**DATA SPRINT COVID19**

**Groupe de travail : mouvements sociaux**

**Table des matières**

[11/05/2020 : Réunion de lancement](#_119345qhojty)

[12/05/2020, 15h : Réunion équipe](#_bid1fikbixan)

[19/05/2020 : Doc de travail équipe](#_o2ffh49a2ghb)

[Questions-Réponses équipe/SM](#_onuewox96iru)

[Point fin de journée 19/5](#_79mdv1749edo)

[Remarques générales SM](#_qxda4uoe7q3s)

[25/05/2020 : Doc de travail équipe](#_v5yuesfbt22i)

[26/05/2020, 10h : Réunion équipe](#_5hnj119sdqb9)

[26/05/2020 : Doc de travail équipe](#_2x07z9tuq2n9)

[02/06/2020 : Doc de travail équipe](#_c9x57t43c4xa)

[08/06/2020 : Doc de travail équipe](#_v3aquqbsb8wq)

[08/06/2020, 16h : réunion équipe](#_4mlx9nkz68m4)

[09/06/2020 : Doc de travail équipe](#_qtnvwp3p5caz)

[15/06/2020 : Doc de travail équipe](#_enfhz3ykn69r)

[15/06/2020, 15h : réunion équipe](#_a1m1b586l1pv)

[16/06/2020 : Doc de travail équipe](#_vrwqkpkkvng3)

[Lien doc rapport:](#_los0hogguv40)

[22/06/2020 : Doc de travail équipe](#_f20y31dh432s)

[23/06/2020 : Doc de travail équipe](#_ks6ivqenqhjl)

[23/06/2020 : Point équipe, 15h30](#_gdeo2z2lga4b)

[29/06/2020 : Doc de travail équipe](#_9uxi5sva2rlc)

# 

# 11/05/2020 : Réunion de lancement

**Présents: Sylvain Mignot (SM), Annick Vignes (AV), Aïchatou Sow (AS), Raïssa Muadi (RS), Martina Catte (MC), Maya Anderson-González (MAG)**

Chaque étudiante s’est présentée ainsi que ses motivations.

* La composition du groupe a été décidé en amont.
* Les membres du groupe sont habitués à travailler ensemble.

Chaque encadrant s’est présenté. Des questions ont été posées aux étudiantes pour connaître leur attentes et leurs expériences en matière technique [à compléter].

* **AV et SM ont convenu de regarder/traiter/préparer les jeux de données avant de les transmettre aux étudiantes**
* **AV va envoyer les article sur le canal Slack dédié**
* **La prochaine réunion de groupe est prévue pour le mardi 12/5 à 15h.**

**A faire pour demain mardi 12/5:**

1. Identification de hashtags (pour identifier des groupes) sur Twitter (toutes les 4)
   1. Ouvrir un compte Twitter si besoin
   2. Suivre l'actualité en lien avec la problématique des mouvements sociaux
   3. En quoi les hashtags repérés ont-ils un rapport avec des mouvements sociaux?
2. Analysez la structure d'un tweet. (AS)
   1. Quelles sont les données que l’on peut récolter sur les tweet?

**A faire pour lundi 18/5:**

1. Deux articles à lire pour une définition des mouvements sociaux et leurs particularités sur le net (MC et MAG: article Granjon ; AS et RM: article Rieder):
   1. Qu’est-ce que c’est qu’un mouvement social?
   2. Comment les mouvements sociaux passe-t-ils de la sphère publique `å la sphère numérique?

**Remarque de contextualisation** : Les citoyens sont pris entre les injonctions du gouvernement et la réalité des contraintes matérielles (dans les transports en commun, notamment).

# 12/05/2020, 15h : Réunion équipe

**Présents:** tous

**Présentation des hashtags**

* Pour la recherche de hashtags, doit-on se limiter au contexte de la France?
  + Oui, on peut filtrer au niveau de la requête pour prendre des les tweets en français et géolocalisés en France

**Discussion sur la méthode envisagée pour la constitution d’une base de tweets**

* Le lien vers [Twint](https://github.com/twintproject/twint) se trouve aussi dans le canal Slack code-data
* Est-ce qu’il est intéressant de prendre #coronavirus qui est un terme très large?
  + Oui, car il s’agira de faire un réseau de cooccurrences de hashtags par la suite
* Autres paramètres à prendre en compte pour filtrer la requête:
  + limiter la requête aux comptes les plus influents (plus de 1000 followers)
  + filtrer par langue et par géo-localisation (France)

**Discussion autour des mouvements sociaux**

* Question des droits du travail, du télétravail
  + remarque AV: Ne pas confondre objets de revendications sociales et mouvements sociaux
* Question des mobilisations pré-coronavirus qui ont “migré” en ligne dû au confinement
  + mobilisations des gilets jaunes
  + mouvements d’extrême droite
  + défenseurs de LREM

**Point à retenir pour l’article de Granjon** : Cet article insiste sur l'idée d'une porosité, en tout cas d'un va-et-vient, entre espaces en ligne et espaces physiques.

**Point à retenir pour l’article de Rieder** : Cet article fait l’état d’un penchant évident vers la gauche au sein de la twittosphère française. Un point qu’il serait intéressant de creuser dans notre propre collection de tweets.

* **SM sera disponible le mardi 19/5 pour répondre aux questions des étudiantes**

# 19/05/2020 : Doc de travail équipe

* **La prochaine réunion de groupe est prévue pour le lundi 25/5, le matin [heure à décider].**

# 

1. **Installer Twint et/ou GetOldTweets3** (toutes les 4)

* si bugs: répertorier toutes les erreures pour pouvoir demander de l’aide (bug de RM sur Windows)
* Lien vers le GitHub de [GetOldTweets3](https://github.com/Jefferson-Henrique/GetOldTweets-python)
  + pour l’installation dans le terminal:
    - pip install GetOldTweets3
    - import GetOldTweets3
* Lien vers le GitHub de [Twint](https://github.com/twintproject/twint)
  + Tutoriels installation Twint:

<https://null-byte.wonderhowto.com/how-to/mine-twitter-for-targeted-information-with-twint-0193853/>

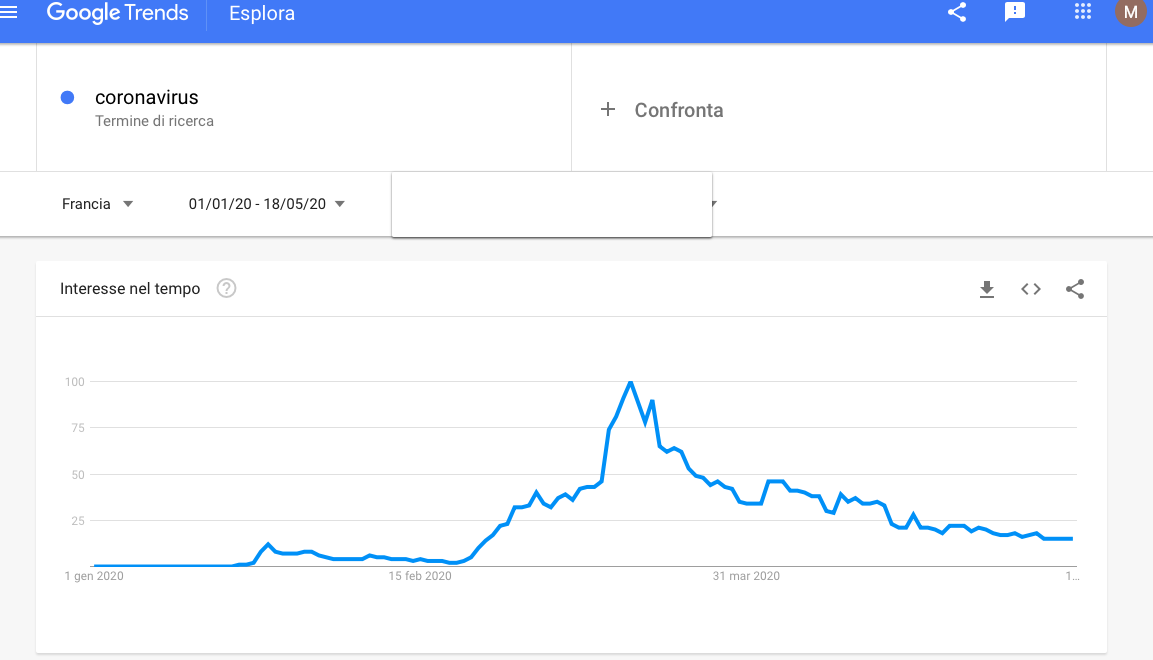
<https://medium.com/analytics-vidhya/how-to-scrape-tweets-from-twitter-with-python-twint-83b4c70c5536>

* + Bug script Twint dans Jupyter :

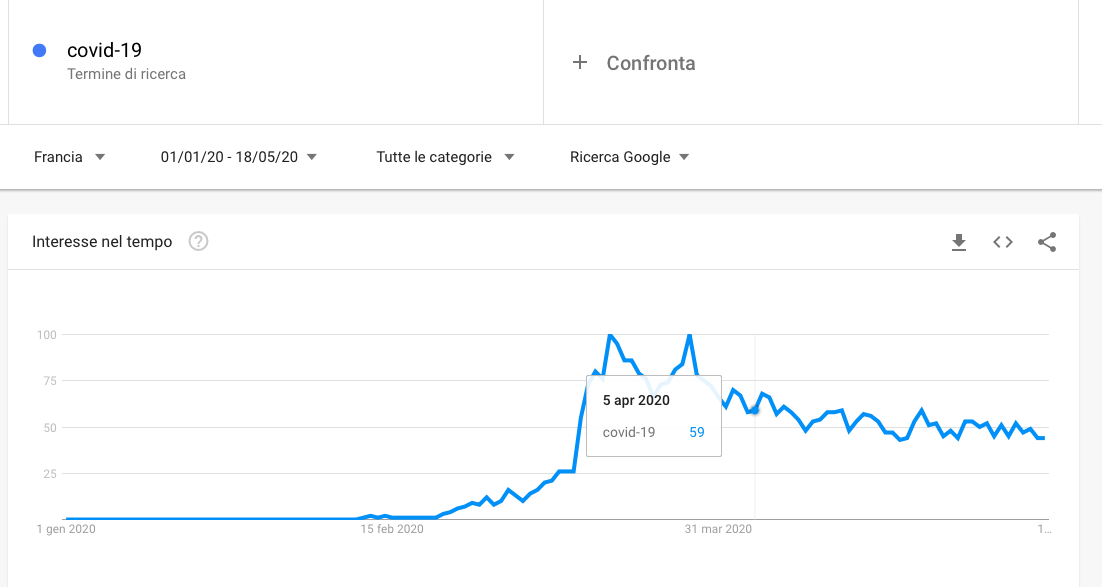
<https://markhneedham.com/blog/2019/05/10/jupyter-runtimeerror-this-event-loop-is-already-running/>

