

# **IC05 : Analyse critique des données numériques**

**Semestre : Automne 2020**

## **Les théories du complot autour de la Covid-19**



**utec**

**Responsable de l'UV :**  
BELLON Anne

**Étudiants :**  
DUTHEIL Rémi  
LAUBY Constance  
MISSAOUI Benjamin  
VIERA Baptiste

<b>I) Introduction</b>	<b>3</b>
I.1) Présentation du sujet	3
I.2) Enjeux et actualités	4
I.3) Etat de la connaissance	4
I.4) Cadre des recherches	6
<b>II) Méthodes et résultats</b>	<b>7</b>
II.1) Twitter	7
II.1.a) Méthode de travail	7
II.1.b) Aspiration et résultats	10
II.1.c) Analyses des résultats : Twitter	14
II.1.d) Pour aller plus loin : la cartographie du réseau complotiste	17
II.2) Youtube	21
II.2.a) Méthode de travail	21
II.2.b) Aspiration et résultats	23
II.2.c) Analyse des résultats : Youtube	24
II.3) Analyse croisée Twitter/Youtube	28
<b>III) Analyse critique</b>	<b>29</b>
III.1) Biais de données	29
III.2) Améliorations	30
<b>IV) Conclusion</b>	<b>31</b>
<b>IV) Références</b>	<b>33</b>
<b>ANNEXE</b>	<b>35</b>
Annexe 1 : Extrait d'un fichier JSON.	35
Annexe 2 : Codes R pour l'analyse textuelle Youtube	38
Code R de la vidéo "Covid-19 on est sauvés le vaccin-viagra arrive"	38
Code R de la vidéo "Quand des complotistes doutent de l'existence du Covid-19"	42
Code R de la vidéo "Hold-up"	46
Annexe 3 : Lien du GitHub	49

# I) Introduction

## I.1) Présentation du sujet

Depuis fin 2019, le monde est frappé par une pandémie causée par le coronavirus SARS-CoV-2, aussi appelé Covid-19. Dans une ambiance anxiogène liée à la peur de tomber malade et aux mois de confinement, une montée des théories du complot peut être observée dans le monde, et particulièrement en France. Ces théories concernent différents sujets autour de la Covid-19 :

- **Origine du virus** : a-t-il été créé par l'homme, est-ce une arme bactériologique échappée d'un laboratoire chinois ?
- **Utilisation de la chloroquine** pour lutter contre le virus : est-ce efficace, pourquoi son utilisation n'est-elle pas généralisée ?
- **Course au vaccin** : sera-t-il obligatoire ? Est-ce un moyen de contrôler la population en injectant, en même temps que le vaccin, d'autres produits ?
- **Efficacité du port du masque** : est-il réellement efficace ? Est-ce un moyen de privation de liberté mis en place par le gouvernement ?
- **Saturation des hôpitaux** : sont-ils vraiment saturés par les patients atteints de la Covid-19 ?

A l'aide des nouvelles technologies et des nouveaux moyens de communication, l'opinion de chacun peut être partagée en un clic. Accessible facilement et offrant une certaine visibilité, les réseaux sociaux tels que Twitter, Facebook ou encore Youtube apparaissent comme le lieu idéal pour la diffusion de théories du complot autour de la Covid-19. Dans le cadre de l'UV **IC05 : Analyse critique des données numériques**, notre groupe a décidé d'étudier la propagation des théories du complot autour du coronavirus sur deux réseaux : Youtube et Twitter. Les réseaux sociaux sont un support idéal pour la liberté d'expression, il nous semble nécessaire d'étudier comment ces réseaux permettent la diffusion des théories. En sélectionnant plusieurs thèmes récurrents dans la sphère complotiste, nous voulons identifier les acteurs principaux dans la propagation des théories du complot.

Nous commencerons par une étude préliminaire d'articles pour délimiter plus précisément notre sujet d'étude, mais également pour définir les premiers acteurs et relais et approcher de la sphère complotiste. Ensuite, nous ferons plusieurs aspirations de données sur Twitter et Youtube pour construire une analyse de réseau. Enfin, nous conclurons par une analyse critique de ces données.

## I.2) Enjeux et actualités

En plein cœur de notre actualité depuis maintenant un an, une étude de l'impact de la pandémie de la Covid-19 présente un intérêt majeur. Le sujet de la Covid-19 génère énormément de réactions de la part des individus. L'incertitude de trouver un vaccin rapidement, la frustration liée au confinement, l'impact économique des mesures sanitaires prises pour endiguer cette pandémie participent au développement des théories autour du virus. De plus, ces théories sont particulièrement présentes sur les réseaux sociaux. Les médias traditionnels ne relaient que très peu d'informations sur ces théories, où bien elles sont réfutées. Nous trouvons, par exemple, des vidéos Youtube réalisées par les chaînes *France 24*, *RTL* ou encore *TV5MONDE Info* qui dénoncent ces théories. La liberté d'expression sur ces supports numériques semble justifier l'utilisation privilégiée des réseaux sociaux par les complotistes. Cependant, ce phénomène ne date pas de l'année 2020. Internet s'est déjà vu inondé de théories du complot, depuis l'attentat du 11 septembre 2001 par exemple. Chaque attentat entraîne une vague de théories du complot qui se propagent sur le Web. Ce fut le cas par exemple lors des attentats du 13 novembre 2015 à Paris, avec notamment "plusieurs vidéos Youtube complotistes comptaient plus de 100 000 vues" d'après un article du Figaro (2015) [1].

## I.3) Etat de la connaissance

Tout d'abord, nous avons réalisé une première recherche bibliographique sur le sujet de la sphère complotiste en général, ainsi que sur les nouvelles théories liées à la Covid-19. Cette étape permet de sélectionner les thèmes principaux sur lesquels nous allons focaliser notre recherche. Si plusieurs médias s'intéressent aux théories du complot autour de la Covid-19, c'est principalement dans le but de les dénoncer, ou bien d'étudier l'impact et la répartition de ces théories dans la population. Un article du Monde (2020) [2] tente, par exemple, de relier l'adhésion d'un individu aux théories du complot autour du Covid et son adhésion à un parti politique. Ainsi, 36% des sympathisants du *Rassemblement National* valideraient l'existence d'un « complot sioniste à l'échelle mondiale ». Cependant, des exceptions au sein des médias sont fréquentes comme le montre cet article de *Valeurs Actuelles* (2020) [3] qui est, à l'inverse de celui du *Monde*, bien moins critique envers ces théories du complot.

La lecture d'articles plus généraux autour de la sphère complotiste nous permet de confirmer notre hypothèse que les réseaux sociaux sont des lieux favorables pour la diffusion de théories complotistes, comme le montre cet article de *Columbia Journalism Review* (2017) [4]. Cet article parle du média américain d'extrême droite *Breitbart* et de son impact sur la présidentielle américaine de 2016. Il met en lumière plusieurs éléments : *Breitbart* a créé tout un écosystème autour de la sphère d'extrême droite américaine en utilisant majoritairement les réseaux sociaux. Le média a diffusé plusieurs articles trompeurs sur cette élection présidentielle. *Breitbart*, un média pro-Trump, est également partisan des théories du complot. L'article *Study: Breitbart-led right-wing media ecosystem*

*altered broader media agenda* (2017) [4] révèle l'écrasante utilisation de Twitter et Facebook pour la diffusion de contenus complotistes. À l'inverse, les partisans d'Hillary Clinton se tournaient, eux, beaucoup plus vers les médias traditionnels (télé et journaux). Ces articles volontairement ambigus peuvent donc être une source de désinformation.



Cependant, cette diffusion n'est peut-être pas aussi optimisée que nous le pensions. Par exemple, *The Conversation* (2019) [5] explique dans cet article que les informations trompeuses diffusées sur les réseaux sociaux ne sont pas les plus propagées. Des contenus visuels de la page Facebook *Santé+Magazine*, comme le montre la figure 1 ci-contre, génèrent 2 à 4 fois plus d'interactions que d'autres contenus. Les posts à base de "proverbes consensuels sur les relations sociales" sont ceux avec le plus d'interactions.

**Figure 1 :** Publication Facebook de la page Santé + Magazine.

Si les théories liées au coronavirus sont nombreuses et concernent de multiples sujets, les théories du complot ne sont pas apparues avec la pandémie de 2020. Les vaccins en général sont par exemple la cible depuis longtemps de théories complotistes. L'essai de l'ex-médecin Québécoise Guylaine Lanctôt, *La Mafia Médicale* sorti en 1996, montre que la méfiance vis-à-vis des vaccins et des antibiotiques sont des phénomènes anciens. Le Covid-19 ne semble être qu'un nouveau tremplin pour ces théories, une nouvelle façon de les renouveler au goût de l'actualité. Nous pouvons également citer les travaux du politologue français [6] Pierre-André Taguieff qui décrivent les théories du complot comme étant "*un appel au peuple contre les élites et une sacralisation du peuple, paré de toutes les vertus, censé être trompé en permanence par les élites cyniques et corrompues*". C'est bien dans cette croyance d'être trompé, manipulé et contrôlé par les personnes ayant du pouvoir, les gouvernements ou les grands chefs d'entreprises par exemple, que s'inscrivent les théories du complot. Ce sentiment d'être contrôlé et que la vérité est cachée par de puissantes personnes est particulièrement mis en avant par les complotistes, qui se font souvent censurer.

Après cette analyse, le profil des complotistes semble se préciser : il s'agit souvent de personnes extrêmement méfiantes vis-à-vis du gouvernement et qui n'ont aucune confiance dans les médias traditionnels. De plus, comme le montre l'article du *Monde* (2020) [1] cité précédemment, il semble avoir un lien entre

l'adhésion à un parti politique d'un individu et son adhésion aux théories du complot sur le coronavirus. Les sphères d'extrême gauche et surtout d'extrême droite semblent liées aux sphères complotistes, que ce soit en France ou aux Etats-Unis.

## I.4) Cadre des recherches

Nous allons définir plusieurs hypothèses et axes de recherches dans le cadre de notre projet. Le climat particulièrement anxiogène provoqué par la Covid-19 peut expliquer l'intérêt, le nombre et l'impact des théories du complot sur ce virus. Malgré les nombreux sujets présents au sein des théories du complot autour du Covid-19, nous avons décidé pour des raisons pratiques de nous focaliser sur les sujet suivant :

- **5G et Covid-19** : cette théorie explique que la Covid-19 permettrait de déployer la 5G avec des puces électroniques 5G implantées à l'aide du vaccin. Ces puces serviraient à nous traquer [7],
- **Le rôle de Bill Gates dans l'origine du virus** : d'après cette théorie, le coronavirus servirait les intérêts de Bill Gates [8],
- **L'hydroxychloroquine** : cette théorie dit que le traitement mis en place par le docteur Didier Raoult serait efficace pour lutter contre la Covid-19 mais qu'il ne serait pas diffusé car utiliser un médicament déjà existant ne serait pas intéressant pour l'industrie pharmaceutique [9].

Nous faisons particulièrement le choix de ne pas traiter les théories anti-vaccin autour de la Covid-19. En effet, la sphère complotiste autour des vaccins en général est très vaste. Elle ne concerne donc pas uniquement le coronavirus. Nous allons également nous restreindre à la sphère complotiste en France. Enfin, nous nous focaliserons sur l'analyse de données à partir de Youtube et de Twitter.

### Objectifs de l'étude

- Les théories sont souvent accompagnées d'acteurs, de personnalités publics qui sont soit, partisans des théories du complot, soit la cible de théories. Dans le cadre de ce projet, nous allons essayer d'identifier certaines d'entre elles, ainsi que les thèmes traités dans les théories du complot autour de la Covid-19.
- La diffusion des théories du complot et les réactions qu'elles provoquent sur les réseaux sociaux ont souvent lieu à la suite d'événements d'actualités précis. Nous allons essayer d'observer ce phénomène.
- Enfin, nous nous attendons à observer un effet de communauté, de groupe, dans la diffusion des théories du complot. Si des individus complotistes partagent des théories autour de la Covid-19 sur les réseaux sociaux, cette diffusion doit s'effectuer principalement entre individus complotistes.

## II) Méthodes et résultats

Notre analyse est réalisée à partir de deux supports numériques : Youtube et Twitter. Pour Twitter, nous avons décidé de nous focaliser sur les trois théories du complot citées précédemment : **le lien entre la Covid-19 et la 5G, la place de Bill Gates dans l'origine du virus et enfin l'intérêt de l'hydroxychloroquine**. Parallèlement au travail réalisé sur Twitter, nous ferons une analyse textuelle des commentaires de trois vidéos Youtube. Chaque vidéo n'a pas pour sujet une théorie en particulier. Elles traitent des théories du complot en général autour de la Covid-19 (utilisation des masques, rôle du confinement, origine du virus, Bill Gates, etc...).

