees-claires-le-podcast/en-quoi-hold-e-is-it-a-complotiste-documentary>

[11] : "Billionaire Elon Musk refuses to get coronavirus vaccine and calls Bill Gates a 'moron,'" PLANETES360.co.uk, published October 1, 2020. Available at :

<https://planetes360.fr/le-milliardaire-elon-musk-refuse-de-se-faire-vacciner-contre-the-coronavirus-and-bill-gates-of-cretin/>

[12] "ANSM publishes its decision on the request for an RTU for hydroxychloroquine in the management of Covid-19 disease - Information Point," ansm.sante.fr, published October 23, 2020. Available at: <https://ansm.sante.fr/S-informer/Communiqués-Communiqués-Points-presse/L-ANSM-publishes-its-decision-on-application-for-an-RTU-for-hydroxychloroquine-in-the-management-of-Covid-19-disease-Information-Point>.

[13] 5G: the City of Nantes opens the debate", metropole.nantes.fr, published on 8 December 2020. Available at [: https://metropole.nantes.fr/actualites/2020/dialogue-citoyen/debat-ville-5g](https://metropole.nantes.fr/actualites/2020/dialogue-citoyen/debat-ville-5g).

[14] : Untitled. <https://www.youtube.com/watch?v=OwKU62WfsRg>²

[15] : "When conspiracists doubt the existence of Covid-19" - YouTube. Published on April 22, 2020 by the FRANCE 24 channel. Available at: <https://www.youtube.com/watch?v=3cwG8wxmSFg>.

[16] : "Covid-19: we're saved, the vaccine-viagra is coming!" - YouTube. Published on November 11, 2020 by the channel Riposte Laïque. Available at: <https://www.youtube.com/watch?v=3YCr8UOAo>.

[17] QAnon: At the roots of the conspiracy theory that is infecting America", LeMonde.fr, published on 14 October 2020. Available at https://www.lemonde.fr/pixels/article/2020/10/14/qanon-aux-racines-de-la-theorie-conspiracy-that-contaminates-america_6055921_4408996.html.

[18] Professor beheaded: Macron accomplice of the Islamists! - YouTube. Published on October 17, 2020 on the channel Riposte Laïque. Available at: <https://www.youtube.com/watch?v=m9CsBH9rO80>.

[19] : "2021 : Year of Dictatorship or French Surge?" - YouTube. Published on December 31, 2020 on the Riposte Laïque channel. Available at: <https://www.youtube.com/watch?v=ilH62gUxVIY>.

[20] : "Coronavirus. Donald Trump remains convinced of the benefits of hydroxychloroquine," Ouest-France.fr, published July 9, 2020. Available at: <https://www.ouest-france.fr/monde/etats-unis/donald-trump/coronavirus-donald-trump-remains-convinced-of-benefits-of-hydroxychloroquine-6901126>.

[21] The Illuminati, the plot that explains all the others Le Temps, published on 10 July 2020. Available at <https://www.letemps.ch/societe/illuminati-complot-explique-autres>.

[22] Rules for monetizing YouTube channels - YouTube Help. Available at: <https://support.google.com/youtube/answer/1311392?hl=en>.

[23] "Twitter to label 'fake news' related to coronavirus". Les Echos, 12 May 2020, <https://www.lesechos.fr/tech-medias/hightech/twitter-va->

labelliser-les-fake-news-lies-au-coronavirus-1202251.

² The "Hold-up" video was removed from Youtube before the fin of this research work.