1. **Construire une requête**

* liste de mots clé à inclure à la requête:
  + covid19
  + coronavirus
  + covid-19
  + coronavirusenfrance
  + covid19fr
  + covid19france
* Contextualisation de la requête avec une recherche sur Google Trends:



<https://trends.google.com/trends/explore?date=2020-01-01%202020-05-18&geo=FR&q=coronavirus>



<https://trends.google.com/trends/explore?date=2020-01-01%202020-05-18&geo=FR&q=covid-19>

* paramètres à inclure à la requête Twint (Maya + Aïcha)
  + Timestamp: --since 2020-01-20 (20 janvier 2020 premier jour pour la France, selon google trends, voir captures ci-dessus)
  + Langue: -l fr
  + Geolocalisation pour le centre de la France: -g 46.769580,2.433050, 1000km
  + Geolocalisation pour Paris et ses alentours: 48.880048,2.385939,200km

Éventuellement:

* + pour avoir les tweets taggé avec un lieu: --location
  + pour avoir les tweets à proximité d’un lieu: --near
  + pour afficher le total des tweets à la fin: --count
  + pour sauvegarder les tweets dans un dataframe Pandas: --pandas-type
  + pour nettoyer automatiquement le dataframe Pandas: --pandas-clean
  + pour récolter les tweets les plus populaires: --popular-tweets
  + pour récolter les tweets d’utilisateurs influents: --verified ???

Exemple de méthodo: <https://fonio.medialab.sciences-po.fr/thinkdolphin/read/19f99d53-d6e7-4248-8ed5-5f4dba7b04f1>

1. **discuter d’une possible répartition des requêtes entres plusieurs ordinateurs (toutes les 4)**
2. **effectuer des test de requêtes (AS, MC, MAG)**

* **Requêtes qui ont marché:**

twint -g="48.880048,2.385939,500km" -o file1.csv --csv

twint -s coronavirus -l fr --limit 20 -o coronatest1.csv --csv

twint -s coronavirus -l fr --limit 20 -g="48.880048,2.385939,500km" -o file2.csv --csv

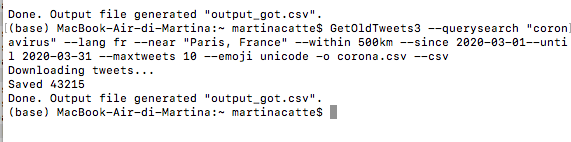
twint -s coronavirus -s covid19 -l fr --limit 20 -g="48.880048,2.385939,500km" -o file3.csv --csv

twint -s coronavirus -s covid19 -l fr --limit 20 -g="48.880048,2.385939,500km" --since "2020-01-20 00:00:00" -o file4.csv --csv

twint -s Coronavirus --verified -l fr -o covidfr.txt

(A la fin de la requête:175 743 lignes obtenues)

GetOldTweets3 --querysearch "coronavirus" --lang fr --near "Paris, France" --within 500km --since 2020-03-01--until 2020-03-31 --maxtweets 10 --emoji unicode -o corona.csv --csv



* **Requête qui n’ont pas marché:**

twint -s coronavirus -l fr --since 2020-01-20 -g="46.1314343,-2.434464" --limit 20 -o coronatest1.csv --csv

twint -s coronavirus --since 2019-12-31 -l fr -o covidfr.txt

twint -s coronavirus --since”2019-12-31” -l fr -o covidfr.txt

twint -s coronavirus -s covid19 -l fr --limit 20 -g="48.880048,2.385939,500km" --since "2020-01-20 00:00:00"--until "2020-01-31 23:59:59 -o file5.csv --csv

twint -s coronavirus -s covid19 -l fr -g="48.880048,2.385939,500km" --since "2020-01-20 00:00:00" --until 2020-02-01 --limit 20 -o file5.csv --csv

twint -s coronavirus -s covid19 -l fr -g="48.880048,2.385939,500km" --since "2020-01-20 00:00:00" --until 2020-02-01 -o file5.csv --csv

GetOldTweets3 --querysearch "coronavirus" --lang fr --near "Paris, France" --within 500km --since 2020-02-01--until 2020-03-01 --maxtweets 10 --emoji unicode -o coronafeb.csv --csv

>>> a pris des tweets de mars, avril et mai

twint -s coronavirus -s covid19 -s covid-19 -s coronavirusenfrance -s covid19fr -s covid19france -l fr --until "2020-03-01 00:00:00" -o file2.csv --csv

>>> n’a pris que le dernier mot clé

twint -s covid-19, -s coronavirusenfrance -l fr --until "2020-03-01 00:00:00" -o file4.csv --csv

>>> n’a pris que le dernier mot clé

## Questions-Réponses équipe/SM

**Installation et utilisation de Twint / GetOldTweets3**

* Y a-t-il un intérêt à utiliser Twint ou GetOldTweets dans un Jupyter Notebook versus en ligne de commande, sachant que nos capacités à coder sont faibles? Nous n’avons pas réussi à utiliser les scripts Python dans le Notebook.
* *je ne connaissais absolument pas ce logiciel donc je vais être moyennement utile :)Quel est l’avantage théorique d'exécuter un script python dans une appli web ? N’est t-il pas plus simple d’exectuer la requete twint via la console et d’analyser les resultats avec R ?*

**Pour bien filtrer les requêtes:**

* quelle démarche suivre pour la géolocalisation?
  + en utilisant le paramètr -g avec les coordonnées des grandes villes et leurs alentours?

*C’est la méthode la plus précise. Avec un inconvénient majeur. les coordonnées GPS ne sont indiqué que si l’utilisateur a donné son accord pour partager sa localisation. Cela concerne une minorité des utilisateurs. A n'utiliser que si la localisation précise est nécessaire en gardant conscience que vous n’obtiendrez qu’un échantillons de données.*

* + en utilisant le paramètre --location pour être certains de prendre des tweets géolocalisés?

*Vous l’avez sans doute déjà lu mais je le met la :*

*when conducting geo searches, the search API will first attempt to find Tweets which have lat/long within the queried geocode, and in case of not having success, it will attempt to find Tweets created by users whose profile location can be reverse geocoded into a lat/long within the queried geocode, meaning that is possible to receive Tweets which do not include lat/long information.*

*--l’option location filtre encore les résultats en ne gardant que celles avec coordonnées gps ? je vais vérifier, je n’ai pas cette option en mémoire.*

* + les deux?

*inutile je pense (mais je vais vérifier car je ne suis pas sur de l’action de --location)*

*A mon avis Le premier filtrage à effectuer est de ne prendre que les tweets en français. 99.9% des tweets écrits par des Français sont écrits en Français. je doute qu’il y ai beaucoup de touristes étrangers qui twittent en france en ce moment.*

*Ca serait votre corpus de grande taille.*

*Après vous pouvez faire la même chose en indiquant (en plus) country = france afin d’avoir un corpus de taille plus réduite. Cela permet de tester vos modèles statistiques plus rapidement.*

* des tweets contenants plusieurs mots-clé (coronavirus et covid19) vont-ils figurer plusieurs fois dans notre base de données?

*Non. une requête ne renvoie pas deux fois un tweet. (ils ont un ID unique, une requête n’affiche pas 2 tweets avec le même ID.)*

*Faire attention à bien réfléchir si vous voulez utiliser ET ou OU (enfin dans votre exemple c’est assez évident :)).*

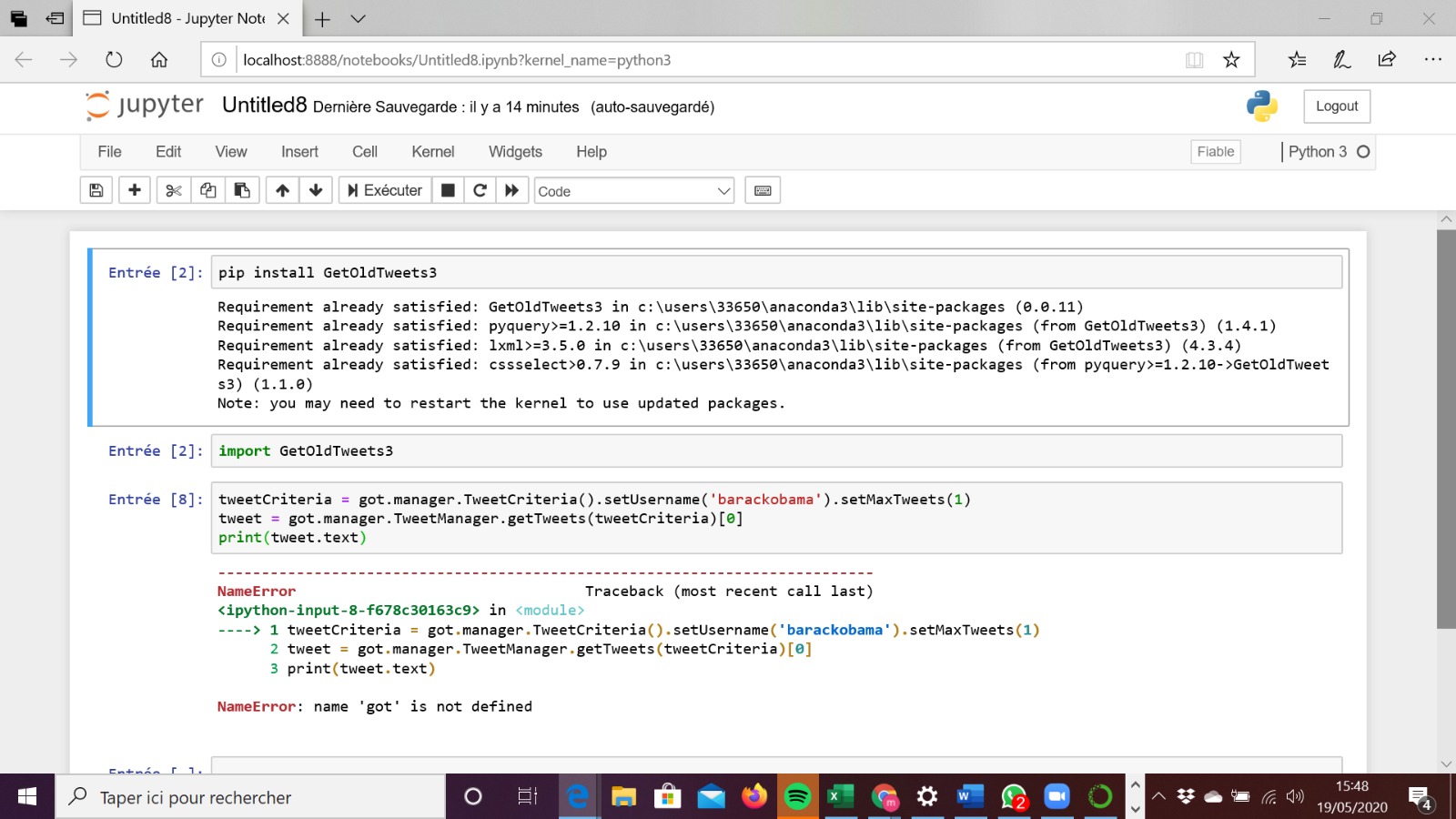
**Workflow**

* Y aurait-il un inconvénient à nous répartir les requêtes? Par exemple, chacune prend un mois différent?