### II.1) Twitter

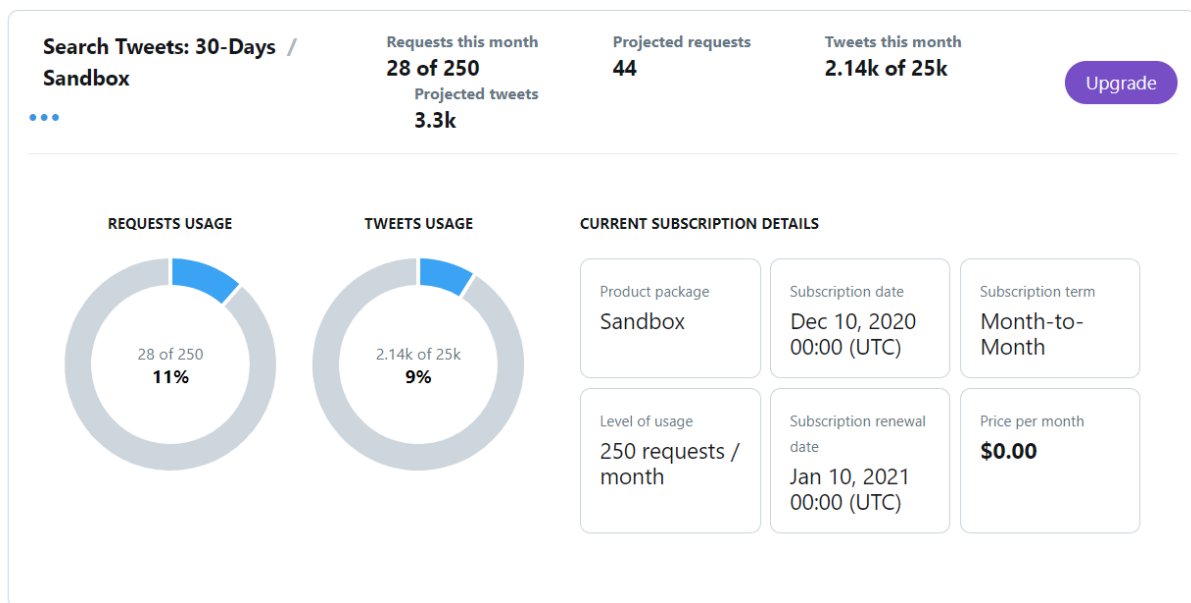
#### *II.1.a) Méthode de travail*

Le travail réalisé à partir de Twitter nous a permis d'aspirer plusieurs milliers de tweets. Les tweets ont été aspirés selon un ordre chronologique, du 6 octobre jusqu'au 6 décembre 2020. Les codes Python sont disponibles en annexe (annexe 3).

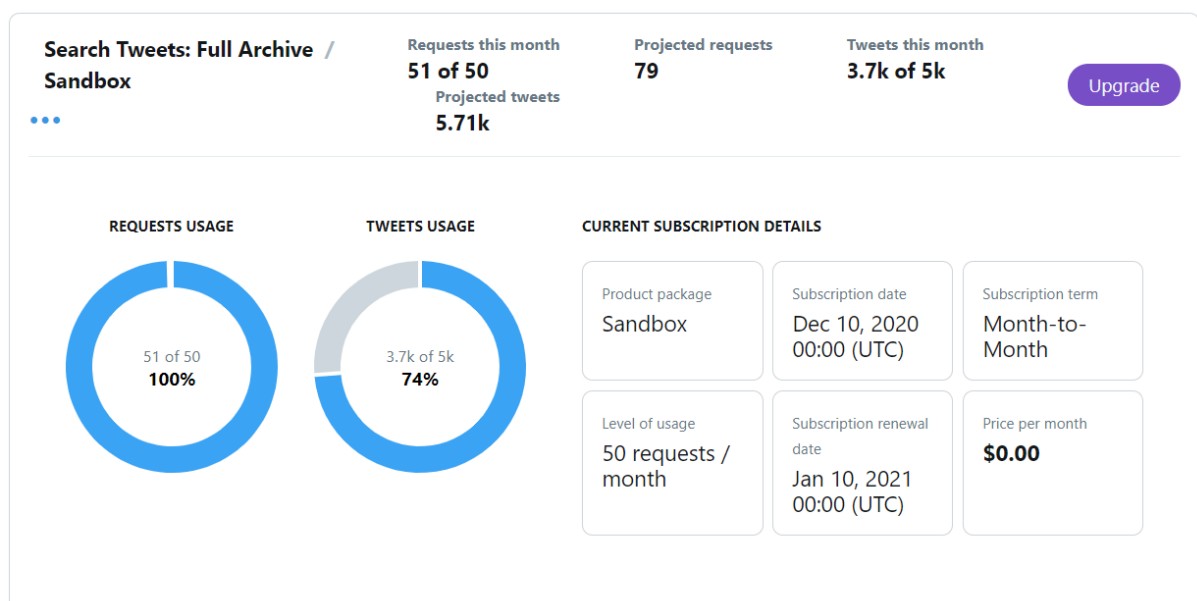
Pour extraire les données, nous avons dans un premier temps essayé d'utiliser l'outil **Gazouilloire** développé par le Media Lab de Sciences Po. Cependant, ce fût rapidement un échec. Nous avons été confrontés à une série de problèmes car l'outil est développé en Python 2 et utilise des modules qui ne sont plus à jour. Cela engendre de nombreuses complications lors de l'exécution du programme. Malgré un essai pour convertir tous les codes python 2 en python 3, le programme n'a pas fonctionné.

Dans un second temps, nous nous sommes dirigés vers un outil différent : un module python appelé **Tweepy**. Ce module nous permet de récupérer des tweets à partir de mots-clés mis en paramètre (exemple : hydroxychloroquine, coronavirus). Cependant, du fait de la jeunesse de ce module, nous n'avions pas accès à toutes les fonctionnalités que l'API de Twitter pouvait nous offrir. En utilisant Tweepy, nous ne pouvions récupérer que des tweets datant de 7 jours au maximum. Cette contrainte temporelle était bien trop forte et nous obligeait à faire chaque semaine des requêtes, ce que nous avons jugé peu productif. Notre objectif était de récupérer des tweets sur une période de deux mois. Nous avons donc pris la décision d'abandonner le module Tweepy.

Malgré ces deux échecs, il restait toute de même une solution : utiliser l'API de Twitter directement, en utilisant les lignes de commande. Cette fois-ci, nous n'avons pas rencontré de problèmes en particulier. L'API de Twitter nous a permis de récolter des tweets datant de 30 jours au maximum et sans contrainte de date limite, en utilisant respectivement l'outil **Search Tweets: 30-Days / Sandbox** et l'outil **Search Tweets: Full Archive / Sandbox**. Nous étions cependant limités dans le nombre de requêtes et de tweets que nous pouvions exécuter et récolter, comme le montre les figures ci-dessous :



**Figure 2 :** Limites du nombre de requêtes autorisées par l'API avec l'outil *Search Tweets: 30-Days / Sandbox*.



**Figure 3 :** Limites du nombre de requêtes autorisées par l'API avec l'outil *Search Tweets: Full Archive / Sandbox*.

Les contraintes étaient particulièrement fortes pour le second outil *Search Tweets: Full Archive / Sandbox* car nous ne pouvions faire que 50 requêtes. De plus, chaque requête permettait de récupérer au maximum 100 tweets, ce qui est relativement faible au vu du nombre de tweets généré en une journée par exemple. Pour palier à ce problème, nous avons ainsi créé 8 comptes Twitter développeur afin de subvenir à nos besoins.

Le code pour lancer les requêtes twitter avait la forme suivante :



```

curl --request POST \
  --url https://api.twitter.com/1.1/tweets/search/30day/dev30Days.json \
  --header 'authorization: Bearer
AAAAAAAAAAAAAAAAAAAAALB7JgEAAAAAnmfvk1wZB6e1kRxyhFqXz6gPZrw%3
D0uEiTXM6WCeiJFe2JHGZFIJ60Vp157FJwkhjreUUWWT89E3z7H' \
  --header 'content-type: application/json' \
  --data '{
    "query":"hydroxychloroquine coronavirus lang:fr",
    "maxResults": "100",
    "fromDate":"20201120000",
    "toDate":"20201121159"
  }' > 12Nov_00_12.json

```

**Figure 4 :** Code en bash permettant de lancer les requêtes twitter.

#### **Explication du code :**

**--url** : spécifie quel type de requête nous souhaitons utiliser, “30day” ou “fullarchive”.

**--header** : inscrit le compte twitter développeur en renseignant simplement le *BearerToken*.

**--header** : le fichier obtenu sera de la forme *json*.

**--data** :

- Dans la partie “**query**”, nous spécifions les mots-clés de la façon suivante, [nom de la théorie coronavirus](#), ainsi que la langue des tweets que nous souhaitons récupérer, qui est le français dans notre cas.
- Dans la partie “**maxResults**”, nous fixons le maximum de tweets que nous souhaitons récupérer, sachant que nous ne pouvons pas aller au-delà de 100 tweets.
- Dans les parties “**fromDate**” et “**toDate**”, nous spécifions la plage sur laquelle nous souhaitons récupérer les tweets suivant ce format précis : années/mois/jours/heures/minutes.

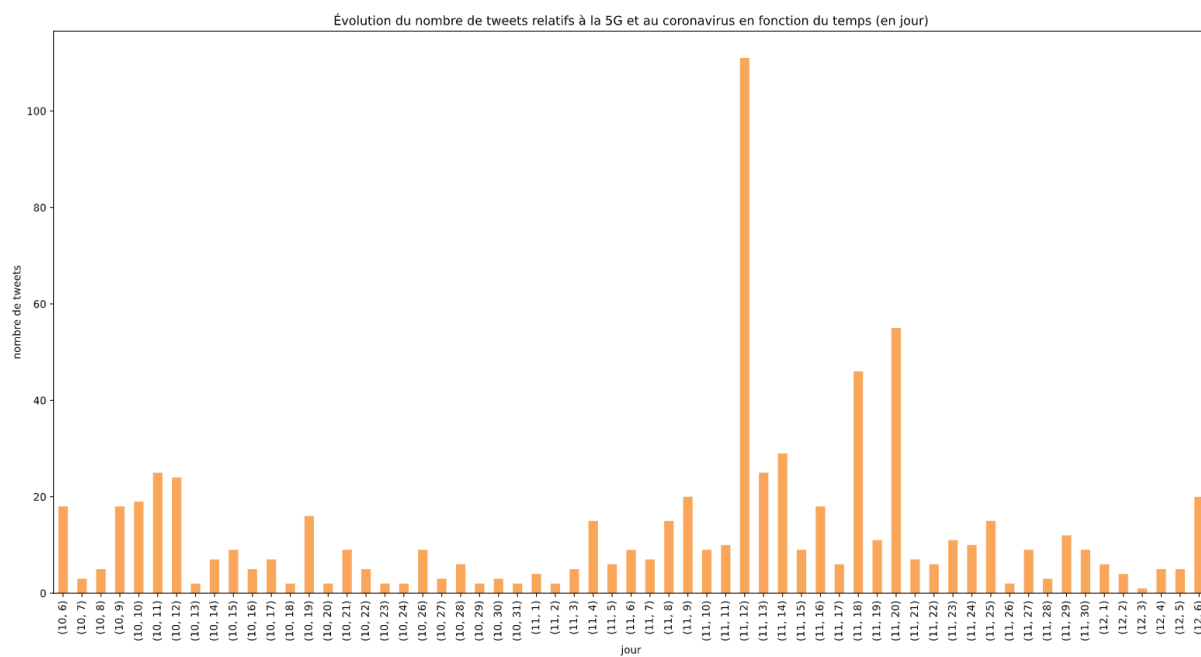
### ***II.1.b) Aspiration et résultats***

Nous obtenons des fichiers *JSON* qui contiennent un certain nombre de tweets (au maximum 100). Un extrait d'un fichier *JSON* est disponible en annexe (annexe 1).