APPENDIX

Appendix 1: Extract from a fichier

```
{
  "created_at": "Sat Nov 07 08:18:54 +0000 2020",
  "id": 1324989691024859100,
  "id_str": "1324989691024859137",
  "text": "Why did Bill Gates gather his buddies in October 2019 to simulate a
coronavirus pandemic worldwide?",
  "source": "<a href='\"https://mobile.twitter.com\"' rel='\"nofollow\"'>Twitter Web
App</a>",
  "truncated": false,
  "in_reply_to_status_id": null,
  "in_reply_to_status_id_str": null,
  "in_reply_to_user_id": null,
  "in_reply_to_user_id_str": null,
  "in_reply_to_screen_name": null,
  "user": {
    "id": 420599591,
    "id_str": "420599591",
    "name": "TriDeVo",
    "screen_name": "BorDQ",
    "location": "Your",
    "url": null,
    "description": null,
    "translator_type": "none",
    "protected": false,
    "verified": false,
    "followers_count": 12,
    "friends_count": 41,
    "listed_count": 0,
    "favorites_count": 452,
    "statuses_count": 177,
    "created_at": "Thu Nov 24 21:04:06 +0000 2011",
    "utc_offset": null,
    "time_zone": null,
    "geo_enabled": false,
    "lang": null,
```

```

    "contributors_enabled": false,
    "is_translator": false,
    "profile_background_color": "1A1B1F",
    "profile_background_image_url":
"http://abs.twimg.com/images/themes/theme9/bg.gif",
    "profile_background_image_url_https":
"https://abs.twimg.com/images/themes/theme9/bg.gif",
    "profile_background_tile": false,
    "profile_link_color": "2FC2EF",
    "profile_sidebar_border_color": "181A1E",
    "profile_sidebar_fill_color": "252429",
    "profile_text_color": "666666",
    "profile_use_background_image": true,
    "profile_image_url":
"http://pbs.twimg.com/profile_images/378800000179923159/e2b0ad5d080085a9fdb1ddebcef90b8_normal.jpeg",
    "profile_image_url_https":
"https://pbs.twimg.com/profile_images/378800000179923159/e2b0ad5d080085a9fdb1ddebcef90b8_normal.jpeg",
    "profile_banner_url":
"https://pbs.twimg.com/profile_banners/420599591/1374651661",
    "default_profile": false,
    "default_profile_image": false,
    "following": null,
    "follow_request_sent": null,
    "notifications": null
  },
  "geo": null,
  "coordinates": null,
  "place": null,
  "contributors": null,
  "is_quote_status": false,
  "quote_count": 0,
  "reply_count": 0,
  "retweet_count": 0,

  "favorite_count": 0,
  "entities": {
    "hashtags": [],
    "urls": [],

```

```
    "user_mentions": [],  
    "symbols": []  
  },  
  "favorited": false,  
  "retweeted": false,  
  "filter_level": "low",  
  "lang": "fr",  
  "matching_rules": [  
    {  
      "tag": null  
    }  
  ]  
}
```

Appendix 2: R codes for textual analysis Youtube

1. *Code R of the video "Covid-19 we are saved the vaccine-viagra is coming"*

```
##### TEXTUAL ANALYSIS

####Install and load working libraries library(tidyr)
library(stringr)
library(dplyr)
library(quanteda)
library(ggplot2)
install.packages("quanteda")
install.packages("ggplot2")
install.packages("dplyr")

##### 1- EXPLORATION OF THE COMMENT BASE

### Load the comments database

base<-read.csv2("videoinfo_3YCrS8U0_Ao&ab_channel=RiposteLaïque_2020_11_
15-11_43_58_comments.tab.csv")

##### 2- CREATION AND EXPLORATION OF THE CORPUS

### Create a corpus object
cp <- corpus(base$text,
              docvars = select(base, Date, authorName, likeCount),
              docnames = base$names)

tk<-tokens(cp, what = "word", remove_punct = TRUE,remove_symbols = TRUE
,remove_numbers= TRUE, remove_url=TRUE)

dfm<-dfm(cp, tolower=TRUE,remove_punct=TRUE, remove_symbols = TRUE,
remove_numbers=TRUE, remove_url = TRUE)

topfeatures(dfm, n=50)

lift_words <- topfeatures(dfm, n=100)
mean(lift_words) summary(lift_words)
```



```

##### 2.B Cleaning and statistics

###Remove linking words, conjunctions or prepositions by enriching
##the list of stopwords already downloadable in quanteda
toremove<-c(stopwords("french"), "a", "like", "of a", "of a", "also",
"does",
        "be", "it", "an", "do", "say", "if", "that",
        "where", "all", "more", "again", "already", "depuis",
        "years", "between", "is", "can", "of which", "therefore",
        "so", "must", "will", "therefore", "all", "then",
        "at", "times", "when", "also", "has", "has not", "that",
        "that", "the one", "have not",
        "the", "the", "that", "or", "from here", "if", "there", "as
        soon as",
        "dit", "pu", "six", "pu", "font", "ceux", "peut",

        "little", "less", "after", "well", "two", "three", "yes",
        "before", "it", "has", "especially", "so", "can",
        "according to", "some", "always", "have", "because", "many",
        "under", "no", "other", "against", "many",
        "other", "any", "end", "time",
        "Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
        "Saturday",
"Sunday",

        "in", "not", "me", "our", "we", "of", "you", "without",
"but", "okay",

        "see", "because", "say", "said", 'go', "nothing", "that
        they",
"what", "just",

        "you're", "ah",

        "why", "too", "can", "less", "since", "under",
"see", "want",

        "will", "really", "have", "go", "bla", "e", "to be",
"months",

"s'en", "<", ">", "br", "href", "=", "\U0001f44d", "quot", "#39", "\U0001f44f", "\U0001f
64f",

"\U0001f602", "\U0001f64f", "\U0001f44f", "c3", "\U0001f44f", "\U0001f64f", "\U0001f
602", "speak",

"de", "#39", "le", "la", "et", "les", "pour", "pas", "en", "est", "br", "bon", "b", "vu", "a
llez", "regardez",