*Non. soyez juste précises pour éviter les doublons.*

## Point fin de journée 19/5

* 3 d’entre nous ont réussi à installer Twint via le terminal et à générer des fichiers csv contenant des tweets.
  + **[bug 1] Avec GetOldTweets, MAG et MC obtiennent une erreur : “An error occured during an HTTP request: HTTP Error 503: Service Temporarily Unavailable” dans le terminal**
  + [bug 2] RM a des problèmes avec l’utilisation de GetOldTweets sur Jupyter Notebook (voir capture ci-dessous)



* Nous allons nous répartir les requêtes de la manière suivante:
* AS a fait le mois de janvier: [à compléter]
* MAG a fait le mois de février: 42700 observations
* MC a fait le mois de mars : 43215 observations
* RM a fait le mois d’avril: [à compléter]
* **[bug] la requête n’a pa pris en compte le paramétrage temporel + les stats obtenues avec pandas\_profiling pour les csv de février et de mars sont identiques**
* Première présentation exploratoire des données collectées par MC : <https://drive.google.com/file/d/101jnFN1jeLLLGR6m1PTkxDGs0x-PzlOx/view?usp=sharing>
* AV a mis sur le canal Slack un exemple de présentation d’analyse de données twitter destiné au grand public qui pourra servir de référence

## Remarques générales SM

*Je met en vrac a la fin. je vous laisse garder/ranger ce que vous voulez.*

*J’aurais du commencer par lire ca :D (le ppt)*

*Je vois que vous avez déjà bien bossé :)*

*C’est quoi votre “bug de dates” ?*

*c’est avec twint / getoldtweets ou les 2 ?*

*Restreindre votre requete en commencant a une date précise est inutile. ne mettez pas de dates pour être sûres de ne pas rater le début.*

*Si vous voulez élargir le dataset : prenez tout les #hashtags qui ont des co occurrences avec #coronavirus, puis refaites une requete pour extraire les tweets contenant ces hashtag. vous pouvez vous amusez a faire un arbre des hashtags avec leurs chronologie. certains apparaissent, d’autre disparaissent montrant le différents événement liés au coronavirus.*

# 25/05/2020 : Doc de travail équipe

1. **Suite aux remarques de SM**

**Chacune doit faire au moins 3 requêtes pour sa période temporelle:**

Rappel temporalités des requêtes avec la syntaxe twint:

--until ‘2020-02-01 00:00:00’ (Maya)

--since 2020-02-1 --until 2020-03-1 (Aïcha)

--since 2020-03-1 --until 2020-04-1 (Martina)

--since 2020-04-1 --until 2020-05-1 (Raïssa)

**Requête 1**:

**Maya:** twint -s coronavirus -l fr --until ‘2020-02-01 00:00:00’ -o file1.csv --csv

**142 435 tweets** (csv partagé sur le drive)

**Aïcha:** twint -s coronavirus -l fr --since 2020-02-1 --until 2020-03-1 -o file1.csv --csv

**311 202 tweets**

**Martina: Nombre de tweets trop volumineux - plusieurs requêtes :**

twint -s coronavirus -l fr --since 2020-03-1 --until 2020-04-1 -o file1.csv --csv

twint -s coronavirus -l fr --since 2020-03-1 --until 2020-03-26 -o fileM2.csv --csv

twint -s coronavirus -l fr --since 2020-03-1 --until 2020-03-17 -o fileM3.csv --csv

twint -s coronavirus -l fr --since 2020-03-1 --until 2020-03-15 -o fileM4.csv --csv

twint -s coronavirus -l fr --since 2020-03-1 --until 2020-03-10 -o fileM5.csv --csv

**Le total de lignes est de 1 861 012 et inclut tout le mois de mars**

**Raïssa:** twint -s coronavirus -l fr --since 2020-04-1 --until 2020-04-30 -o fileRaissa.csv --csv

twint -s coronavirus -l fr --since 2020-04-1 --until 2020-04-23 -o fileRaissa2.csv --csv

twint -s coronavirus -l fr --since 2020-04-1 --until 2020-04-17 -o fileRaissa3.csv --csv

twint -s coronavirus -l fr --since 2020-04-1 --until 2020-04-08 -o fileRaissa4.csv --csv

twint -s coronavirus -l fr --since 2020-04-1 --until 2020-04-02 -o fileRaissa5.csv --csv

**113 102 tweets**

**Requête 2**:

*> twint n’a pas de commande ‘country=France’, donc on prend les coordonnées gps de Paris et 500 km aux alentours*

*C’est pas propre ca :)*

*Donc pour trouver tous les tweets en france c’est légèrement plus compliqué.*

*Il va vous falloir un logiciel spécifique (un geographic information system).*

*Il y en a un paquet. ca peut aussi se faire sous R mais je n’ai jamais testé.*

*(* [*https://quanti.hypotheses.org/795*](https://quanti.hypotheses.org/795) *)*

*Personnellement j’utilise QGIS (mais à mon époque la cartographie sous R était balbutiante. et j’ai eu la flemme d’apprendre autre chose après.)*

*Dans ce logiciel vous chargez le fond de carte de la France, puis des points correspondants aux coordonnées GPS des tweets.*

*Ensuite vous “regardez” (enfin le logiciel regarde) quels points sont à l’intérieur du polygone “france”. ensuite soit vous leurs ajoutez un attribut “france” soit vous effacez les autres.*

> Option à tester : télécharger via <https://crisisnlp.qcri.org/covid19>

**Maya:** avant février

**Aïcha:** février

**Martina:** mars

**Raïssa:** avril

**Requête 3**:

*> on peut en faire autant qu’il y a de hashtags cooccurrents… peut-être s’arrêter aux 5/10 premiers? Inclure par exemple covid19, coronavirus, covid-19, coronavirusenfrance, covid19fr?*

**Maya:** twint -s hashtag\_cooccurrent -l fr --until ‘2020-02-01 00:00:00’ -o file3.csv --csv

**Aïcha:** twint -s hashtag\_cooccurrent -l fr --since 2020-02-1 --until 2020-03-1 -o file3.csv --csv

**Martina:** twint -s hashtag\_cooccurrent -l fr --since 2020-03-1 --until 2020-04-1 -o file3.csv --csv

**Raïssa:** twint -s hashtag\_cooccurrent -l fr --since 2020-04-1 --until 2020-05-1 -o file3.csv --csv

1. **Nouvelles questions**

* Souhaite-t-on filtrer les tweets issus de comptes influents dès la requête? Si oui, quel critère utiliser? Verified? Popular tweets?

La seule raison de filtrer au niveau de la requête c’est de limiter le volume de données. l’idée étant de scraper au maximum. mieux vaut trop que pas assez.

Donc la question est : quel volume de données pouvez vous récupérer ? avez vous besoin de filtrer les informations ?

* + Sachant que le statut de ‘verified’ est attribué par Twitter aux comptes jugés ‘d’utilité publique’, est-ce un bon critère à prendre?

**Réunion d’équipe reportée au 26/5/2020, 10h**

# 26/05/2020, 10h : Réunion équipe

**Présents :** RM, MC, AS, AV, SM, Bilel Benbouzid

**Absente** : Maya

**Présentation des premières explorations :** <https://drive.google.com/file/d/101jnFN1jeLLLGR6m1PTkxDGs0x-PzlOx/view?usp=sharing>

**Résumé des prochaines actions à effectuer:**

1. Créer le réseau de coocurrence des mots et hashtags les plus récurrents
   1. Procéder selon les indications données plus haut par SM: *prenez tout les #hashtags qui ont des co occurrences avec #coronavirus, puis refaites une requete pour extraire les tweets contenant ces hashtag. vous pouvez vous amusez a faire un arbre des hashtags avec leurs chronologie. certains apparaissent, d’autre disparaissent montrant le différents événement liés au coronavirus.*

*Remarque : a + b ca ne sert pas a grand chose je pense. lorsque vous allez extraire tous les tweets avec #coronavirus vous devriez retrouver tous les # cité en b. Faire un graph sur la totalité de la période puis un graph par mois. En ne prenant que les tweets avec un # visiblement vous ne gardez que 40 % des tweets. ca devrait simplifier l’extraction.*

*Si c’est encore trop ne gardez que les tweets importants, ou émis par des utilisateurs influents. un # qui n’est jamais utilisé en même temps que #coronavirus n’a sans doute aucun lien avec le coronavirus.*

*(si vous pouviez faire une recherche avec (#coronavirus, #covid19 et covid\_19 ca serait pas mal, ce sont les # “généraux” en france)*

* 1. Inclure les hashtags fournis par Bilel : #onnoublierapas, #guillotine2020, #Guillotine2020, #OnNoublierapas, #OnOublierapas, #IlsSavaient, #ilssavaient, #gouvernementcriminel #plaintecovid, #GouvernementCriminel, #PlainteCovid, Gilets jaunes.

(Annick) *A priori, si vous avez fait vos extractions correctement, vous devriez retrouver les # des listes envoyés par Bilel. C’est une bonne façon de vérifier la qualité de vos extractions.*

* + 1. **Remarque**: effectuer un test pour voir si les mots clé des requêtes twint sont sensibles à la casse (car il se peut que non).

1. Procéder à une analyse qualitative de chaque base de données:
   1. Lire systématiquement les 10 tweets les plus retweetés de chaque requête pour relever des éléments de contexte. *J’ai dit 10 un peu au hasard le mieux serait de tracer la distribution du nombre de RT et de lire ceux qui ont une probabilité non nulle d’être RT.*
   2. Identifier les mouvements sociaux les plus intéressants (variables hashtags + tweet)
   3. Identifier les tweets les plus importants (variable retweet)
   4. Identifier les acteurs les plus importants (variables username + mention) *et les qualifier (hommes, femmes, métier, mots les décrivant etc.).*
   5. Prendre note de ces éléments dans un tableau quali (?)