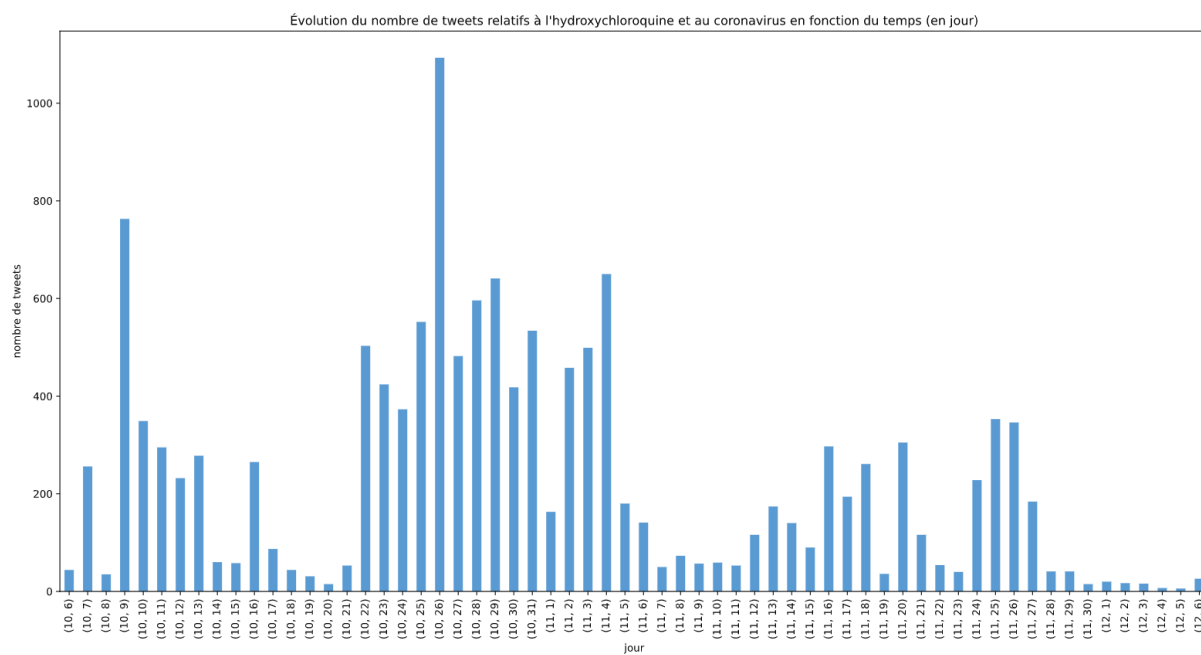
Ensuite, à l'aide de différentes fonctions en python, nous avons regroupé tous les fichiers *JSON* pour chacune de nos trois théories dans une dataframe. Nous obtenons *in fine* trois bases de données correspondant aux trois théories (hydroxychloroquine, 5G, Bill Gates). Chacune de ces bases de données contiennent les tweets et leurs caractéristiques, qui sont nombreuses comme le souligne l'exemple de l'annexe 1.

A partir des ces bases de données, nous allons faire deux analyses distinctes. Le premier traitement consiste à la réalisation de plusieurs histogrammes qui représentent l'évolution du nombre de tweets en fonction du temps pour chacune des théories. De plus, un diagramme circulaire permet de représenter la répartition en pourcentage des tweets pour chacune des trois théories traitées ici.

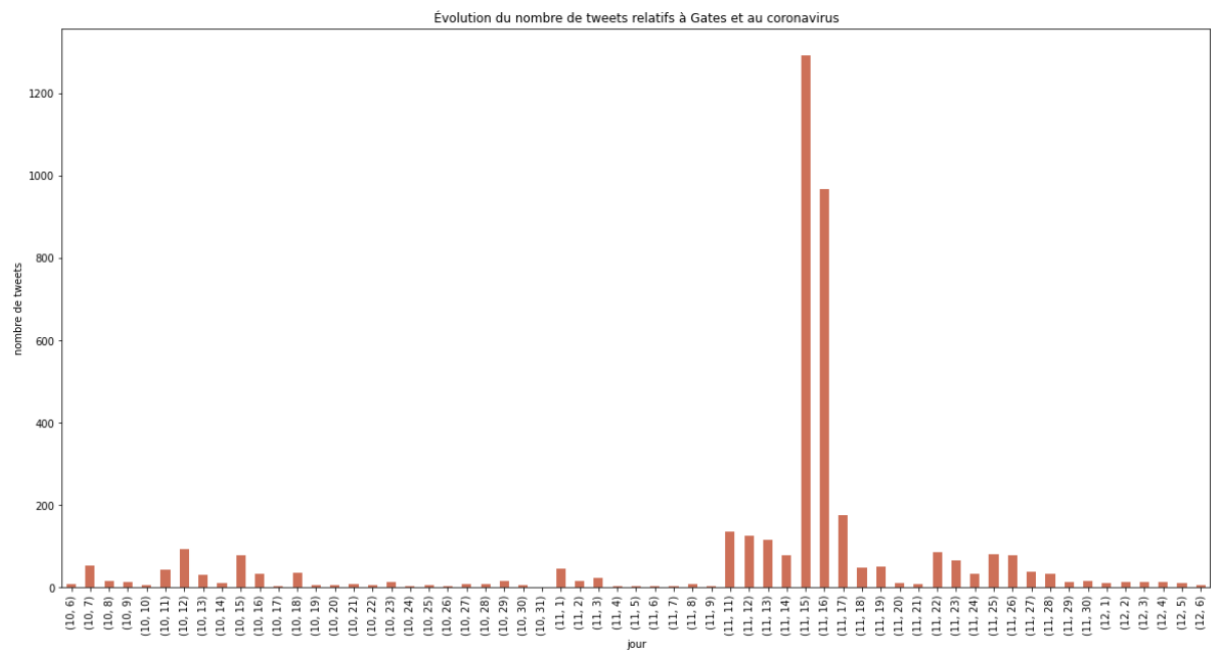
Les graphiques obtenus sont les suivants :



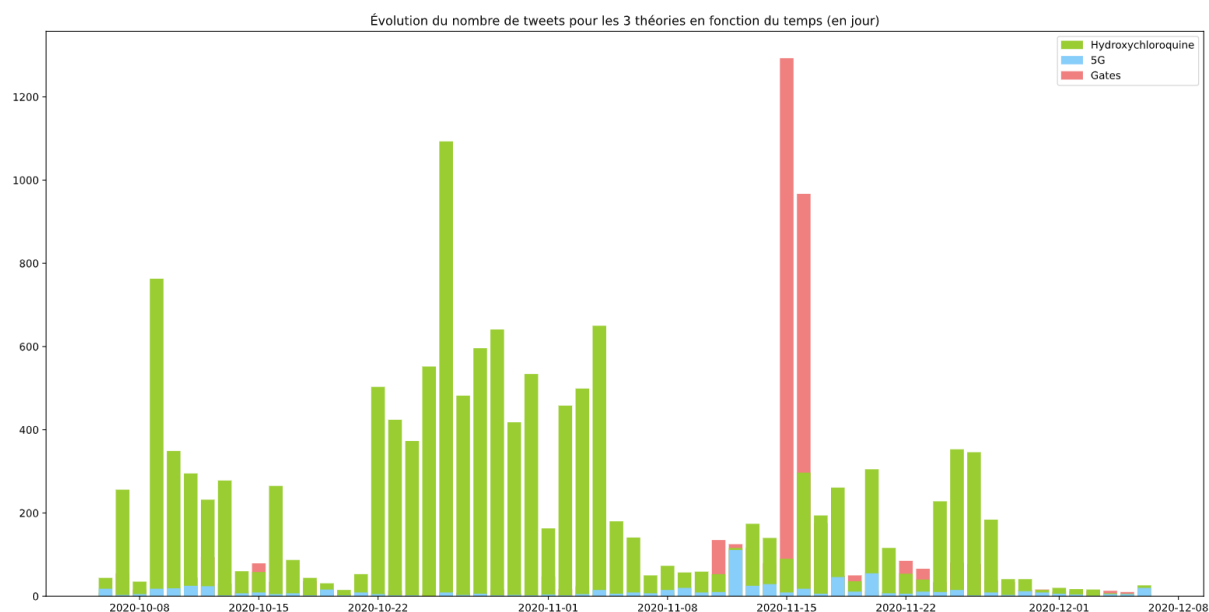
**Figure 6 :** Évolution du nombre de tweets relatifs à la 5G et au coronavirus.



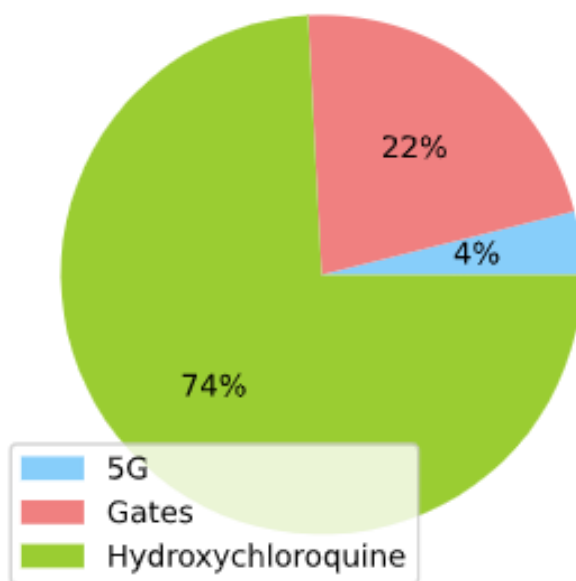
**Figure 7 :** Évolution du nombre de tweets relatifs à l'hydroxychloroquine et au coronavirus.



**Figure 8 :** Évolution du nombre de tweets relatifs à Gates et au coronavirus.



**Figure 9 :** Évolution du nombre de tweets relatifs à 3 théories et au coronavirus.



**Figure 10 :** Proportion (%) des tweets par théories.

### **II.1.c) Analyses des résultats : Twitter**

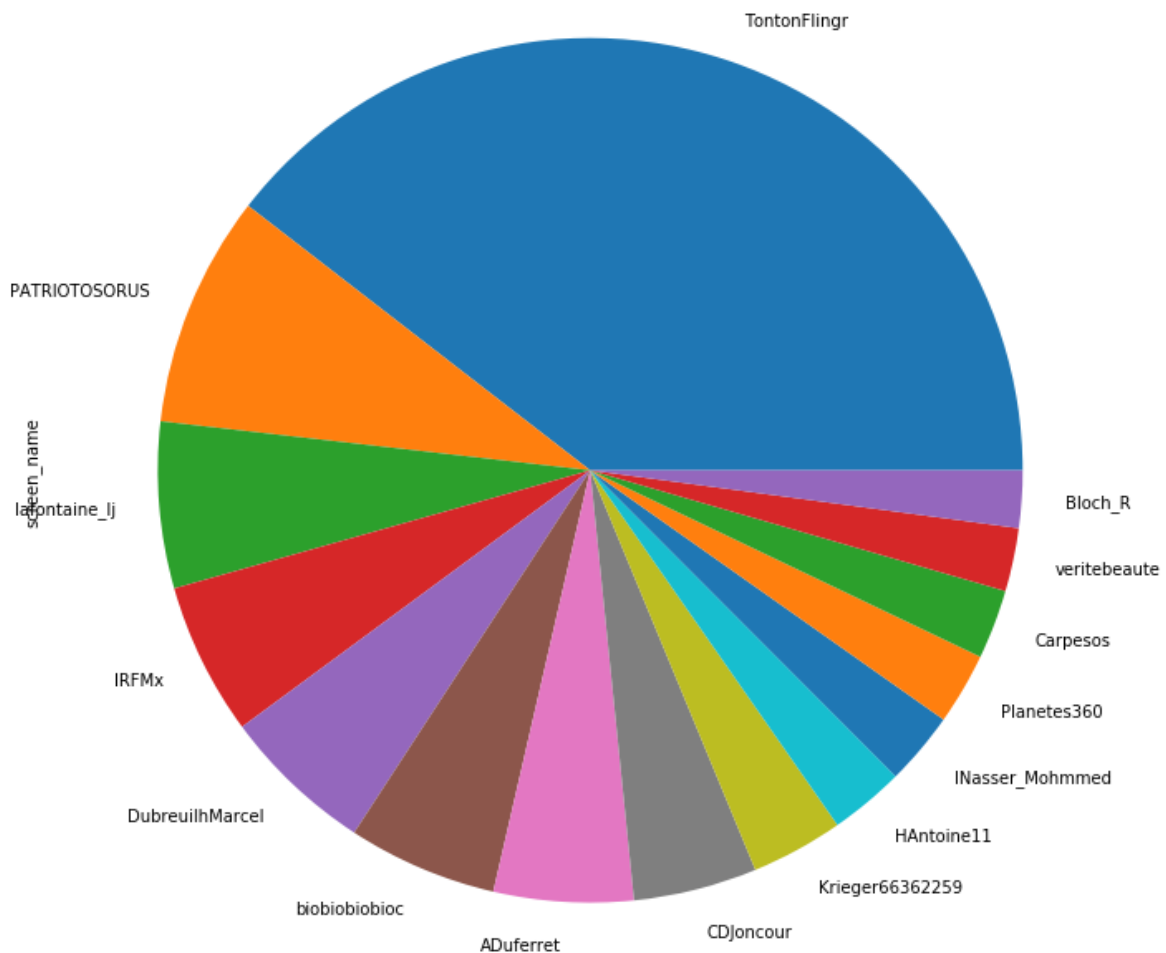
Le premier élément que nous pouvons remarquer est que la proportion de tweets mentionnant à la fois l'hydroxychloroquine et le coronavirus est largement supérieure aux deux autres. En effet, la figure 10 montre que parmi les plus de 20 000 tweets que nous avons aspiré, 14 000 (74%) se rapportent à la théorie complotiste sur l'hydroxychloroquine, 5 000 (22%) à celle de Bill Gates et 1 000 (4%) au sujet de la 5G. Il reste cependant à déterminer si les théories sur l'hydroxychloroquine sont effectivement majoritaires par rapport aux deux autres, ou si notre base contient plus de faux positifs que les autres (tweets non complotistes traitant les sujets du coronavirus et de l'hydroxychloroquine, à des fins purement scientifiques par exemple). Nous développerons ce point ultérieurement.