```

```

"prendre", "pense", "veulent", "salut", "lol", "voilà", "\U0001f923", "\U0001f1eb\U00
01f1f7",

"\U0001f1f2\U0001f1eb", "\U0001f471\U0001f3fb9", "\U0001f413", "\U0001f471\U0001f
3fb9", "\U0001f609",

"\U0001f914", "\U0001f603", "\U0001f644", "\U0001f471\U0001f3fb9", "toujours", "pre
mier",

"come", "go", "arrive", "soon", "jadot", "alain", "good", "some", "especially"
,

"so", "great", "hello", "go", "all", "thing", "mr", "mdr", "none", "never"
)

##Create a dictionary of the most used expressions from
textstat_collocations
textstat_collocations(tk, min_count = 5, size = 2L) %>%
  arrange(desc(count)) %>%
  slice(1:500)

dict <- dictionary(list(
  covid= c("covid", "covid-19", "coronavirus", "corona"),
  vaccine=c("vaccine" , "vaccines") ,
  vaccinate=c("vaccinate", "vaccinated", "vaccinated"),
  Pierre_CASSEN=c("pierre", "cassen"), people=c("person", "people",
"individuals", "peoples"), death=c("death", "deaths"),
  Bill_GATES=c("Bill", "gates"),
  conspiracy=c("conspiracy",
"conspiracies"), politics=c("politics",
"policies"))
)

## Recreate the matrix and token objects, this time removing the
uninteresting words
## Draw the word cloud

tk <- tokens_remove(tk, toremove)

```

```
dfm<-dfm(cp, tolower=T, remove_punct=T, remove_numbers=T, remove_url =  
T, remove= toremove, thesaurus=dict)  
topfeatures(dfm, n=100)  
  
textplot_wordcloud(dfm, min_count=10)  
#vary the min_count to have more or less words on the graph
```

2. R code of the video "When conspiracy theorists doubt the existence of Covid-19"

```
##### TEXTUAL ANALYSIS

#### Install and load working libraries library(tidyr)
library(stringr)
library(dplyr)
library(quanteda)
library(ggplot2)
install.packages("quanteda")
install.packages("ggplot2")
install.packages("dplyr")

##### 1- EXPLORATION OF THE COMMENT BASE

### Load the comments database

base<-read.csv2("videoinfo_3cwG8wxmSFg&ab_channel=FRANCE24_2020_11_15-11_37_16
_comments.tab.csv")

### Create a corpus object
cp <- corpus(base$text,
             docvars = select(base, Date, authorName, likeCount),
             docnames = base$names)

tk<-tokens(cp, what = "word", remove_punct = TRUE,remove_symbols = TRUE
,remove_numbers= TRUE, remove_url=TRUE)

dfm<-dfm(cp, tolower=TRUE,remove_punct=TRUE, remove_symbols = TRUE,
remove_numbers=TRUE, remove_url = TRUE)

topfeatures(dfm, n=50)

lift_words <- topfeatures(dfm, n=100)
mean(lift_words) summary(lift_words)

##### 2.B Cleaning and statistics
```