**Notes d’Aïcha:**

Comment la gauche se positionne? La droite? Les gilets jaunes? (idée générale: Identifier les mouvements sociaux)

Partir du jeux de données de Bilel, repartir des hashtags (proposition de Bilel)

Beau jeu de données avec requête (commentaire)

Idée créer notre propre corpus et à partir des Mots-clés=> prendre hashtags et les non hastaghs (les 2 à la fois)====Requête qui mélange mots clés et hashtags

identifier les mouvements sociaux en 1er

Aller voir les tweets les plus récurrents, les classer par nombre, voir ce qui domine, faire du qualitative

10 plus les plus retweetés, qui parle et de quoi il parle? (Vignes)

Réseaux de cooccurrences, faire la somme du poids du lien avec…. notebook r dispo avec S. Mignot au besoin

Construire la liste des hashtags qui ressortent le plus souvent, puis avoir la communauté (se référer au ppt sur slack posté par A. Vignes) Utiliser Cortext pour avoir communauté et identifier s’il y a des groupes qui forment des mouvements sociaux

\* hashtags coronavirus pour réseau de cooccurrence

\* tweet les plus retweetés

\* filtrer le plus possible

\* visualiser et voir les communautés

\*Quels sont les tweets les plus importants et aller dans les détails

\*comprendre les données

**Prochaine réunion d’équipe prévue pour le 8/6/2020, à 14h**

# 26/05/2020 : Doc de travail équipe

1. Terminer l’extraction de la requête 1, peut-être utile pour la suite (Martina + Raïssa?)
2. Créer une nouvelle requête 2.1 comprenant plusieurs mots clé (Maya avec Twint + Aïcha avec GetOldTweets)

Ci-dessous le code Python (merci Raïssa) pour lancer une requête twint depuis un Notebook Jupyter **en créant directement un pandas dataframe avec tous les tweets**:

import twint

!pip install nest\_asyncio

import nest\_asyncio

nest\_asyncio.apply()

c = twint.Config()

c.Search = "coronavirus"

c.Since = "2020-04-01"

c.Until = "2020-05-01"

c.Lang = "fr"

​

c.Pandas = True

​

twint.run.Search(c)

​

Tweets\_df = twint.storage.panda.Tweets\_df

1. Une fois la nouvelle requête stabilisée : lancer la collecte (en attente)

* **Problèmes pour les requêtes testées avec Twint et GetOldTweets :** 
  + **impossible d’effectuer une requête comprenant plusieurs mots-clé avec Twint en ligne de commande, donc obligées de passer par Jupyter Notebook**

# 02/06/2020 : Doc de travail équipe

1. **Organisation**

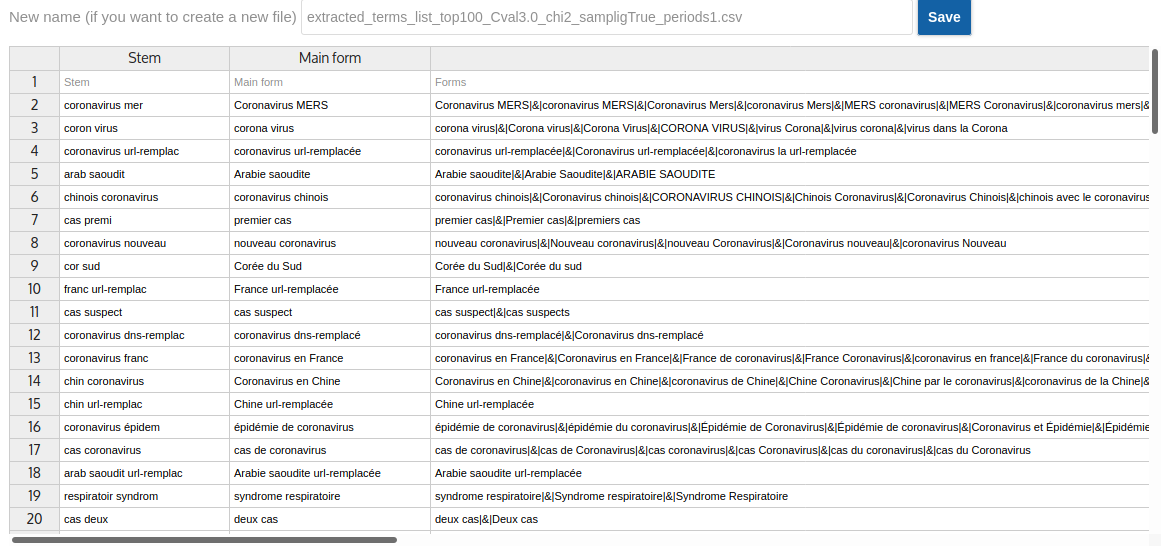
* création du dossier ‘CSV-MouvementsSociaux-CoronaSprint sur Drive
* partage avec les membres de l’équipe et les encadrants
* contient les 4 csv correspondant aux requêtes faites pour le mot-clé “coronavirus”
  + Requete1\_coronavirus\_avantfevrier.csv (Maya)
  + Requete1\_coronavirus\_fevrier.csv (Aïcha)
  + Requete1\_coronavirus\_mars.csv (Martina)
  + Requete1\_coronavirus\_avril.csv (Raïssa)
* lancement script Python (ci-dessus) dans Jupyter Notebook pour recollecter les données du mois d’avril

1. **Construction des réseaux de cooccurrences**

* création du projet Mouvements Sociaux Corona Sprint sur Cortext
* import csv avant février, février, mars
* parsing des fichiers csv + premières visualisations

**Extractions de termes:**

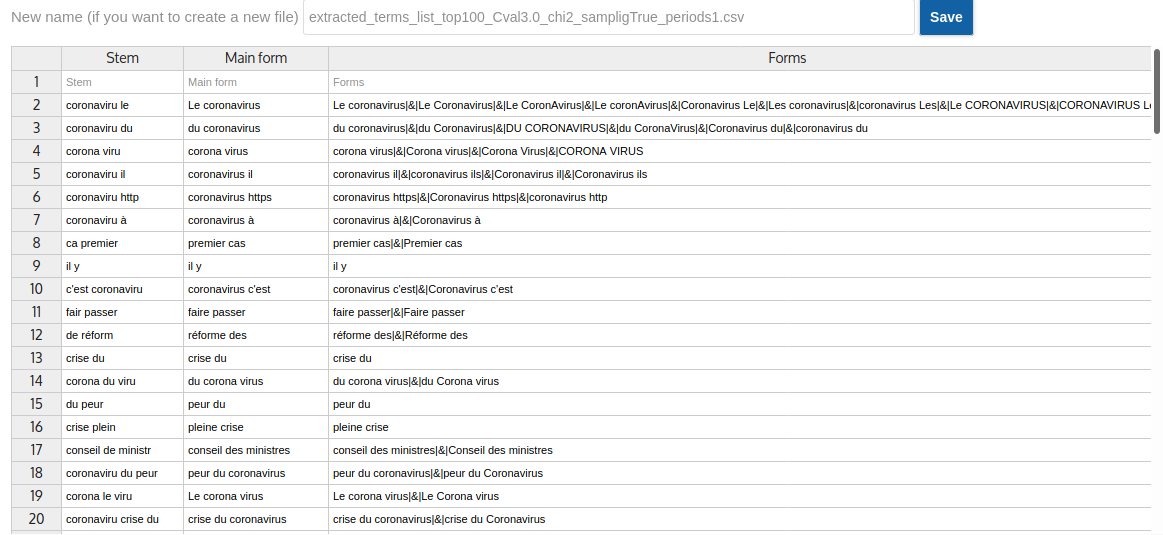
*Top 20 pour avant février*



Remarques:

* occurrences les plus parlantes : “coronavirus chinois”, “premier cas”, “nouveau coronavirus”, “cas suspect”, “syndrome respiratoire”
* noms propres à forte occurrence : “Arabie Saoudite”, “Corée du Sud”, “Chine”

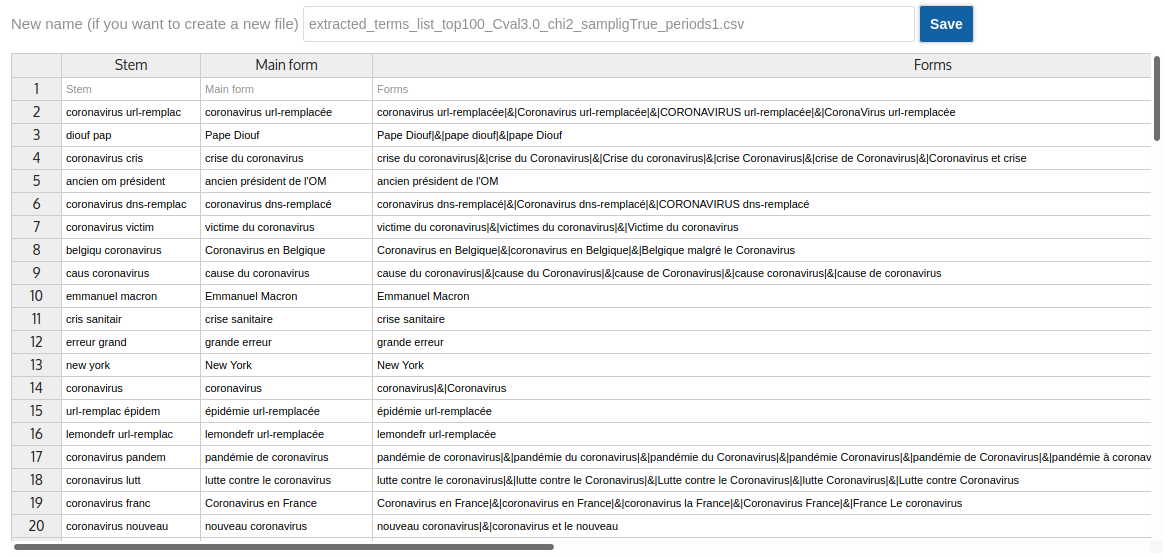
*Top 20 pour février*



Remarques:

* occurrences les plus parlantes : “premiers cas”, “réforme des”, “crise du coronavirus”, “conseil des ministres”, “peur du coronavirus”
* aucun nom propre à forte occurrence