Les diagrammes en bandes des figures 6, 7, 8 et 9 permettent d'identifier des "pics" du nombre de tweets sur un sujet. Dans le cas de la 5G, un pic est observé le 12 novembre 2020 sur la figure 6, avec 111 tweets publiés. Cette date correspond au lendemain de la sortie du documentaire indépendant conspirationniste ***Hold-Up : Retour sur un chaos***. Ce documentaire assure montrer que le coronavirus a été mis au point dans le but de créer "*une dictature mondiale technologique basée sur le sans contact, l'Internet des objets et donc la 5G*", comme le montre cet article de *France Culture* (2020) [10]. Dès lors, ce documentaire nous semble une bonne cible pour réaliser une analyse textuelle des réactions et commentaires qu'il suscite. Nous explorerons plus en détail ce documentaire dans la partie Youtube de cette étude.

Ensuite, le diagramme sur la théorie complotiste autour de Bill Gates (figure 8) montre également un pic de 1293 tweets le 15 novembre 2020. Ce pic semble être dû aux retweets d'un message de "@TontonFlingr" :

*"Le milliardaire Elon Musk refuse de se faire vacciner contre le coronavirus et traite Bill Gates de "crétin" ", @TontonFlingr.*

Ce tweet est en réalité le titre d'un article du site *PLANETES360* (2020) [11]. Il est accompagné du lien vers ce dernier. Il est difficile de dire pourquoi ce tweet a suscité autant de réactions. Cependant, le compte twitter *@TontonFlingr* se trouve être le compte le plus retweeté de toute notre base de données au sujet de la théorie autour de Bill Gates. En effet, 31% des retweets de notre base de données proviennent du compte *@TontonFlingr*, comme le montre la figure 11 ci-dessous.



**Figure 11 :** Répartition des 15 utilisateurs les plus retweetés parmi les tweets autour de la théorie du complot au sujet de Bill GATES.

Le compte *@TontonFlingr* a par la suite été supprimé. Il était détenu par un individu nommé Pascal Laborie, suivi par plus de 4000 personnes avant sa suppression. Le fait qu'un individu détient un compte aussi influent montre que les théories du complot peuvent rapidement se propager à l'aide des réseaux sociaux tels que Twitter du fait de la grande visibilité auprès d'un large public. Ce public semble cependant, dans le cas des théories du complot, n'être composé que de complotistes. Ces théories complotistes sont donc diffusées entre les individus d'un même groupe, à l'intérieur d'une communauté et avec un éventuel leader. Les opinions de ce dernier sont directement acceptées et propagées massivement par ceux qui le suivent, sans remise en question. De plus, aucun compte officiel de médias n'est présent parmi les comptes les plus retweetés. Seuls des comptes d'utilisateurs sont présents, ce qui confirme bien cette idée de **communauté du complot**.

Finalement, le diagramme en bandes de la figure 7 montre que la théorie complotiste autour de l'hydroxychloroquine est globalement plus active et plus discutée par rapport à nos deux autres théories. Le nombre de tweets journaliers au sujet de cette théorie est plus important par rapport aux deux autres théories, comme le montre la figure 9. Un pic du nombre de tweets est présent le 26



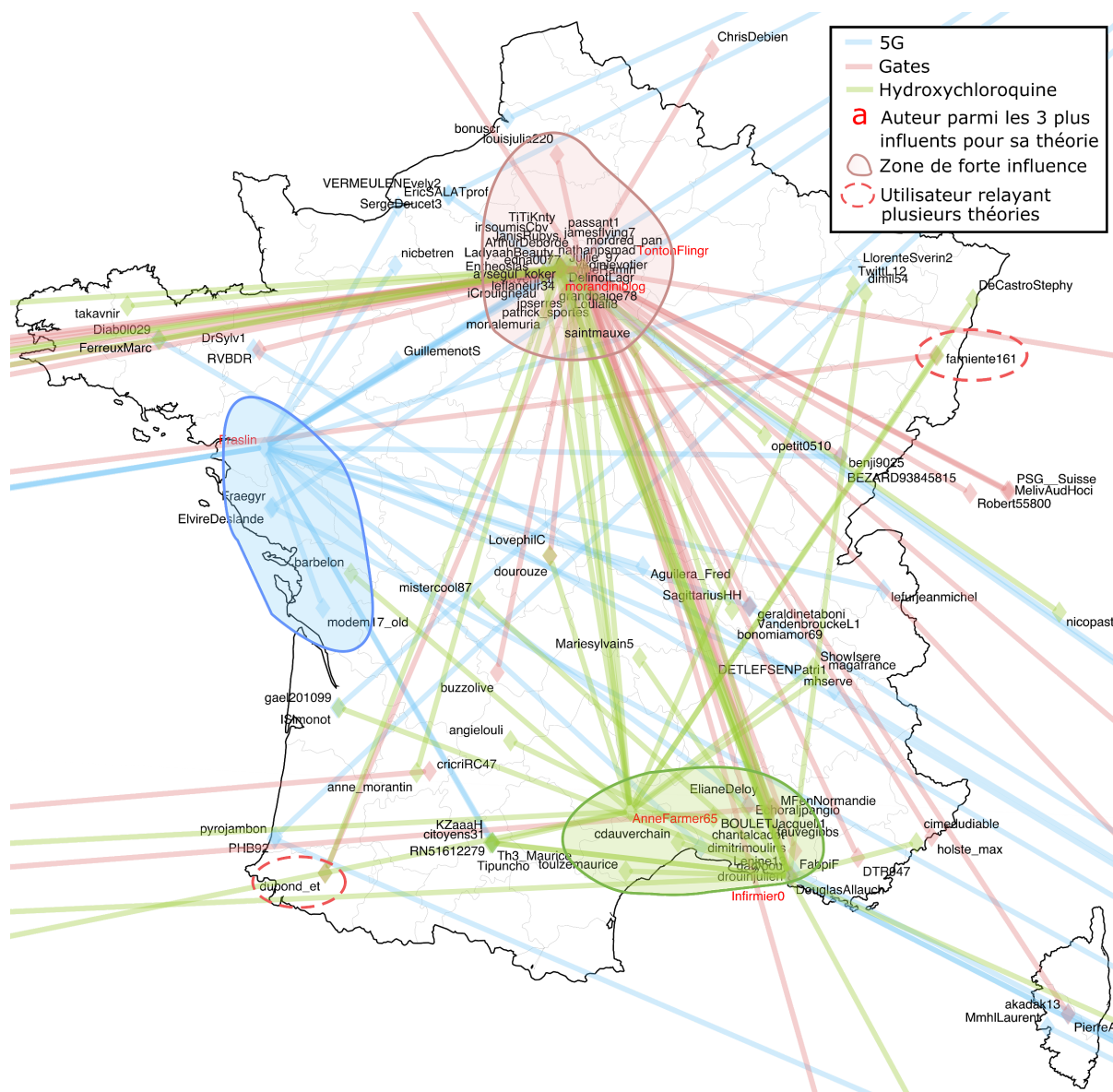
octobre 2020. Ce jour correspond à la réaction de Didier Raoult, professeur de microbiologie, sur Twitter après que l'ANSM (Agence nationale de sécurité du médicament et des produits de santé) ait décidé de stopper l'utilisation extensive de l'hydroxychloroquine dans le cadre des tests (ANSM, 2020) [12]. Didier Raoult semble lui-même très peu retweeté par les complotistes. Dans notre aspiration de données, seuls 7 utilisateurs l'ont retweeté en mentionnant les mots "coronavirus" et "hydroxychloroquine", entre le 6 octobre et le 6 décembre 2020. La théorie de l'hydroxychloroquine possède elle aussi un utilisateur plutôt influent, dont le pseudo est *@biobiobiobio*. Il possède actuellement plus de 15 000 abonnés.

Enfin, la base de tweets sur cette dernière théorie comporte quelques comptes twitters de médias officiels. Néanmoins ces comptes ne font pas autant réagir les internautes par rapport aux autres comptes twitters. En effet, aucun des 13 comptes les plus retweetés pour l'hydroxychloroquine n'est le compte d'un média officiel. Cela montre la facilité que les théories du complot ont à se propager par rapport aux informations transmises par des médias officiels.

Pour conclure, ces premières analyses révèlent que les théories du complot sur Twitter se propagent majoritairement entre utilisateurs complotistes, dans les groupes assez fermés telles des communautés, et non pas par l'intermédiaire de commentaires sous les posts de certains grands médias par exemple. Ces utilisateurs profitent de certains événements de l'actualité liés à la théorie en laquelle ils croient, mais déforment largement la réalité pour que celle-ci corresponde à leur vision. Ces fausses informations ont alors tendances à se propager beaucoup plus facilement, surtout dans ce contexte incertain actuel.

#### ***II.1.d) Pour aller plus loin : la cartographie du réseau complotiste***

Les bases de données élaborées jusque-là nous ont permis d'identifier quelques acteurs importants dans la propagation des théories du complot comme *@TontonFlingr* pour la théorie sur Bill Gates, et *@biobiobiobio* pour celle au sujet de l'hydroxychloroquine. Chaque tweet comporte des informations sur son auteur, mais également sur l'auteur qu'il retweete le cas échéant. Ainsi, en extrayant les tweets qui sont effectivement des retweets, et pour lesquels Twitter nous fournit un indice de localisation, nous pouvons afficher le réseau complotiste sous forme de graphe sur une carte, présentée dans la figure 12 ci-dessous.



**Figure 12 :** Réseau des différentes théories du complot en France.

Afin d'améliorer la lisibilité tout en transmettant un maximum d'informations, nous avons d'abord décidé de ne garder que les arcs reliés à l'un des trois auteurs les plus influents pour chaque théorie. Nous n'avons également gardé que 50 arcs par théorie (en respectant bien entendu les proportions, c'est à dire que si les auteurs A, B et C sont les plus importants pour la 5G, et qu'ils ont dans été retweeté 500, 2000 et 2500 fois, nous afficherons donc respectivement 5, 20 et 25 arcs pour chacun d'entre eux).

Les 3 auteurs les plus influents pour chaque théorie (ceux dont nous avons affiché les liens) peuvent être résumés par le tableau suivant :

Théorie	Auteur	Nombre d'arcs	Localisation
5G	Fraslin (16 600 abonnés)	21	Nantes, France
	RaoulDidierOff (1 736 abonnés)	18	Tel Aviv, Israël
	BobLeCentriste (16 300 abonnés)	9	Koszalin, Pologne
Gates	TontonFlingr (4 137 abonnés)	35	Paris, France
	PATRIOTOSORUS (8 522 abonnés)	7	Bryce, Utah, USA
	lafontaine_lj (16 abonnés)	5	Québec, Canada
Hydroxychloroquine	Infirmier0 (16 900 abonnés)	19	Marseille, France
	AnneFarmer65 (15 200 abonnés)	15	Le Caylar, France
	morandiniblog (662 400 abonnés)	12	Paris, France

**Tableau 1 :** Localisation géographique, nombres d'arcs et auteurs des retweets de la figure 12 par théorie du complot autour de la Covid-19.

## Interprétation

Cette carte soulève plusieurs points intéressants. Chaque trait de couleur représente le fait qu'un utilisateur en a retweeté un autre (en indiquant les termes "coronavirus" et "5G"/"Gates"/"Hydroxychloroquine" dans son retweet). Si le trait est bleu, il relie deux utilisateurs supposés vecteurs de théories autour de la 5G ; le trait rouge relie deux utilisateurs relayant des théories sur Bill et Melinda Gates ; et le trait vert relie des utilisateurs propageant des théories sur l'hydroxychloroquine. Au bout de chaque arc est inscrit le pseudo Twitter de l'utilisateur concerné, et ce texte est coloré en rouge s'il s'agit d'un des utilisateurs les plus influents (un des plus retweetés) pour sa théorie. Enfin, nous avons ajouté des zones colorées autour des lieux où se situent le plus d'utilisateurs influents pour chacune des théories.