###Lifting words of liasons, conjunctions or prepositions by enriching

```
to remove <- c(stopwords("french"), "a", "as", "of a", "of a", "also",  
"does",  
"be", "it", "an", "do", "say", "if", "that",  
"where", "all", "more", "again", "already", "depuis",  
"years", "between", "is", "can", "of which", "therefore",  
"so", "must", "will", "therefore", "all", "then",  
"at", "times", "when", "also", "has", "has not", "that",  
"that", "the one", "have not",  
"the", "the", "that", "or", "from here", "if", "there", "as soon  
as",  
"dit", "pu", "six", "pu", "font", "ceux", "peut",  
  
"little", "less", "after", "well", "two", "three", "yes",  
"before", "it", "has", "especially", "so", "can",  
"according to", "some", "always", "have", "because", "many",  
"under", "no", "other", "against", "many",  
"other", "any", "end", "time",  
"Monday", "Tuesday", "Wednesday", "Thursday", "Friday",  
"Sunday",  
"Saturday", "in", "not", "me", "our", "we", "of", "you",  
"okay."  
"without", "but."  
"just."  
"see", "because", "say", "said", "will", "nothing", "that they",  
"what",  
  
"why", "too", "can", "less", "since", "under", "are",  
"ah", "see",  
"will", "really", "there", "go", "bla", "e", "to be", "want",  
"month",  
  
"s'en", "<", ">", "br", "href", "=", "\U0001f44d", "quot", "#39", "\U0001f44f", "\U0001f  
64f",  
  
"\U0001f602", "\U0001f64f", "\U0001f44f", "c3", "\U0001f44f", "\U0001f64f", "\U0001f  
602", "speak",  
  
"de", "#39", "le", "la", "et", "les", "pour", "pas", "en", "est", "br", "bon", "b", "vu", "a  
llez", "regardez",  
  
"take", "think", "want", "hi", "lol", "here", "@corentin", "@max", "thanks",
```

```

"km", "especially", "@christian", "@dobonne", "n'ai", "better", "the", "sabrina",
"gt", "must", "things", "other", "never", "until", "how", "what", "thing",
"choses", "point", "\U0001f601", "besoin", "nombre", "existe", "martin", "manuel",
"sai", "big", "always", "stay", "shape", "moment", "†", "totally", "pass", "r
egard"
)

##Create a dictionary of the most used expressions from textstat_collocations
textstat_collocations(tk, min_count = 10, size = 2L) %>%
  arrange(desc(count)) %>%
  slice(1:500)

dict <- dictionary(list(
  covid=c("covid", "covid 19", "coronavirus"),
  vaccine=c("vaccine", "vaccines", "vaccinate",
"vaccinated*"), believe=c("believe", "believe",
"believe"), people=c("people", "people", "person",
"people"), hold_up=c("hold up", "report"),
evidence=c("evidence", "proof"), mask=c("mask",
"masks"), containment=c("containment", "contain"),
truth=c("know", "truth", "know", "true"),
theory=c("theory", "theories"), video=c("videos",
"video")
)

## Recreate the matrix and token objects, this time removing the uninteresting
words
## Draw the word cloud

tk <- tokens_remove(tk, toremove)
dfm<-dfm(cp, tolower=T, remove_punct=T, remove_numbers=T, remove_url = T,
remove= toremove, thesaurus=dict)
topfeatures(dfm, n=100)

textplot_wordcloud(dfm, min_count=10)
#vary the min_count to have more or less words on the graph