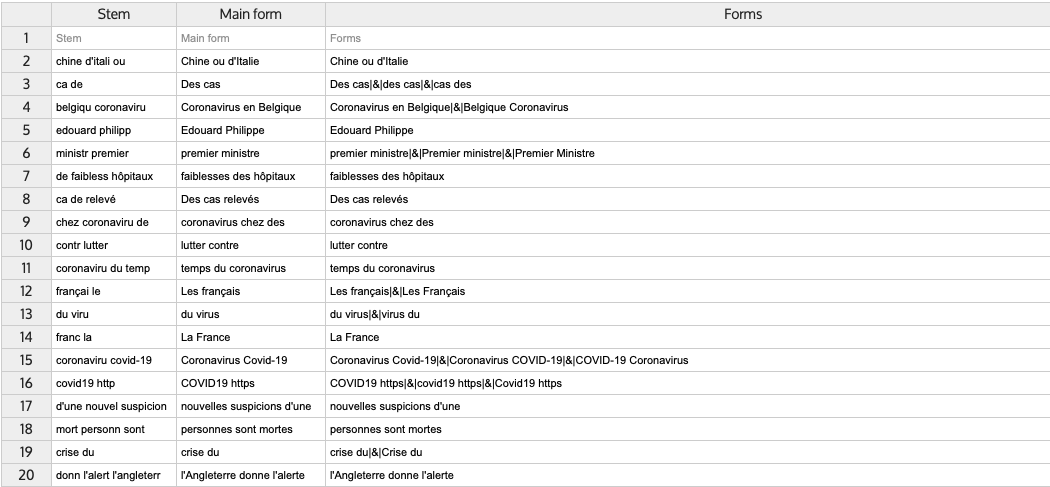
*Top 20 pour mars:*



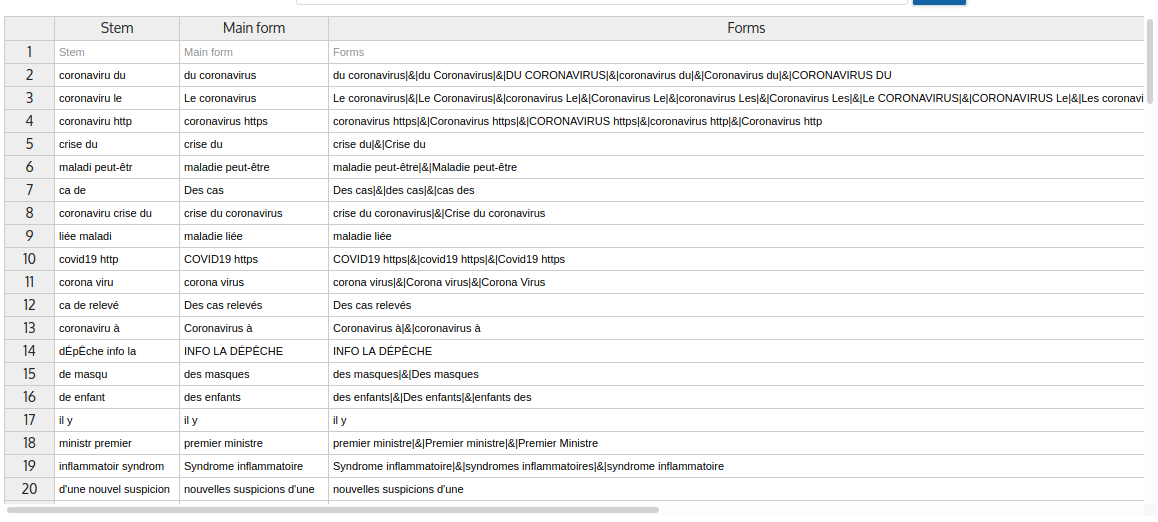
Remarques:

* occurrences les plus parlantes : “ancien président de l’OM”, “cause du coronavirus”, “grande erreur”, “crise sanitaire”
* noms propres à forte occurrence : “Pape Diouf”, “Emmanuel Macron”, “New York”

*Top 20 pour avril:*



*Top 20 pour avril (trié par rapport au nombre d'occurrences):*



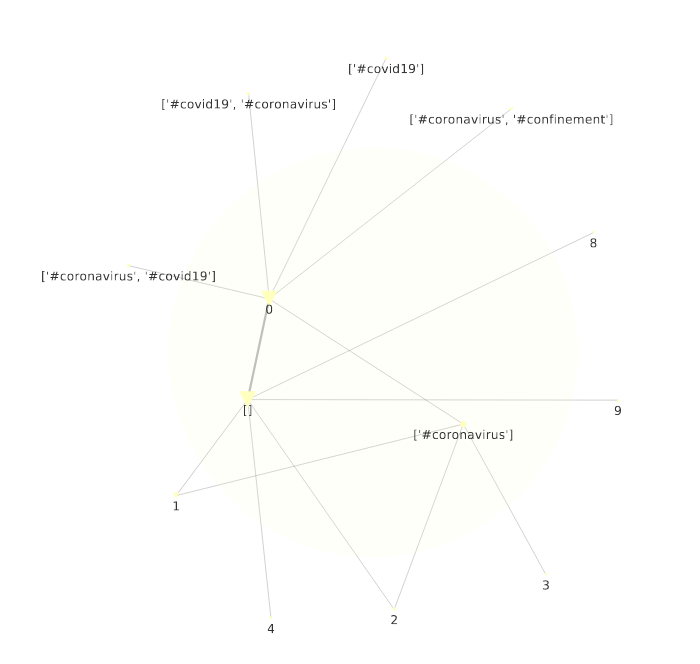
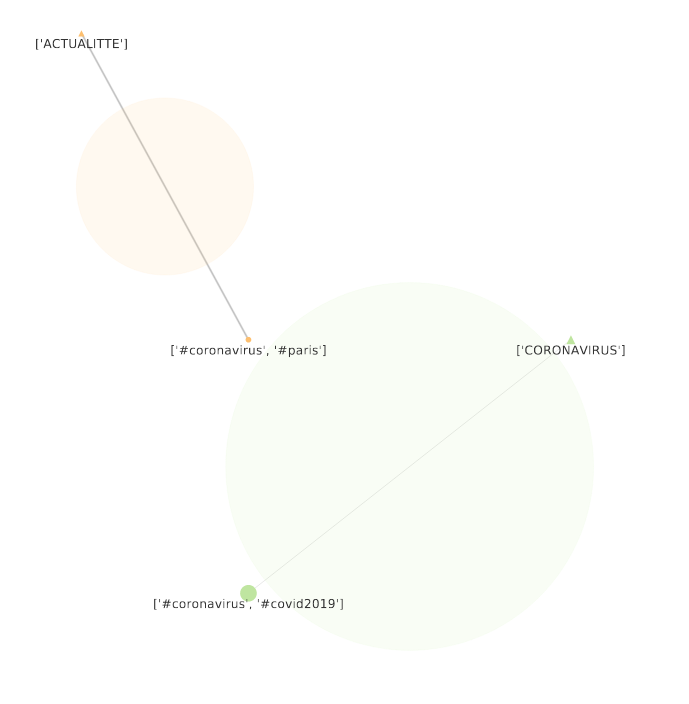
Remarques:

* occurrences les plus parlantes: “Chine ou d’Italie”, “faiblesses des hôpitaux”, “l’Angleterre donne l’alerte”, “Des cas relevés”, “Coronavirus en Belgique”, “des masques”, “des enfants”
* noms propres à forte occurrence: “Chine ou d’Italie”, “Edouard Philippe”, “La France”

|  |
| --- |
| *Quelques éléments d’analyse préliminaire:*  Entre le début de la crise en France et le pique d’infection au mois de mars, on remarque quelques indices d’évolutions dans les termes les plus cooccurrents dans les échanges sur Twitter.  Avant que la crise ne touche la France, le coronavirus semble être une thématique dénuée de sens politique. En effet, l’extraction de termes révèle l’historique de la maladie (MERS-CoV) et de ses différentes mutations. Par contre, dès la fin du mois de janvier 2020, les conversations autour du virus prennent un ton plus politisé (reste à voir si militant) lorsque la crise sanitaire s’ajoute à la crise sociale déjà en cours, notamment les mobilisations contre la réformes des retraites. En mars, la mort de personnalités célèbres font beaucoup parler et la situation de crise semble bel et bien s’être installée. Au mois d’avril, on sent l’évolution de l’épidémie dans d’autres pays (Chine, Belgique, Italie, Angleterre). Les cas explosent un peu partout dans le monde. Les structures d’accueil montrent leur incapacité à traiter tous les malades face à la rapide évolution du nombre de cas. |

**Réseaux sémantiques :**

**Problème pour générer le réseau de cooccurrences : il faudra nettoyer les données pour obtenir des résultats significatifs**



*Réseau février (gauche) et mars (droite)*

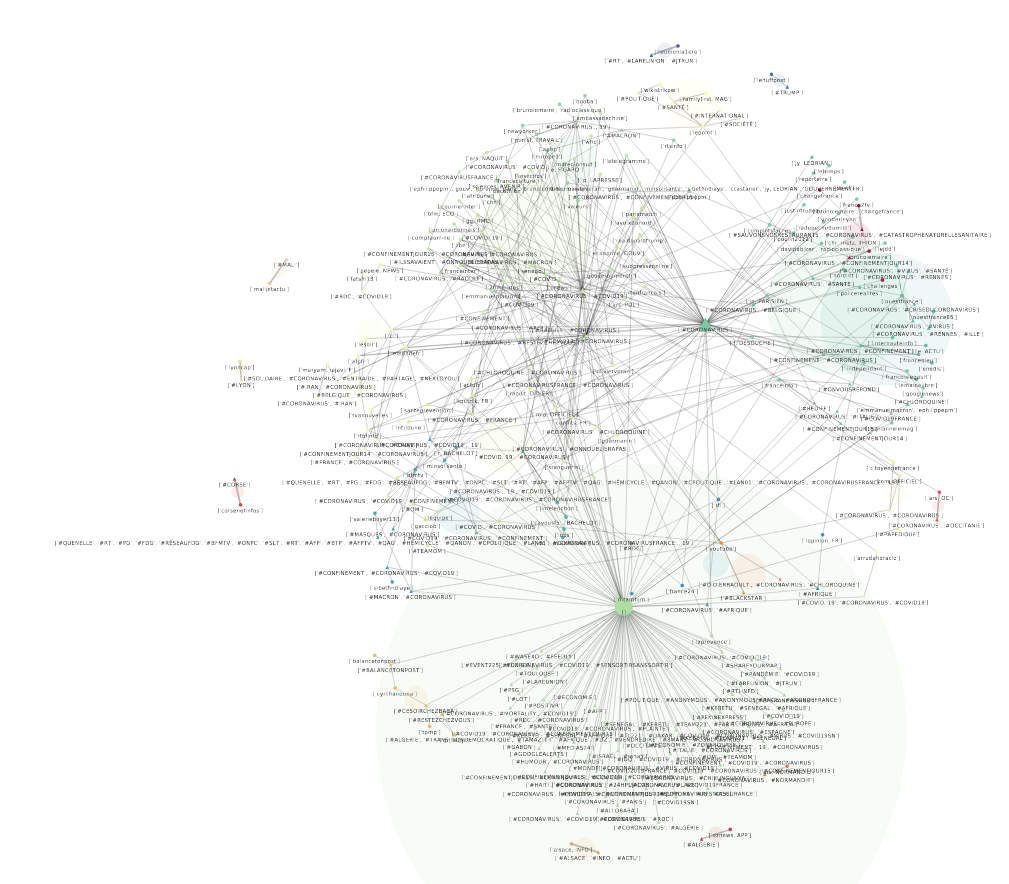
# 08/06/2020 : Doc de travail équipe

1. **Rédaction des analyses préliminaires pour l’extraction de termes (encadré gris, ci-dessus)**
2. **Analyse qualitative des tweets**