En analysant cette carte, la première chose que nous pouvons remarquer, ce sont les tendances. Deux des trois utilisateurs les plus retweetés pour l'hydroxychloroquine (@AnneFarmer65 et @Infirmier0) sont situés au sud-est, non loin de Marseille, tandis que @TontonFlingr et d'autres comptes importants

pour les théories sur Bill Gates, sont plutôt situés à Paris. Ces zones font particulièrement sens : Didier Raoult étant basé à l'IHU de Marseille, il paraît normal de constater que les individus habitant non loin de cette zone se sentent plus concernés par la théorie autour de l'hydroxychloroquine. Nous remarquons aussi que de nombreux utilisateurs sont situés aux États-Unis pour les théories autour de Bill Gates (plusieurs arcs rouges sortent de la France à l'Ouest, en direction des USA).

Enfin, la zone d'influence à l'ouest fait également sens pour la théorie autour de la 5G. Nantes est l'une des seules villes qui devrait être couverte par le réseau 5G dans cette zone majoritairement rurale. La 5G est un sujet assez controversé dans cette ville, comme le montre cet article de *metropole.nantes.fr* [13] rédigé le 8 décembre 2020.

Enfin, nous remarquons que malgré le peu d'arcs affichés sur cette carte (seulement 50 pour chaque théorie, soit environ 0.8% des arcs des bases de données initiales), nous pouvons tout de même observer deux utilisateurs, entourés par un cercle un rouge pointillé, qui ont retweeté et donc relayé plusieurs théories différentes : @farniente161 et @dupond\_et. Si ce phénomène s'observe déjà avec ce nombre extrêmement réduit de nœuds et d'arcs, il ne fait aucun doute qu'il est d'autant plus visible à plus grande échelle. Si un compte twitter partage une théorie complotiste, il doit certainement partager d'autres théories du complot sur des sujets différents. Cette observation contribue une nouvelle fois à notre hypothèse qu'il existe une "communauté du complot", un ensemble d'utilisateurs dans lequel les théories du complot sont particulièrement populaires et relayées.

Pour conclure sur cette première partie, l'analyse révèle que les principaux relais des théories du complot autour de la Covid-19 ne sont pas des médias traditionnels en ligne, mais un certain nombre d'utilisateurs particulièrement suivis, qui instrumentalisent certains faits et propagent de fausses informations. De plus, ce partage de fausses informations semble se faire majoritairement et uniquement au sein de la communauté du complot. Les complotistes partagent peu d'informations en dehors de la sphère complotiste. Nous avons également appris que la capacité d'une théorie à se propager dans une zone géographique est d'autant plus grande que les individus de cette zone se sentent concernés, inquiets ou mal informés sur le sujet. Cela peut expliquer le phénomène de **zone géographique** pour un sujet précis d'une théorie du complot. Ce phénomène est par exemple présent sur la figure 12, avec une zone verte représentant une tendance des théories du complot au sujet de l'hydroxychloroquine proche de Marseille, ville où est situé Didier Raoult, un acteur majeur du débat autour du rôle de ce médicament dans la lutte contre le coronavirus.

Cependant, il nous reste encore à expliquer les mécanismes qui permettent à ces théories de se propager autant. Pour cela, nous réalisons une analyse textuelle sur Youtube.

## II.2) Youtube

Sur Youtube, les chaînes et vidéos parlant des théories complotistes du coronavirus sont principalement celles de médias traditionnels tels que *France culture*, *Le Parisien* ou encore *Le Monde*. Ces chaînes et vidéos dénoncent ces théories et n'y adhèrent pas. Au vu de leur position contre les théories du complot, nous pouvons penser que l'aspiration des données à partir de ces médias pourrait entraîner un biais de sélection dans nos données. Nous pouvons donc supposer que les commentaires sous de telles vidéos seront principalement **contre** les théories du complot. Cependant, les partisans des théories du complot sont souvent actifs dans les commentaires de ces vidéos, même s'ils ne sont que minoritaires. Pour palier à ce possible biais, nous allons également aspirer des données à partir de vidéos ou de chaînes youtube adhérant aux théories. Les chaînes *WhataBeautifulWorld Youtube Channel* et *Riposte laïque* totalisent de nombreuses vidéos promouvant les théories autour de la Covid-19, notamment sur le port du masque et l'origine du virus. Il est important de souligner que ces chaînes sont difficilement accessibles à partir de la barre de recherche youtube. Elles ne possèdent pas un bon référencement, ce qui explique pourquoi elles ne sont pas mises en avant par la plateforme. Cette difficulté d'accès à ce type de contenu sur Youtube peut être dû à un phénomène de modération de la part de la plateforme. Une étude sur ce sujet serait particulièrement intéressante, notamment dans le cadre de la lutte des réseaux sociaux contre les Fake-news et les théories du complot. Notre projet ne traitera cependant pas de cet aspect.

### II.2.a) Méthode de travail

Nous avons analysé les commentaires sous trois vidéos. Nous avons deux vidéos complotistes et une vidéo non-complotiste :

- **“Hold-up”** [14] : reportage mis gratuitement sur la chaîne *WhataBeautifulWorld Youtube Channel*<sup>1</sup>. Ce documentaire français sorti le 11 novembre 2020 a été produit par Pierre Barnérias, Nicolas Réoutsky et Christophe Cossé. Il a eu un grand retentissement dans les sphères complotistes car il remet en cause les décisions du gouvernement (port du masque, utilité du confinement, hydroxychloroquine). Il évoque également la théorie d'un complot mondial à l'origine de la Covid-19 [10]. La vidéo totalise 524 commentaires et la chaîne *WhataBeautifulWorld Youtube Channel* comptabilise 2 860 abonnés.
- **“Covid-19 on est sauvés le vaccin-viagra arrive”** [15] : vidéo totalisant 10 198 vues et 364 commentaires. Elle est postée sur la chaîne youtube *Riposte Laïque*, qui compte 42 400 abonnés environ. *Riposte Laïque* est avant tout un site web d'extrême droite créé en 2007, dont le créateur, Pierre Cassen, est arrivé sur Youtube dès 2007. Sur sa chaîne, il partage ses idées complotistes et islamophobes.

---

<sup>1</sup> La vidéo a été supprimée avant la fin de notre projet.

- **“Quand des complotistes doutent de l'existence du Covid-19”** [16] : vidéo comptant 31 848 vues et 416 commentaires, issue de la chaîne youtube du média *FRANCE 24*, avec 3,08 millions d'abonnés. Ce court reportage dénonce les théories du complot autour de la Covid-19, en présentant notamment des théories affirmant que le virus est inoffensif, voire qu'il n'existe pas.

Les deux premières vidéos sont issues de chaînes youtube ouvertement complotistes. Nous posons donc l'hypothèse que les commentaires sont rédigés par des individus complotistes. A l'inverse, la dernière vidéo est postée par une chaîne d'informations. Nous émettons donc l'hypothèse que les commentaires seront écrits par des individus qui ne sont pas complotistes. Ces trois vidéos sont donc séparées en deux groupes bien distincts : vidéos complotistes et non complotistes. Ces hypothèses seront réfutées ou bien validées à l'aide de l'analyse textuelle de ces commentaires.

Nous utilisons *RStudio* pour réaliser une analyse textuelle à partir des commentaires des trois vidéos Youtube sélectionnées. Nous avons expérimenté la méthode d'analyse dans le cadre du TD n°8 réalisé le jeudi 19 novembre. *Rstudio* nous permet de traiter les données brutes récupérées par *Youtube Data Tools*. Une étape de traitement et de filtrage nous permet d'extraire les mots et expressions les plus récurrentes dans les commentaires, mais également de retirer de cette liste des mots parasites tels que “le”, “la” ou encore “dans”. En comparant les résultats obtenus avec les trois vidéos, l'analyse textuelle nous permet d'identifier des thématiques communes. De plus, nous pouvons également identifier si une vidéo mentionne un sujet ou des personnalités en particulier. C'est ainsi que des liens peuvent apparaître entre ces théories et un parti politique ou une personnalité publique par exemple.

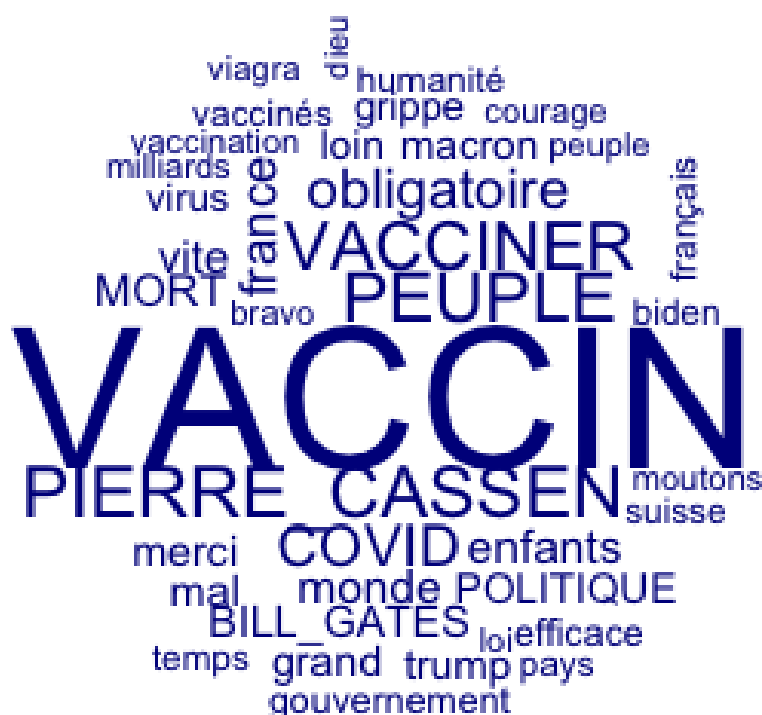
L'aspiration des commentaires se fait à l'aide de l'outil *Youtube Data Tools*, présenté en TD dans le cadre de l'enseignement IC05. Nous avons utilisé l'option *Video Info and Comments Module* pour obtenir des fichiers CSV regroupant les commentaires. Ensuite, à l'aide de l'application *Rstudio*, nous avons réalisé une analyse textuelle afin mettre en évidence les mots et expressions les plus utilisés dans l'espace commentaire. Les codes R utilisés sont présentés en annexe (annexe 2).

## II.2.b) Aspiration et résultats

Après un filtrage des mots les plus utilisés, nous obtenons les résultats suivants :



**Figure 13** : Analyse textuelle de la vidéo "hold-up".



**Figure 14** : Analyse textuelle de la vidéo "Covid-19: on est sauvé le vaccin viagra arrive".



**Figure 15 :** Analyse textuelle de la vidéo “Quand des complotistes doutent de l’existence du Covid-19”.

### **II.2.c) Analyse des résultats : Youtube**

Dans un premier temps, il est important de différencier les individus ayant commenté sous les trois vidéos. Nous supposons que les commentaires des vidéos “hold-up” (figure 13) et “Covid-19 : nous sommes sauvés le vaccin viagra arrive” (figure 14) ont été écrits par des partisans des théories du complot. Cette hypothèse se base sur le caractère très complotiste des chaînes youtube possédant ces vidéos. A l’inverse, nous considérons les commentaires sous la vidéo “Quand des complotistes doutent de l’existence du Covid-19” (figure 15), issus du média *FRANCE 24*, ont été écrits par des individus qui n’adhèrent pas à ces théories du complot. Nous avons donc analysé deux vidéos complotistes et une vidéo contre ces théories du complot.

De nombreux points communs apparaissent dans ces trois analyses textuelles. Tout d’abord, le champ lexical de la **peur** est présent dans ces trois figures. Les mots tels que “peur”, “mort”, “mal” ou encore “mourir” montrent bien la crainte de la population face à la menace que représente la Covid-19. L’ambiance anxiogène de la société actuelle apparaît clairement dans les commentaires des vidéos.

Le thème du **vaccin** est également présent dans les trois figures, avec les mots “vaccin” et “vacciner”. Ici, le vaccin est employé pour parler d’un thème précis des théories du complot autour du coronavirus. Cependant, il peut



également faire référence au complot autour des vaccins en général. Cela montre que la sphère complotiste autour de la Covid et les complots anti-vaccins sont liés.

De plus, le "**gouvernement**" [français] est un mot commun aux trois vidéos. Ce dernier est notamment critiqué pour sa gestion de la crise sanitaire, ou bien à cause des mesures sanitaires prises, impactant énormément l'économie du pays mais également la liberté individuelle.

Ensuite, des thèmes communs à seulement deux vidéos sur trois sont révélés à l'aide de notre analyse textuelle. Le documentaire "hold-up" est présent sur les figures 13 et 14. Ce documentaire de 2020 très controversé et critiqué a eu un impact retentissant dans la sphère complotiste. Il est cité de nombreuses fois.