```



3. Code R of the "Hold-up" video

```
##### TEXTUAL ANALYSIS

#### Install and load working libraries library(tidyr)
library(stringr)
library(dplyr)
library(quanteda)
library(ggplot2)
install.packages("quanteda")
install.packages("ggplot2")
install.packages("dplyr")

##### 1- EXPLORATION OF THE COMMENT BASE

### Load the comments database

base<-read.csv2("videoinfo_CeHQQSTQMeU&ab_channel=WhataBeautifulWorldYoutubeCh
annel_2020_11_26-14_45_28_comments.tab.csv")

##### 2- CREATION AND EXPLORATION OF THE CORPUS

### Create a corpus object
cp <- corpus(base$text,
             docvars = select(base, Date, authorName, likeCount),
             docnames = base$names)

tk<-tokens(cp, what = "word", remove_punct = TRUE,remove_symbols = TRUE
,remove_numbers= TRUE, remove_url=TRUE)

dfm<-dfm(cp, tolower=TRUE,remove_punct=TRUE, remove_symbols = TRUE,
remove_numbers=TRUE, remove_url = TRUE)

topfeatures(dfm, n=50)
```



```

lift_words <- topfeatures(dfm, n=100)
mean(lift_words) summary(lift_words)

##### 2.B Cleaning and statistics

toremove<-c(stopwords("french"), "a", "as", "of a", "of a", "also",
"does",
      "be", "it", "an", "do", "say", "if", "that",
      "where", "all", "more", "again", "already", "depuis",
      "years", "between", "is", "can", "of which", "therefore",
      "so", "must", "will", "therefore", "all", "then",
      "at", "times", "when", "also", "has", "has not", "that",
      "that", "the one", "have not",
      "the", "the", "that", "or", "from here", "if", "there", "as soon
      as",
      "dit", "pu", "six", "pu", "font", "ceux", "peut",
      "little", "less", "after", "well", "two", "three", "yes",
      "before", "it", "has", "especially", "so", "can",
      "according to", "some", "always", "have", "because", "many",
      "under", "no", "other", "against", "many",
      "other", "any", "end", "time",
      "Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
"Sunday",
      "Saturday", "in", "not", "me", "our", "we", "of", "you",
"okay."
      "without", "but."
"just."
      "see", "because", "say", "said", 'will', "nothing", "that they",
      "what",
      "why", "too", "can", "less", "since", "under", "are",
"ah", "see",
      "will", "really", "there", "go", "bla", "e", "to be", "want",
"month",
"s'en", "<", ">", "br", "href", "=", "\U0001f44d", "quot", "#39", "\U0001f44f", "\U0001f
64f",
"\U0001f602", "\U0001f64f", "\U0001f44f", "c3", "\U0001f44f", "\U0001f64f", "\U0001f
602", "speak",
      "amp", "the", "case", "gives", "until", "want", "seen", "now",
      "look", "sabrina", "gt", "b", "put", "absolutely", "while",
"know", "except", "how",

```

```

"good," "finally," "quickly," "when," "soon," "point," "put," "♡," "must,"
"good," "al lez," "all."

"some", "always", "other", "other", "none", "q", "small", "large", "time"
)

##Create a dictionary of the most used expressions from textstat_collocations
textstat_collocations(tk, min_count = 10, size = 2L) %>%
  arrange(desc(count)) %>%
  slice(1:500)

dict <- dictionary(list(
  covid= c("covid", "covid 19", "coronavirus", "virus", "COVID"),
          hold_up=c("hold up", "hold-up", "documentary*",
"film*", "report*", "article*"),
  people = c("people*", "person*", "humanity*", "people", "peoples"),
  truth = c("true", "truth"),
  vaccine=c("vaccine", "vaccines"),
  politics=c("politics*", "politician*"),
  world=c("world", "global", "planet"),
  humanity=c("humanity", "human", "human*"),
  conspiracy=c("conspiracy*", "conspiracy")
)

## Recreate the matrix and token objects, this time removing the uninteresting
words
## Draw the word cloud

tk <- tokens_remove(tk, toremove)
dfm<-dfm(cp, tolower=T, remove_punct=T, remove_numbers=T, remove_url = T,
remove= toremove, thesaurus=dict)
topfeatures(dfm, n=100)

#textplot_wordcloud(dfm, min_count=20) ==> 20 too big textplot_wordcloud(dfm,
min_count=9)
#vary the min_count to have more or less words on the graph

```

Appendix 3: Link from GitHub

All of the code, high-definition visualizations, aspirated tweets, and assembled dataframes for Twitter aspirations are available on GitHub :

https://github.com/bapt380/IC05_Covid19