Ressource:

<https://www.bu.edu/tech/files/2017/09/Python-for-Data-Analysis.pdf>

1. **Création de la présentation Power Point + préparation réunion (toutes)**
2. **Réduction du dataset du mois de mars aux Tweets contenants des hashtags + création des réseaux de cooccurrence des hashtags et des mentions (Martina)**



L’image ci-dessus correspond au fichier [PDF](https://drive.google.com/drive/folders/13xbTvqTjcUhHffAENYiEeza1RUzFAJvL) partagé sur le Drive

# 08/06/2020, 16h : réunion équipe

**Absentes**: AV., RM.

**Présentation des résultats:**

<https://drive.google.com/drive/folders/1jyoeF5h-wYUA-BLEt8qJmYyB48zbpuhl>

**Remarques de SM:**

Avec la requête large, uniquement sur coronavirus, on retrace bien la prise de conscience autour du virus via les tweets, mais ce n’est pas la meilleur manière d'appréhender les mouvements sociaux

Pour cela, il faudra extraire uniquement des tweets contenants des hashtags. ça réduira le volume de données et nous renseignera uniquement sur les mouvements sociaux actifs pendant la période.

Faire attention à la structure des réseaux (celui de mars pour comprendre d’où viennent les chiffres)

D’après la présentation faite, on peut dire que les tweets qui vont nous intéresser commencent en mars (revendications sociales, les salaires, etc.)

On pourra donc se limiter à la période mars-avril 2020.

Pour construire une requête Twint avec plusieurs hashtags: [https://github.com/twintproject/twint/issues/165](https://slack-redir.net/link?url=https%3A%2F%2Fgithub.com%2Ftwintproject%2Ftwint%2Fissues%2F165&v=3)

**Remarques de AV postées dans Slack:**

“Je pense que étant donné que les nouvelles changeaient très rapidement, il faudrait considérer des étapes de temps d'une semaine et non pas d'un mois. Donc, refaire le même travail mais par semaine. [...] et peut-être de prendre les 5 tweets les plus RT.”

# 09/06/2020 : Doc de travail équipe

1. **Effectuer de nouvelles requêtes (Requête 3) à partir de Jupyter Notebook.** 
   1. **mars-avril : Maya**
   2. **mai-juin : Aïcha**

**Référence**: le Notebook Jupyter “test\_twint.ipynb”, [accessible sur le Drive](https://colab.research.google.com/drive/1EreOEPWMHSPQh76fJdfW8gWB_66W6aJ4), contient la totalité des scripts Python exécutés pour cette requête.

import twint

!pip install nest\_asyncio

import nest\_asyncio

nest\_asyncio.apply()

c = twint.Config()

c.Search = "#coronavirus #covid\_19 #covid19" <- Faites super gaffe ici, je crois qu’il faut utiliser "#coronavirus OR #covid\_19 OR #covid19" sinon par défaut ca sera AND. vérifiez. (SM)

c.Hashtags = True

c.Format = "Tweet id: {id} | Hashtags: {hashtags} | Tweet: {tweet}"

c.Since = "2020-03-01"

c.Until = "2020-05-01"

c.Lang = "fr"

c.Pandas = True

twint.run.Search(c)

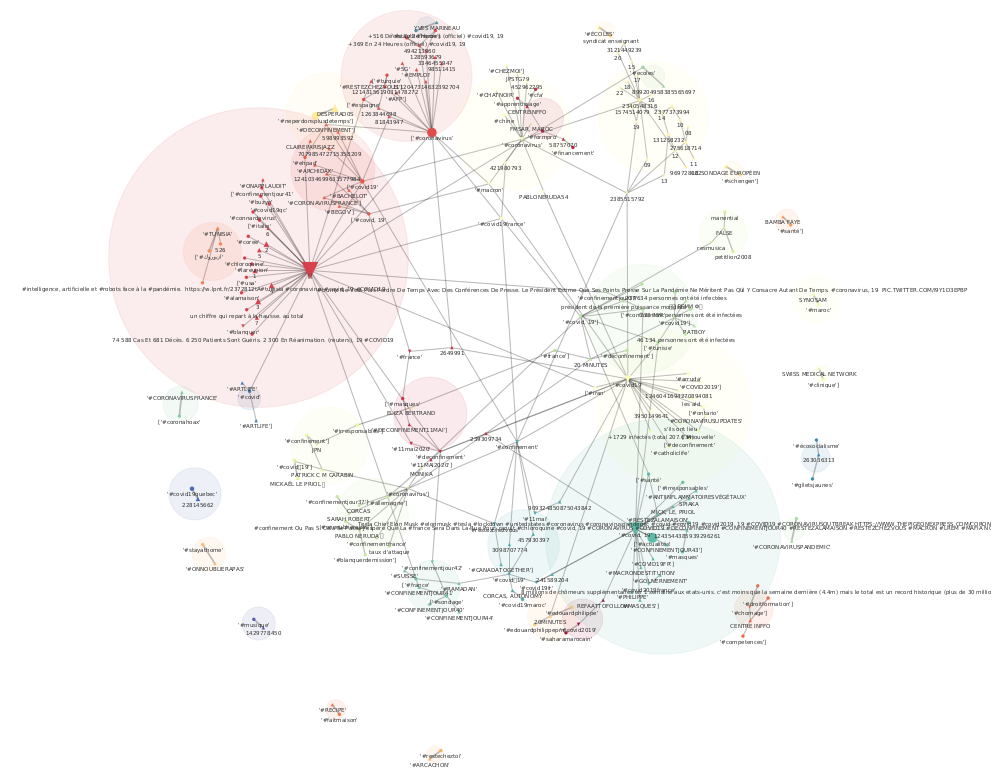
Tweets\_df = twint.storage.panda.Tweets\_df

1. **Générer le réseau de cooccurrence des hashtags des nouveaux datasets avec Cortext (Maya + Aïcha)**

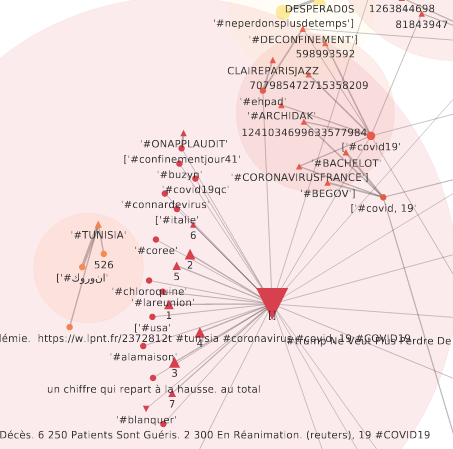
**BUG: Problème rencontré lors du paramétrage du Network Mapping dans Cortext : une erreur se produit lorsqu’on génère un réseau homogène (en indiquant la même variable dans le champ 1 et le champs 2 de l’onglet ‘Nodes’).**

**Nous arrivons cependant à générer des réseaux hétérogènes comme celui ci-dessous:**

**Requête 3 : mars + avril**

****

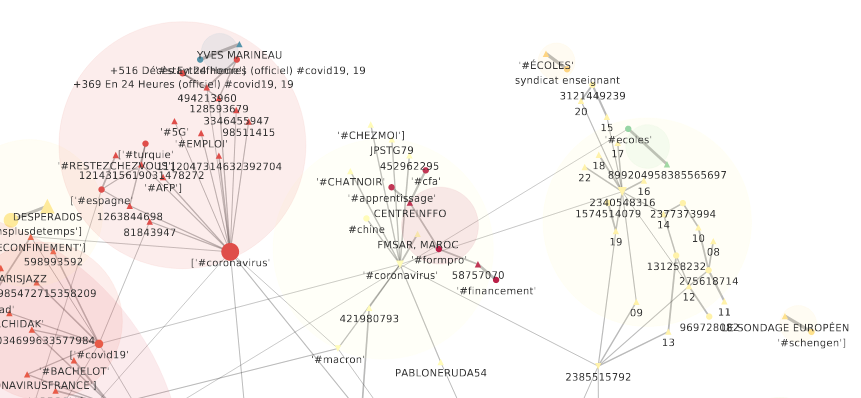
***Graph: Réseau hashtag + name***

******

***Cluster 1 #onapplaudit #connardvirus #chloroquine #alamaison***

***Cluster 2 #Tunisia***

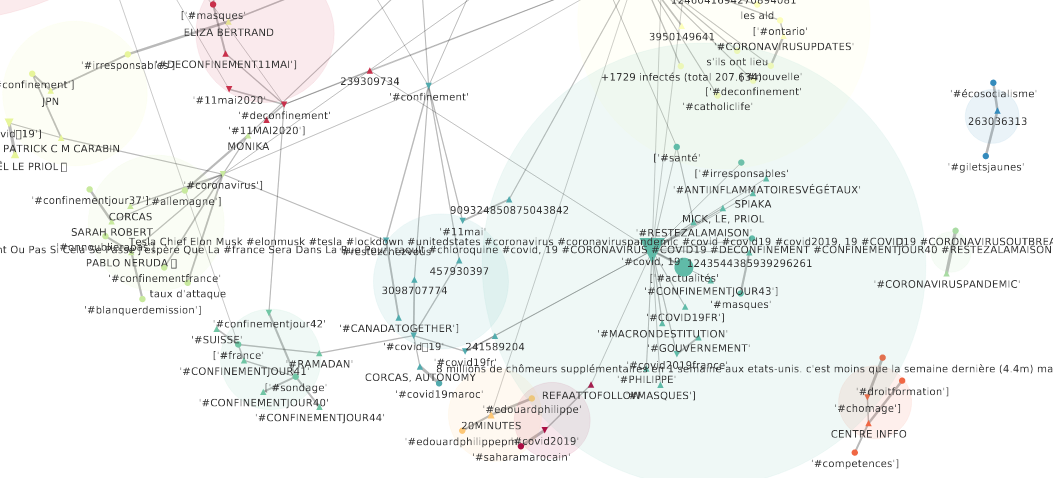
***Cluster 3 #ehpad #archidak #bachelot #begov #coronavirusfrance***