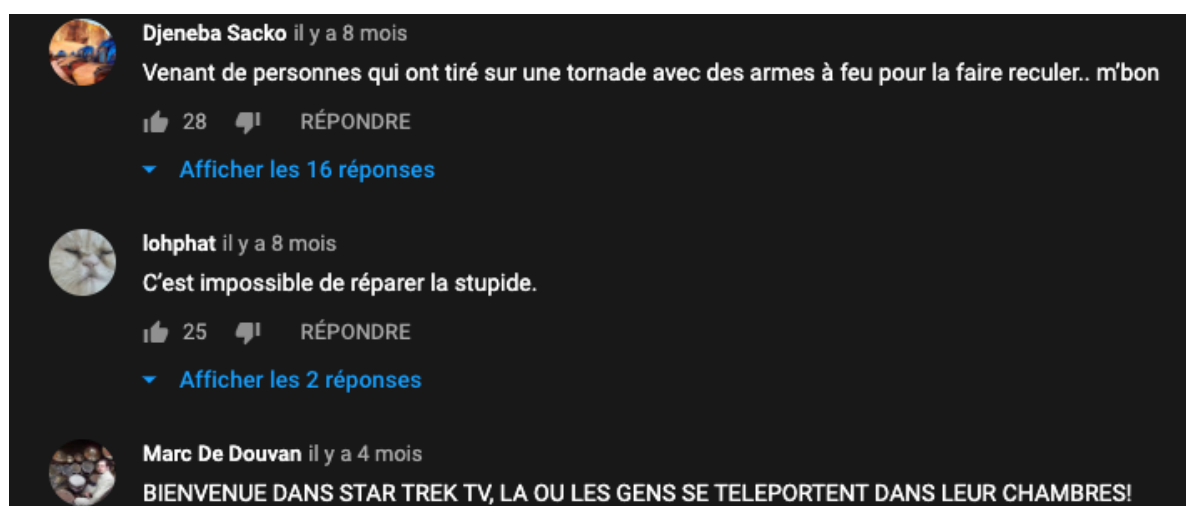
Les remerciements avec les mots "merci" et "bravo" des figures 13 et 14 semblent valider notre hypothèse que ce sont bien des complotistes qui ont commenté ces mêmes vidéos. Ils remercient la publication de ce documentaire. Un désir de "partage" et de connaître la "vérité" montre que les individus complotistes appellent à la diffusion du lien de la vidéo "hold-up" afin de propager la soit-disante vérité qu'apporterait ce documentaire.

La méfiance des individus vis à vis des informations communiquées par les médias est marquée par les termes "mensonges", "faux", et "vérité". La possibilité d'un "complot" est mentionnée dans les figures 13 et 15.

De plus, plusieurs acteurs semblent être identifiés à l'aide de cette analyse textuelle. Trois personnalités publiques sont mentionnées dans les figures 13 et 14 : "Macron", "Trump" et "Bill Gates". Le président français est critiqué pour ses choix et sa gestion de la crise. Cela est accentué par la critique générale faite envers le gouvernement français. Donald Trump, lui, semble être un acteur majeur dans la diffusion des théories du complot. Joe Biden est également mentionné dans la figure 14. Il est intéressant de souligner que la vidéo "Covid-19 : on est sauvé le vaccin viagra arrive" a été postée sur la plateforme avant la fin des élections présidentielles américaines. Nous savons que les complotistes sont souvent favorables à la vision du monde véhiculée par Donald Trump. Le mouvement "QAnon" (LeMonde, 2020) [17] témoigne par exemple de cette association complotiste et pro-Trump. Nous pouvons donc supposer que les commentaires mentionnant Trump acclament de dernier, tandis que son adversaire Joe Biden est critiqué, voire moqué. Bill Gates est également mentionné à cause de la théorie qui entoure son implication/rôle dans la pandémie du Covid-19.

Des thèmes récurrents et liés aux théories du complot autour de la covid-19 semblent donc se dégager de cette analyse. Le gouvernement est particulièrement pointé du doigt dans sa prise de décision et sa gestion de la crise de 2020. La présence de "masque" dans les figures 13 et 15 montre que leur utilisation et leur utilité sont très controversées et constitue un autre thème majeur au sein des complots autour du coronavirus.

Si l'analyse textuelle des commentaires Youtube nous montre des points communs entre ces vidéos, des différences apparaissent cependant. La vidéo de *France 24* est la seule pour laquelle nous faisons l'hypothèse que les commentaires ont été écrits par des individus qui ne sont pas partisans des théories du complot autour de la Covid-19. La théorie liée au virus semble, pour la figure 15, liée à de nombreuses autres théories du complot. Les termes "lune", "ISS", "banques" ou encore "[terre] plate" ne semblent pas faire partie du champ lexical autour de la Covid-19. De nombreuses théories complotistes sont ici évoquées. Elles sont très populaires, comme la théorie de la terre plate, ou encore celle affirmant que l'humanité n'a jamais posé le pied sur la lune. Si ces différentes théories sont présentes dans la figure 15, les commentateurs se moquent ouvertement d'elles. La figure 16 montre les top-commentaires, tournant au ridicule les complots.



**Figure 16 :** Extrait des commentaires de la vidéo "Quand des complotistes doutent de l'existence du Covid-19".

Les complots évoqués sont associés à des "conneries" comme le montre la figure 15. Ici, les individus sont majoritairement des personnes n'y adhérant pas. Cette critique des théories du complot semble correspondre au profil d'un abonné à la chaîne de *France 24* qui est, à la base, une chaîne télévisée d'informations. Cela accentue le fait que ce ne sont pas les mêmes individus qui consomment du contenu de la chaîne *Riposte Laïque* et celui d'une chaîne d'informations.

Contrairement à la vidéo de *France 24*, le champ lexical du documentaire "Hold-up" et de la vidéo de la chaîne *Riposte Laïque* laisse à penser que les commentateurs sont favorables aux théories du complot autour de la covid-19. Pour le documentaire "Hold-up", l'analyse textuelle est marquée par des remerciements, avec des mots tels que "merci" et "bravo". Les commentateurs semblent féliciter les réalisateurs du documentaire. Cela confirme notre hypothèse que ces commentaires sont rédigés par des partisans complotistes. Ces derniers remercient le documentaire de dire la "vérité" au sujet de la Covid-19.

La différence des commentaires entre ces vidéos peut notamment être expliquée par l'opinion politique des commentateurs, mais également leur statut social. Sous la vidéo de *France 24*, les personnes semblent s'indigner face à ces théories. Au contraire, sous les deux autres vidéos, les individus adhèrent et désirent partager les "vérités" promues par les théories du complot. Le public de ces vidéos est donc très différent.

Enfin, la troisième vidéo utilise un registre très provocateur, comme beaucoup de vidéos issues de la chaîne youtube *Riposte Laïque*. Le titre de la vidéo se veut choquant : le vaccin est associé à du "viagra". Les commentaires sont favorables aux théories du complot et critiquent le vaccin. Donald Trump, personnage public très actif sur les réseaux sociaux et sur le sujet du coronavirus, et Bill Gates, au cœur d'une théorie complotiste, sont évoqués. Les individus semblent soutenir Pierre Cassen, le détenteur de la chaîne Youtube. Le public de *Riposte Laïque* semble être complotiste et adhère aux idées de Pierre Cassen.

### **II.3) Analyse croisée Twitter/Youtube**

Les analyses de Twitter et Youtube peuvent être mises en parallèle. L'interprétation des résultats nous montre d'abord que les réseaux sociaux semblent être en effet un terrain propice à la diffusion des théories du complot. La liberté d'expression, la possibilité de l'anonymat explique ce phénomène. Cependant, ces théories semblent se diffuser majoritairement au sein même de la communauté complotiste. L'analyse des théories du complot sur Twitter montrent que les tweets complotistes ne touchent principalement qu'un public de complotistes initié à ces théories. Les théories du complot atteignent donc que très peu les autres utilisateurs des réseaux sociaux.

Ensuite, les complotistes profitent de sujets d'actualité ou de mesures gouvernementales plus ou moins polémiques pour propager leurs théories et interagir sur les réseaux sociaux. Les pics sur tweets correspondent à des dates précis où des événements ont provoqué énormément de réactions des internautes, comme avec la sortie du documentaire *Hold-up* par exemple. Les publications de vidéos Youtube du complot coïncident également avec des événements d'actualité.

De plus, nos analyses nous permettent de cerner plusieurs acteurs des théories du complot. Elles mettent notamment en évidence l'implication de la sphère d'extrême droite. Sur Youtube, les chaînes complotistes les plus populaires se revendiquent souvent d'extrême droite. C'est le cas de la chaîne *Riposte Laïque* dont nous avons parlé précédemment. Les vidéos de la chaîne parlent des sujets de prédilection de l'extrême droite : immigration, islam [18], nationalisme ou encore antiparlementarisme [19]. Une majorité des personnes qui approuvent ces vidéos complotistes semblent aussi adhérer aux idées d'extrême droite de *Riposte Laïque*.

D'une façon générale, nous pouvons affirmer que les personnalités politiques sont des acteurs de ces théories. C'est le cas de Donald Trump par exemple. Son nom se retrouve dans les analyses textuelles de chaque vidéo Youtube. Ce dernier a défendu à plusieurs reprises le traitement du Docteur Raoult à l'hydroxychloroquine [20], ce qui va dans le sens de la théorie complotiste sur ce médicament. De même, ce n'est pas un hasard si une personne aussi influente que Bill Gates soit au cœur d'une des théories. En effet, pour les personnes complotistes, les personnalités les plus riches et influentes de ce monde sont toutes liées de près ou de loin à des théories du complot ou des organisations secrètes [21].

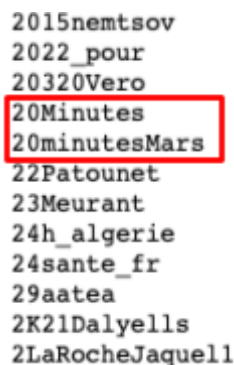
Enfin, nous comprenons que les réseaux sociaux sont à la fois acteurs et relais des théories du complot. Ce sont des relais car aujourd'hui, ce sont sur les réseaux sociaux que se propagent principalement et le plus efficacement les théories du complot. Mais les réseaux sont également des acteurs dans cette diffusion, car leurs algorithmes ont un rôle à jouer dans la mise en avant ou non des théories du complot. Nous avons vu par exemple que sur Youtube, il était bien plus facile de trouver à l'aide de la barre de recherche Youtube des vidéos de médias traditionnels critiquant les théories du complot que des vidéos complotistes.

## III) Analyse critique

### III.1) Biais de données

Dans notre projet, nous avons aspiré une quantité de données assez faible dans la partie Youtube. Nous devons émettre l'hypothèse que nos données présentent **un biais de sélection**. Introduit dans le cours n°6 de l'UV IC05, intitulé "Biais de données", le biais de sélection peut être défini comme étant une erreur systématique liée à la sélection des sujets étudiés. Cette observation peut être particulièrement présente dans la partie Youtube de notre projet. En effet, comme nous l'avons présenté, la majorité des vidéos sur la plateforme traitant des théories du complot sur la covid-19 proviennent de médias n'adhérant pas à ces théories. Bien au contraire, les vidéos critiquent et démontent ces théories en montrant leur limite ou en présentant les preuves permettant de les réfuter. Nous pouvons penser que les consommateurs de ces chaînes Youtube sont en accord avec le message véhiculé par ces vidéos, à savoir que ces théories sont fausses voire ridicules. Les commentaires peuvent donc ne pas être représentatifs de la population favorable à ses théories, mais au contraire être majoritairement constitué de personnes critiquant les théories du complot. Il est aussi difficile de trouver des vidéos sur Youtube de complotistes même en tapant des mots clés ou en faisant une recherche avancée. Nous pouvons supposer que les algorithmes de Youtube ont pour but de ne pas référencer correctement les vidéos de complotistes et de les faire tomber "dans l'oubli".

De plus, le biais de sélection peut aussi s'observer avec les données aspirées sur Twitter. En effet l'aspiration fonctionne par mot-clés, et renvoient l'ensemble des tweets correspondant à ces derniers. Notre base sur l'hydroxychloroquine par exemple, représente les tweets dont le texte comprend "coronavirus" et "hydroxychloroquine". Notre but est donc de cibler les tweets conspirationnistes en rapport avec l'hydroxychloroquine, mais naturellement, les complotistes ne sont pas les seuls à utiliser ces deux mots ensembles. En parcourant notre base, nous pouvons voir que les auteurs de certains tweets ne sont effectivement pas complotistes du tout, mais des médias "sérieux", qui se contentent de relater l'actualité :



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

**Figure 17** : Extrait des auteurs de la base hydroxychloroquine.

Ainsi, bien que ces médias traditionnels soient (en apparence) largement minoritaires, leur présence pourrait nuancer nos résultats. Le meilleur moyen de tous les écarter serait de parcourir toutes les bases à la main, mais cela impliquerait de parcourir 20 000 tweets un par un.

## III.2) Améliorations

La quantité de vidéos et de tweets aspirés pourrait être augmentée dans une nouvelle étude afin d'avoir des résultats plus représentatifs de l'ensemble de la population. Si l'aspiration de données sur Youtube ne dépend que du temps et des compétences techniques de l'équipe en charge du projet, une difficulté particulière est à souligner quant à l'utilisation de Twitter comme support. L'API du site nécessite une autorisation de Twitter pour que l'équipe de recherche puisse utiliser les données numériques de la plateforme. Sans cette autorisation, l'aspiration des données est quasiment impossible.

Ensuite, bien que notre équipe ait eu accès à l'API de Twitter, nous avons été confrontés à plusieurs limites techniques. En effet, le nombre de tweet qui peut être aspiré est prédéfini, et cette limite ne peut être dépassée.

Un second point important à souligner est la temporalité des tweets aspirés. L'API du site permet d'acquérir des tweets datant de moins d'un mois sans restrictions sur leur nombre. Si cela peut permettre une analyse temporelle, au fil de l'actualité, nous ne pouvons cependant pas analyser des données issues du début de la pandémie par exemple. Twitter permet des aspirations plus anciennes, mais les requêtes sont extrêmement limitées. Nous avons eu besoin de

8 comptes développeurs différents pour couvrir une période de seulement deux mois. Cette pandémie mondiale s'étend maintenant depuis plus d'un an, et les théories du complot ont commencé à émerger très tôt. Les informations fiables et vérifiées au sujet de la Covid-19 étaient rares, ce qui rendait la propagation de rumeurs et de fausses informations plus facile. Malgré tout, notre connaissance de ce virus et notre vision de la crise ont largement évolué au fil des mois. Il est donc possible que l'aspiration de tweets plus anciens (datant du premier confinement par exemple) mettent en exergue des résultats complètement différents, avec de nouveaux acteurs et de nouvelles communautés.

## IV) Conclusion

Les réseaux sociaux apparaissent dans un premier temps comme l'eldorado pour le partage massif d'informations, complotistes ou non. La liberté d'expression et la promesse de pouvoir toucher un maximum de personnes grâce à la visibilité des publications sur les réseaux semblent être les bons ingrédients nécessaires aux partages des théories du complot autour de la Covid-19. Le fonctionnement de Twitter est particulièrement favorable à cette large diffusion : un événement de l'actualité peut engendrer des réactions de la part des internautes qui seront partagées instantanément.

Notre étude a pu identifier des personnalités publiques telles qu'Emmanuel Macron, Bill Gates et Donald Trump au centre de ces théories complotistes autour de la Covid-19. Des acteurs majeurs dans la diffusion sont également identifiés sur Twitter. Un nombre restreint de comptes centralisent les réactions et les retweets sur le réseau social, au sujet du coronavirus.

De plus, les théories complotistes sur la covid-19 sont évoquées parallèlement à de nombreux autres sujets d'actualités, mais également à d'autres théories du complot. Le gouvernement et la gestion de la crise sanitaire sont souvent montrés du doigt, et la covid-19 met sur le devant de la scène le complot général autour des vaccins par exemple.

Enfin, un effet communautaire est observé sur Twitter. Les tweets de théories du complot sont diffusés dans un cercle d'individus précis. Un phénomène géographique est également observé. Les réactions sur un thème particulier auront tendance à être concentrées dans une zone géographique où des personnalités publiques relaient des informations concernant les théories complotistes. Cela est observé avec les réactions au sujet de l'hydroxychloroquine autour de Marseille, ville où réside Didier Raoult, acteur majeur pour ce thème précis.

Cependant, une lutte contre les Fake news et, par la même occasion, contre les théories du complot, est visible sur ces plateformes [9]. Sur Youtube, nous savons qu'une vidéo non-monétisée aura un référencement inférieur par rapport à un contenu monétisé. Cette variation de référencement peut favoriser la mise en avant des vidéos, mais également inverser ce processus. Cela semble par exemple

être le cas pour les vidéos de la chaîne *Riposte Laïque*. Elle compte 42 500 abonnés et totalise près de 30 millions de vues. Elle pourrait donc théoriquement avoir accès à la monétisation, mais ce n'est pas le cas. La monétisation d'une chaîne s'obtient après vérification à la conformité des règles de la plateforme par des évaluateurs de Youtube [22]. Le contenu, le sujet traité et le langage sont notamment analysés. Les sujets des vidéos de Pierre Cassen ne semblent pas lui donner accès à cette monétisation. Son référencement est donc mauvais. L'algorithme de Youtube ne met pas ces vidéos en avant, et cela participe à la non diffusion des théories puisque seule les abonnés de la chaîne sont prévenus de la mise en ligne d'une nouvelle vidéo de sa chaîne.

Enfin, de nouvelles méthodes de lutte contre les Fake news sont apparues sur Internet. Récemment mise en place, la labellisation des Fakes news sur Twitter [23] est un exemple de ces mesures. Une étude sur la lutte des réseaux sociaux contre la diffusion de fausses informations, comme les théories du complot, semble donc être nécessaire.

## IV) Références

[1] « Après les attentats du 13 novembre, le fléau du complotisme », LeFigaro.fr, publié le 01 janvier 2015. Disponible sur : <https://www.lefigaro.fr/actualite-france/2015/12/01/01016-20151201ARTFIG00046-apres-les-attentats-du-13-novembre-le-fleau-du-complotisme.php>

[2] « Comment le coronavirus est devenu un terrain fertile pour les théories du complot ». LeMonde.fr, publié le 28 mars 2020. Disponible sur : [https://www.lemonde.fr/politique/article/2020/03/28/le-coronavirus-terrain-fertile-pour-le-conspirationnisme\\_6034764\\_823448.html](https://www.lemonde.fr/politique/article/2020/03/28/le-coronavirus-terrain-fertile-pour-le-conspirationnisme_6034764_823448.html)

[3] « A Wuhan, l'arbre du complot cache une forêt de preuves », Valeursactuelles.fr, publié le 21 avril 2020. Disponible sur : <https://www.valeursactuelles.com/societe/wuhan-larbre-du-complot-cache-une-for-et-de-preuves-118394>

[4] « Study : Breitbart-Led Right-Wing Media Ecosystem Altered Broader Media Agenda », Columbia Journalism Review, publié le 3 mars 2017. Disponible sur : <https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>

[5] « Fake News et Santé : Des Internautas Pas Si Crédules », publié le 16 juin 2019. Disponible sur : <https://theconversation.com/fake-news-et-sante-des-internautes-pas-si-credules-118786>

[6] « Taguieff décortique les théories du complot », LExpress.fr, publié le 12 mai 2018. Disponible sur : [https://www.lexpress.fr/actualite/societe/taguieff-les-complots-repondent-a-un-be-soin-d-ordre\\_2007876.html](https://www.lexpress.fr/actualite/societe/taguieff-les-complots-repondent-a-un-be-soin-d-ordre_2007876.html)

[7] « BigPharma, puces 5G... Les cinq théories complotistes sur le Covid-19 les plus répandues », LAVDN.fr, publié le 6 octobre 2020. Disponible sur : <https://www.lavoixdunord.fr/875498/article/2020-10-06/bigpharma-puces-5g-les-cinq-theories-complotistes-sur-le-covid-19-les-plus>

[8] : « Covid-19 : pourquoi Bill Gates est devenu la cible de tant de théories du complot », lejdd.fr, publié le 4 décembre 2020. Disponible sur : <https://www.lejdd.fr/Societe/covid-19-pourquoi-bill-gates-est-devenu-la-cible-de-tant-de-theories-du-complot-4010118>

[9] : « Coronavirus : comment le professeur Didier Raoult est devenu une figure centrale des théories complotistes », LeMonde.fr, publié le 28 mars 2020. Disponible sur : [https://www.lemonde.fr/pixels/article/2020/03/28/comment-didier-raoult-est-devenu-la-nouvelle-egerie-des-complotistes\\_6034761\\_4408996.html](https://www.lemonde.fr/pixels/article/2020/03/28/comment-didier-raoult-est-devenu-la-nouvelle-egerie-des-complotistes_6034761_4408996.html)

[10] : « En quoi "Hold-up" est-il un documentaire complotiste ? » FranceCulture.fr, publié le 18 novembre 2020. Disponible sur : <https://www.franceculture.fr/emissions/les-idees-claires-le-podcast/en-quoi-hold-up-st-il-un-documentaire-complotiste>

[11] : « Le milliardaire Elon Musk refuse de se faire vacciner contre le coronavirus et traite Bill Gates de « crétin » », PLANETES360.fr, publié le 1 octobre 2020. Disponible sur :



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

[12] : « L'ANSM publie sa décision sur la demande d'une RTU pour l'hydroxychloroquine dans la prise en charge de la maladie Covid-19 - Point d'Information », [ansm.sante.fr](https://ansm.sante.fr/S-informer/Communiqués-Communiqués-Points-presse/L-A-NSM-publie-sa-decision-sur-la-demande-d-une-RTU-pour-l-hydroxychloroquine-dans-la-prise-en-charge-de-la-maladie-Covid-19-Point-d-Information), publié le 23 octobre 2020. Disponible sur : <https://ansm.sante.fr/S-informer/Communiqués-Communiqués-Points-presse/L-A-NSM-publie-sa-decision-sur-la-demande-d-une-RTU-pour-l-hydroxychloroquine-dans-la-prise-en-charge-de-la-maladie-Covid-19-Point-d-Information>.

[13] : « 5G : la Ville de Nantes ouvre le débat. », [metropole.nantes.fr](https://metropole.nantes.fr), publié le 8 décembre 2020. Disponible sur : <https://metropole.nantes.fr/actualites/2020/dialogue-citoyen/debat-ville-5g>.

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

[15] : « Quand des complotistes doutent de l'existence du Covid-19 » - YouTube. Publié le 22 avril 2020 par la chaîne FRANCE 24. Disponible sur : <https://www.youtube.com/watch?v=3cwG8wxmSFg>.

[16] : « Covid-19 : on est sauvés, le vaccin-viagra arrive ! » - YouTube. Publié le 11 novembre 2020 par la chaîne Riposte Laïque. Disponible sur : [https://www.youtube.com/watch?v=3YCrs8UO\\_Ao](https://www.youtube.com/watch?v=3YCrs8UO_Ao).

[17] : « QAnon : aux racines de la théorie conspirationniste qui contamine l'Amérique », [LeMonde.fr](https://www.lemonde.fr), publié le 14 octobre 2020. Disponible sur : [https://www.lemonde.fr/pixels/article/2020/10/14/qanon-aux-racines-de-la-theorie-conspirationniste-qui-contamine-l-amerique\\_6055921\\_4408996.html](https://www.lemonde.fr/pixels/article/2020/10/14/qanon-aux-racines-de-la-theorie-conspirationniste-qui-contamine-l-amerique_6055921_4408996.html).

[18] : Professeur décapité : Macron complice des islamistes ! - YouTube. Publié le 17 octobre 2020 sur la chaîne Riposte Laïque. Disponible sur : <https://www.youtube.com/watch?v=m9CsBH9r080>.

[19] : « 2021 : Année de la dictature ou du sursaut français ? » - YouTube. Publié le 31 décembre 2020 sur la chaîne Riposte Laïque. Disponible sur : <https://www.youtube.com/watch?v=ilH62gUxVIY>.

[20] : « Coronavirus. Donald Trump reste convaincu des bienfaits de l'hydroxychloroquine », [Ouest-France.fr](https://www.ouest-france.fr), publié le 9 juillet 2020. Disponible sur : <https://www.ouest-france.fr/monde/etats-unis/donald-trump/coronavirus-donald-trump-reste-convaincu-des-bienfaits-de-l-hydroxychloroquine-6901126>.

[21] : « Les Illuminati, le complot qui explique tous les autres ». [Le Temps](https://www.letemps.ch), publié le 10 juillet 2020. Disponible sur : <https://www.letemps.ch/societe/illuminati-complot-explique-autres>.

[22] Règles de monétisation des chaînes YouTube - Aide YouTube. Disponible sur : <https://support.google.com/youtube/answer/1311392?hl=fr>.

[23] « Twitter va labelliser les « fake news » liées au coronavirus ». [Les Echos](https://www.lesechos.fr), 12 mai 2020, <https://www.lesechos.fr/tech-medias/hightech/twitter-va-labelliser-les-fake-news-liees-au-coronavirus-1202251>.