******

***Cluster 1 #espagne #emploi #turquie #restezchezvous #AFP***

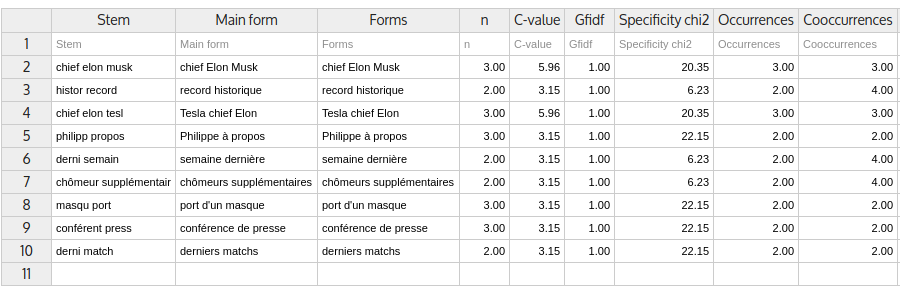
***Cluster 2 #apprentissage #chezmoi #chatnoir #centreinffo #chine #FMSAR #Maroc***

***Cluster 3 #ecoles #syndicat enseignant***

******

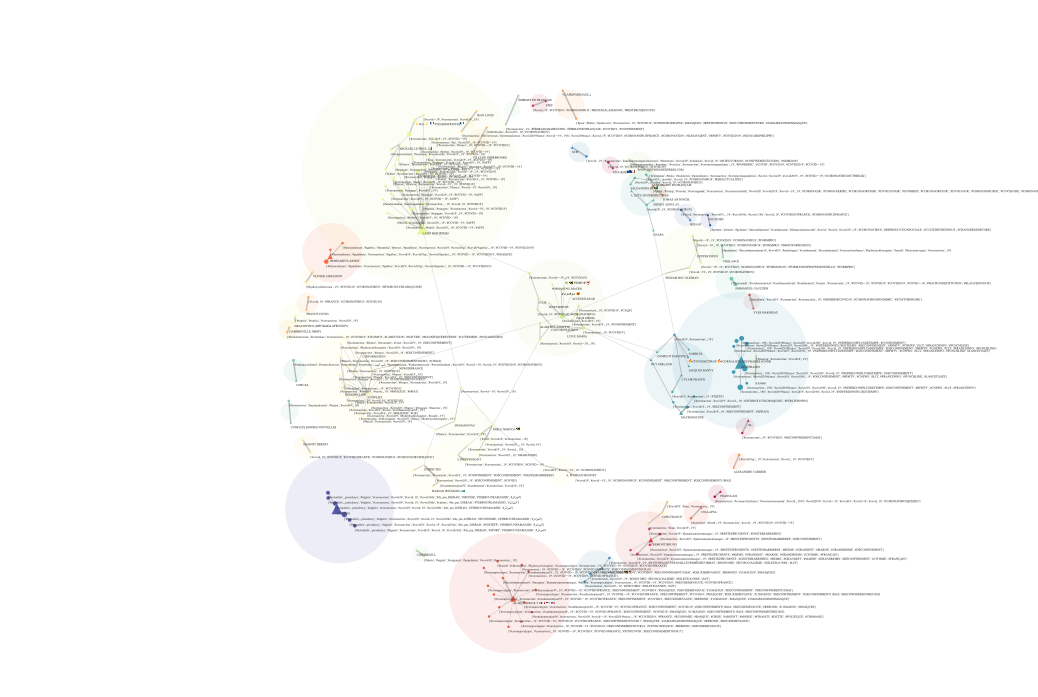
***Cluster 1 #Antiinflammatoiresvégétaux #irresponsables #santé #macrondestitution #gouvernement #masques***

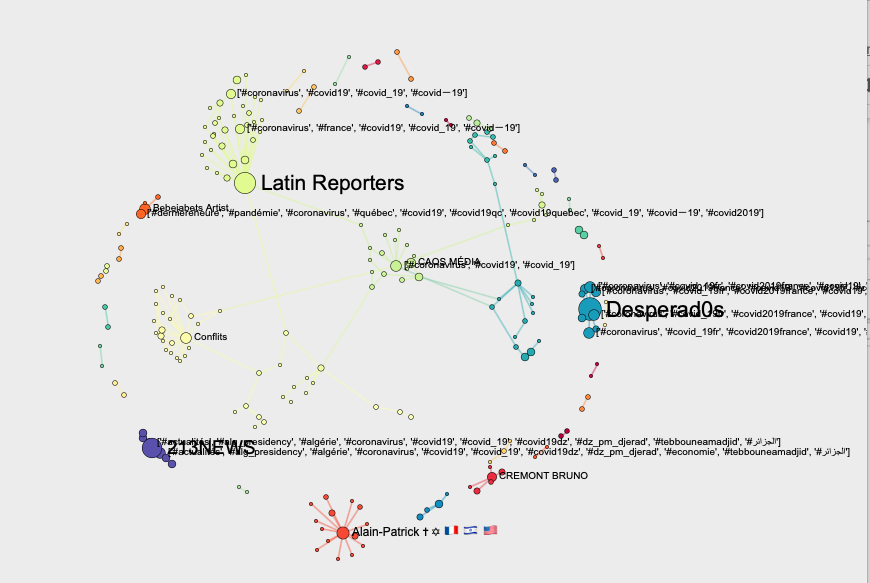
***Cluster 2 # confinementfrance #blanquerdemission***

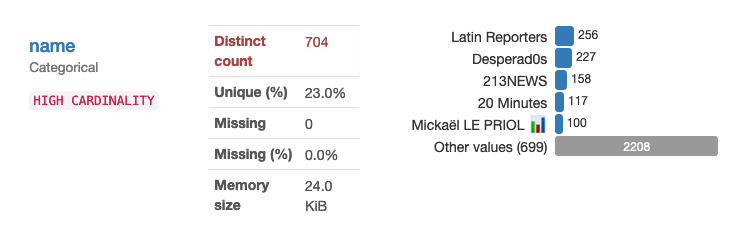
******

***Tableau: Extraction de terms (top 10)***

**Requête 3 : mai + juin**

******

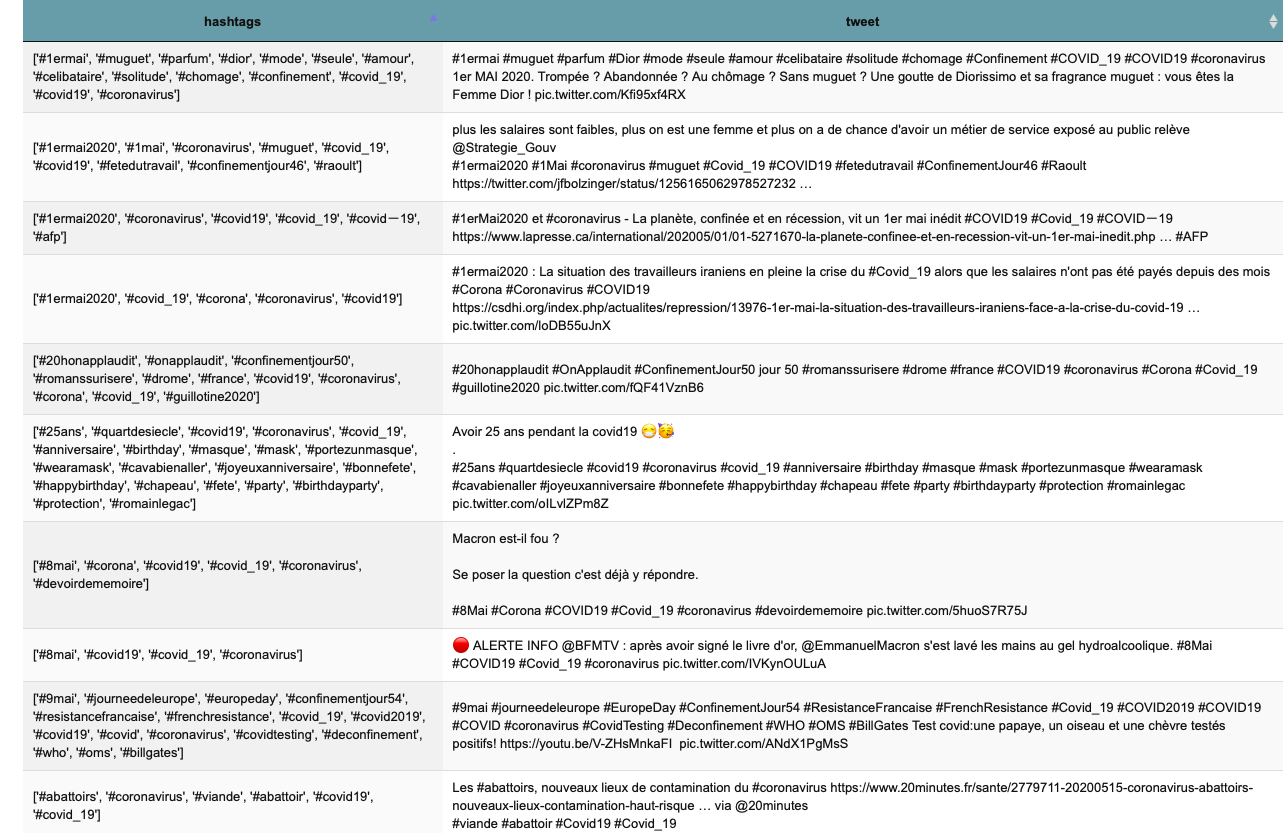
******

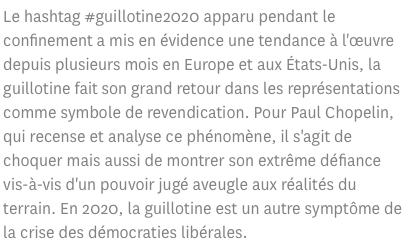
****

**Graph : Réseau hashtag + name**

**Croisement Hashtag et tweet** [**https://drive.google.com/file/d/1sis7YKOZnSR4HrsWGQRPYHJzoQ-E2oIo/view?usp=sharing**](https://drive.google.com/file/d/1sis7YKOZnSR4HrsWGQRPYHJzoQ-E2oIo/view?usp=sharing)