---

<sup>2</sup> La vidéo "Hold-up" a été supprimée de Youtube avant la fin de ce travail de recherche.

# ANNEXE

## Annexe 1 : Extrait d'un fichier JSON.

```
{
  "created_at": "Sat Nov 07 08:18:54 +0000 2020",
  "id": 1324989691024859100,
  "id_str": "1324989691024859137",
  "text": "Pourquoi Bill Gates a réuni ses copains en octobre 2019 pour simuler
une pandémie de coronavirus dans le monde entier ?",
  "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,
    "favourites_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/e2b0ad5d080085a9fdbcddebcef90b8_normal.jpeg",
    "profile_image_url_https":
"https://pbs.twimg.com/profile_images/378800000179923159/e2b0ad5d080085a9fdbcddebcef90b8_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  
    }  
  ]  
}
```

## Annexe 2 : Codes R pour l'analyse textuelle Youtube

### 1. Code R de la vidéo "Covid-19 on est sauvés le vaccin-viagra arrive"

```
##### ANALYSE TEXTUELLE

####Installer et charger les librairies de travail

library(tidyr)
library(stringr)
library(dplyr)
library(quanteda)
library(ggplot2)
install.packages("quanteda")
install.packages("ggplot2")
install.packages("dplyr")

##### 1- EXPLORATION DE LA BASE DES COMMENTAIRES

### Charger la base des commentaires

base<-read.csv2("videoinfo_3YCrS8U0_Ao&ab_channel=RiposteLaiïque_2020_11_
15-11_43_58_comments.tab.csv")

##### 2- CREATION ET EXPLORATION DU CORPUS

### Creer un objet corpus
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)

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

```

##### 2.B Nettoyage et statistiques

###Enlever les mots de liaisons, conjonctions ou préposition en
enrichissant
##la liste des stopwords déjà téléchargeable dans quanteda
toremove<-c(stopwords("french"),"a", "comme", "d'un", "d'une", "aussi",
"fait",
        "être", "c'est", "an", "faire", "dire", "si", "qu'il",
        "où", "tout", "plus", "encore", "déjà", "depuis",
        "ans", "entre", "n'est", "peut", "dont", "donc",
        "ainsi", "faut", "va", "donc", "tous", "alors",
        "chez", "fois", "quand", "également", "n'a", "n'y",
        "celui", "celle", "l'un", "n'ont",
        "l'a", "l'on", "qu'on", "or", "d'ici", "s'il", "là", "dès",
        "dit", "pu", "six", "pu", "font", "ceux", "peut",
        "j'ai", "ni", "très", "l'une", "lors", "puis", "etc", "tel",
        "chaque", "ca", "veut", "toute", "qu'elle"
        , "peu", "moin", "après", "bien", "deux", "troi", "oui",
        "avant", "ça", "s'est", "notamment", "tant", "peuvent",
        "selon", "quelque", "toujour", "avoir", "car", "beaucoup",
        "sou", "non", "d'autre", "contre", "plusieurs",
        "autre", "toute", "fin", "heure",
        "lundi", "mardi", "mercredi", "jeudi", "vendredi", "samedi",
"dimanche",
        "dans", "pas", "me", "nos", "nous", "de", "vous", "sans",
"mais", "d'accord",
        "voir", "parce", "dis", "dit", "vont", "rien", "qu'ils",
"quoi", "juste",
        "pourquoi", "trop", "peux", "moins", "depuis", "sous",
"t'es", "ah", "vois",
        "vais", "vraiment", "y'a", "vas", "bla", "e", "d'être",
"veux", "mois",

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

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

"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\U0001f3fb♀","\U0001f413","\U0001f471\U0001f
3fb♀","\U0001f609",

"\U0001f914","\U0001f603","\U0001f644","\U0001f471\U0001f3fb♀","toujours","pre
mier",

"venir","part","arrive","bientôt","jadot","alain","bonne","certains","surtout"
,

"tellement","super","bonjour","aller","toutes","chose","mr","mdr","aucun","jam
ais"
)

##Créer un dictionnaire des expressions les plus usités à partir de
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"),
  vaccin=c("vaccin" ,"vaccins") ,
  vacciner=c("vacciner" ,"vacciné","vaccinées", "vaccinée"),
  Pierre_CASSEN=c("pierre","cassen"),
  peuple=c("personne","personnes","gens","individus","peuples"),
  mort=c("mort","morts"),
  Bill_GATES=c("Bill","gates"),
  complot=c("complot","complots"),
  politique=c("politique","politiques"))
)

## Recréer les objets matrices et tokens en enlevant cette fois les mots
inintéressants
## Tracer le nuage de mots

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)  
#varier le min_count pour avoir plus ou moins de mots sur le graphique
```



## 2. Code R de la vidéo “Quand des complotistes doutent de l'existence du Covid-19”

```
##### ANALYSE TEXTUELLE

####Installer et charger les librairies de travail

library(tidyr)
library(stringr)
library(dplyr)
library(quanteda)
library(ggplot2)
install.packages("quanteda")
install.packages("ggplot2")
install.packages("dplyr")

##### 1- EXPLORATION DE LA BASE DES COMMENTAIRES

### Charger la base des commentaires

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

### Créer un objet corpus
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)

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

##### 2.B Nettoyage et statistiques
```

```

###En lever les mots de liaisons, conjonctions ou préposition en enrichissant

toremove<-c(stopwords("french"),"a", "comme", "d'un", "d'une", "aussi",
"fait",
      "être", "c'est", "an", "faire", "dire", "si", "qu'il",
      "où", "tout", "plus", "encore", "déjà", "depuis",
      "ans", "entre", "n'est", "peut", "dont", "donc",
      "ainsi", "faut", "va", "donc", "tous", "alors",
      "chez", "fois", "quand", "également", "n'a", "n'y",
      "celui", "celle", "l'un", "n'ont",
      "l'a", "l'on", "qu'on", "or", "d'ici", "s'il", "là", "dès",
      "dit", "pu", "six", "pu", "font", "ceux", "peut",
      "j'ai", "ni", "très", "l'une", "lors", "puis", "etc", "tel",
      "chaque", "ca", "veut", "toute", "qu'elle",
      , "peu", "moin", "après", "bien", "deux", "troi", "oui",
      "avant", "ça", "s'est", "notamment", "tant", "peuvent",
      "selon", "quelque", "toujour", "avoir", "car", "beaucoup",
      "sou", "non", "d'autre", "contre", "plusieurs",
      "autre", "toute", "fin", "heure",
      "lundi", "mardi", "mercredi", "jeudi", "vendredi", "samedi",
"dimanche",
      "dans", "pas", "me", "nos", "nous", "de", "vous", "sans", "mais",
"d'accord",
      "voir", "parce", "dis", "dit", "vont", "rien", "qu'ils", "quoi",
"juste",
      "pourquoi", "trop", "peux", "moins", "depuis", "sous", "t'es",
"ah", "vois",
      "vais", "vraiment", "y'a", "vas", "bla", "e", "d'être", "veux",
"mois",
"s'en", "<", ">", "br", "href", "=", "\U0001f44d", "quot", "#39", "\U0001f44f", "\U0001f
64f",
"\U0001f602", "\U0001f64f", "\U0001f44f", "c3", "\U0001f44f", "\U0001f64f", "\U0001f
602", "parle",
"de", "#39", "le", "la", "et", "les", "pour", "pas", "en", "est", "br", "bon", "b", "vu", "a
llez", "regardez",
"prendre", "pense", "veulent", "salut", "lol", "voilà", "@corentin", "@max", "merci",

```

```

"km", "surtout", "@christian", "@dobonne", "n'ai", "mieux", "the", "sabrina",

"gt", "doit", "choses", "autres", "jamais", "jusqu", "comment", "qu'en", "chose",

"choses", "point", "\U0001f601", "besoin", "nombre", "existe", "martin", "manuel",

"sai", "gros", "toujours", "reste", "forme", "moment", "t", "totalement", "passer", "r
egarder"
)

##Créer un dictionnaire des expressions les plus usités à partir de
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"),
  vaccin=c("vaccin", "vaccins", "vacciner", "vacciné*"),
  croire=c("crois", "croit", "croire"),
  peuple=c("peuples", "personnes", "personne", "gens"),
  hold_up=c("hold up", "reportage"),
  preuve=c("preuve", "preuves"),
  masque=c("masque", "masques"),
  confinement=c("confinement", "confiner"),
  vérité=c("sais", "vérité", "connais", "vrai"),
  théorie=c("théorie", "théories"),
  vidéo=c("vidéos", "vidéo"))
)

## Recréer les objets matrices et tokens en enlevant cette fois les mots
inintéressants
## Tracer le nuage de mots

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)
#varier le min_count pour avoir plus ou moins de mots sur le graphique

```



### 3. Code R de la vidéo “Hold-up”

```
##### ANALYSE TEXTUELLE

####Installer et charger les librairies de travail

library(tidyr)
library(stringr)
library(dplyr)
library(quanteda)
library(ggplot2)
install.packages("quanteda")
install.packages("ggplot2")
install.packages("dplyr")

##### 1- EXPLORATION DE LA BASE DES COMMENTAIRES

### Charger la base des commentaires

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

##### 2- CREATION ET EXPLORATION DU CORPUS

### Creer un objet corpus
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)
```

```

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

##### 2.B Nettoyage et statistiques

toremove<-c(stopwords("french"),"a", "comme", "d'un", "d'une", "aussi",
"fait",
        "être", "c'est", "an", "faire", "dire", "si", "qu'il",
        "où", "tout", "plus", "encore", "déjà", "depuis",
        "ans", "entre", "n'est", "peut", "dont", "donc",
        "ainsi", "faut", "va", "donc", "tous", "alors",
        "chez", "fois", "quand", "également", "n'a", "n'y",
        "celui", "celle", "l'un", "n'ont",
        "l'a", "l'on", "qu'on", "or", "d'ici", "s'il", "là", "dès",
        "dit", "pu", "six", "pu", "font", "ceux", "peut",
        "j'ai", "ni", "très", "l'une", "lors", "puis", "etc", "tel",
        "chaque", "ca", "veut", "toute", "qu'elle",
        , "peu", "moin", "après", "bien", "deux", "troi", "oui",
        "avant", "ça", "s'est", "notamment", "tant", "peuvent",
        "selon", "quelque", "toujour", "avoir", "car", "beaucoup",
        "sou", "non", "d'autre", "contre", "plusieurs",
        "autre", "toute", "fin", "heure",
        "lundi", "mardi", "mercredi", "jeudi", "vendredi", "samedi",
"dimanche",
        "dans", "pas", "me", "nos", "nous", "de", "vous", "sans", "mais",
"d'accord",
        "voir", "parce", "dis", "dit", "vont", "rien", "qu'ils", "quoi",
"juste",
        "pourquoi", "trop", "peux", "moins", "depuis", "sous", "t'es",
"ah", "vois",
        "vais", "vraiment", "y'a", "vas", "bla", "e", "d'être", "veux",
"mois",
"s'en", "<", ">", "br", "href", "=", "\U0001f44d", "quot", "#39", "\U0001f44f", "\U0001f
64f",
"\U0001f602", "\U0001f64f", "\U0001f44f", "c3", "\U0001f44f", "\U0001f64f", "\U0001f
602", "parle",
        "amp", "the", "cas", "donne", "jusqu", "veulent", "vu", "maintenant",
        "regarder", "sabrina", "gt", "b", "mettre", "absolument", "pendant",
"sais", "sauf", "comment",

```

```

"bonne", "enfin", "vite", "lorsque", "bientôt", "point", "mis", "♥", "doit", "bon", "allez", "toutes",

"certains", "toujours", "autres", "autre", "aucun", "q", "petites", "grand", "temps"
)

##Créer un dictionnaire des expressions les plus usités à partir de
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", "documentaire*",
"film*", "reportage*", "article*"),
  peuple = c("peuple*", "personne*", "humanité*", "gens", "peuples"),
  vérité = c("vrai", "vraie", "vérité"),
  vaccin=c("vaccin", "vaccins"),
  politique=c("politique*", "politicien*"),
  monde=c("monde", "mondiale", "mondiale", "planète"),
  humanité=c("humanité", "homme", "humain*"),
  complot=c("complot*", "complotisme"))
)

## Recréer les objets matrices et tokens en enlevant cette fois les mots
inintéressants
## Tracer le nuage de mots

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 trop grand
textplot_wordcloud(dfm, min_count=9)
#varier le min_count pour avoir plus ou moins de mots sur le graphique

```

## **Annexe 3 : Lien du GitHub**

L'ensemble des codes, des visualisations en haute définition, des tweets aspirés et des dataframes assemblées pour les aspirations sur Twitter sont disponibles sur GitHub :

[https://github.com/bapt380/IC05\\_Covid19](https://github.com/bapt380/IC05_Covid19)