En mai, nous avons des hashtags tels que #guillotine2020, #frenchresistance, #resistancefrancaise, #adamatraore, #justiceforwazizi, qui font référence à de la revendication sociale





1. **Comparer avec le df réduit de mars (et avril)?**
2. **Faire une analyse qualitative des 5 tweets les plus retweetés par semaine pour la période mars-avril. (à faire)**

# 15/06/2020 : Doc de travail équipe

1. **Lancer la Requête 4** 
   1. **mars : Martina**
   2. **avril : Maya**
   3. **mai : Aïcha**

import twint

!pip install nest\_asyncio

import nest\_asyncio

nest\_asyncio.apply()

c = twint.Config()

c.Search = "#coronavirus OR #covid\_19 OR #covid19"

c.Hashtags = True

c.Format = "Tweet id: {id} | Hashtags: {hashtags} | Tweet: {tweet}"

c.Since = "2020-03-01"

c.Until = "2020-05-01"

c.Lang = "fr"

c.Pandas = True

twint.run.Search(c)

Tweets\_df = twint.storage.panda.Tweets\_df

1. **Lancer les analyses suivantes sur Cortext**
   1. Corpus Explorer
   2. Network Mapping
2. **Préparer la réunion**

# 15/06/2020, 15h : réunion équipe

* Passage en revue des graphs

*“même dans le reseau bazar que vous avez on peut deviner des cluster” (SM)*

* Les encadrants partagent le projet #balancetonmetro avec les étudiantes pour montrer un exemple de réseau de hashtags réussi
* Les étudiantes partagent le projet #mouvements sociaux et sollicitent les encadrants afin de d’avancer sur la question du paramétrage du script Network Mapping sur Cortext

**Consignes:**

* Contacter Lionel Villard, [lionel.villard@esiee.fr](mailto:lionel.villard@esiee.fr) pour debug sur Cortext
  + **Fait le 15/06 à 18h24 + invitation projet Cortext**
* Commencer à rédiger votre rapport en expliquant la problématique et la chronologie pendant que ca finit de scrapper

*“récapituler les événements entre disons le 13 mars et le 15 juin. Discours de E. :::::::::::::::::::::::!Macron, discours de E. Philippe, nombre de morts symboliquement forts (quand on dépasse les 10 000 puis quand on dépasse les 15000 puis les 20000 morts), déclarations de Sibeth NDiaye sur les masques inutiles, sur les profs et les fraises, etc... Maladie de Johnson, déclarations de Trump” (AV)*

* Faire le graph pour filtrer les données et identifier les mouvement sociaux, et seulement ensuite faire de la recherche fine semaine par semaine

# 16/06/2020 : Doc de travail équipe

1. **Construire une chronologie (Martina)**

*Ressources*:

<https://fr.wikipedia.org/wiki/Chronologie_de_la_pand%C3%A9mie_de_Covid-19>

<https://fr.wikipedia.org/wiki/Chronologie_de_la_pand%C3%A9mie_de_Covid-19_en_France>

*Chronologie*:

<https://drive.google.com/drive/folders/13xbTvqTjcUhHffAENYiEeza1RUzFAJvL>

1. **Collecter les données de la requête 4 (toutes)**

mars :

avril : 418651 observations (du 12 au 30)

mai : 357 389 observations

1. **Nettoyer les données pour les mettre dans cortext (toutes)**
   1. Suivre la démarche indiquée par Lionel Villard dans le projet #Mouvements sociaux corona sprint

1/ Ouvrir le fichier csv avec Calc Open office (virgule comme séparateur, Séparateur de chaine de caractère : guillemets)

2/ Rechercher & Remplacer les retours de chariots en expression régulière : rechercher \n et le remplacer par un espace

3/ Rechercher & Remplacer les doubles espaces et les remplacer par un espace simple (il y en a bcp, faire la manip plusieurs fois)

4/ Dans la colonne hashtags : Rechercher & Remplacer [' par rien. Et Rechercher & Remplacer '] par rien. Et en Rechercher & Remplacer ', ' par *\**\*\*

5/ Enregistrer sous « texte csv », changer de nom, cliquer sur éditer les filtres csv, choisir « UTF8 » et « Tabulation » comme séparateur de colonne, et ne pas utiliser de séparateur de chaine de caractères (supprimer le '').

1. **Scripts à lancer dans CorText:**
   1. Network mapping (chacune pour ses données)

***Import dans Cortext et paramètres***

1/ Zipper le fichier csv

2/ Uploader dans CorText Manager

3/ choisir robust Csv

4/ définir date dans le paramètre Time Field (ne rien mettre ici)

5/ Dans l’option « Date Format,… » Choisir yes

6/ Préciser la granularité des analyses (day, je pense), et choisissez l’année de début (2020)

1. **Documentation de l’échange sur le forum de Cortext (Maya)**
2. **Rédaction rapport (Maya)**

|  |
| --- |
| *Rappel de la présentation du sujet rédigé par les encadrants:*  “La diffusion des technologies numériques d’information et de communication a suscité très tôt un ensemble de réflexions quant à ce que leurs usages pouvaient éventuellement modifier dans l’implication des citoyens, la mobilisation du consensus (et de l’action), ainsi que dans la construction publique des revendications portées par les acteurs des mouvement sociaux. Nous connaissons maintenant le rôle des réseaux sociaux dans l’amplification de mouvement sociaux récents, qu’il s’agisse de #metoo, des gilets jaunes ou des mouvements pour la lutte contre le réchauffement climatique. Mais l’espace public est aussi un espace physique de coprésence où des individus se rendent visibles les uns aux autres, se croisent et, éventuellement, entrent en relation et communiquent. Les mouvements d’occupation des place ou le succès des manifestations, Gilets Jaunes, Nous Toutes ou les Marches pour le Climat en sont des preuves claires. En ces temps de confinement qui sont aussi des temps d’incertitude, aussi bien économique que sociale ou même démocratique et alors que les mouvements précédents ne se sont pas éteints mais que d’autres contestations émergent (manque de matériel médical, conditions de travail, difficultés économiques….), il paraît particulièrement important de comprendre, à travers l’analyse de réseaux sociaux, comment ces revendications continuent de se faire entendre. Quels sont les acteurs de ces mouvements sociaux? Comment s’expriment-ils?  A partir de l’exploitation d’un corpus de tweets, le projet cherchera à évaluer l’émergence ou la continuation de mouvements sociaux autour des thèmes larges de la gestion de crise, du travail et des conditions de travail et de la démocratie. Il sera important de distinguer entre des “vieux” thèmes, tels que les droits des femmes ou les revendications concernant la défense du droit du travail et les thèmes exclusivement reliés à la crise du covid 19 comme la pénurie de masques. Il s’agira d’identifier les principaux acteurs (politiques, médias, twittos influenceurs…), leurs thèmes privilégiés et la temporalité des messages. Dans cette classification d’acteurs, on s’attachera à distinguer entre acteurs “installés” et reconnus sur certaines thématiques (par exemple Caroline de Haas sur les droits des femmes ou François Ruffin dans la défense des d |

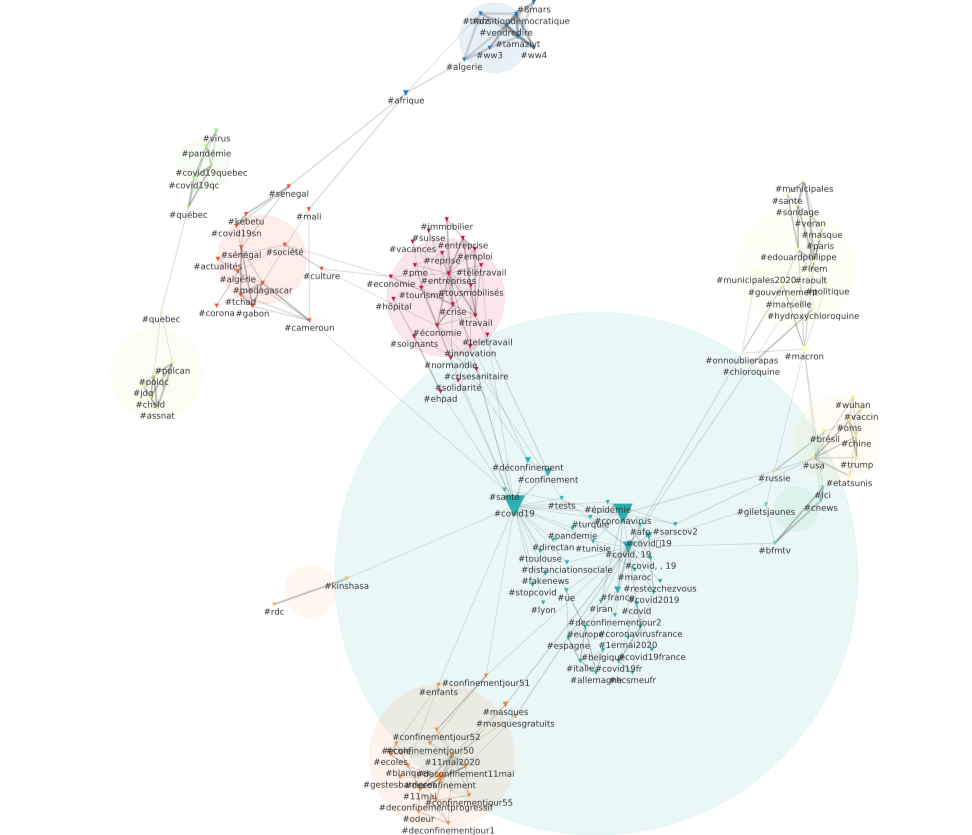
### Lien doc rapport:

<https://docs.google.com/document/d/17p_m8VcONgYaxgn2Kkwt3dUZt9ZYLFWBy57PLEgQecg/edit?usp=sharing>

# 22/06/2020 : Doc de travail équipe

1. **Visualisations sur CorTexte (voir les instructions ci-dessus) (Aïcha)**

Visualisation réseau de cooccurrence des hashtags pour le mois de mai 2020

****

**Paramètres:**

Number of nodes: 200

Automatically define the Proximity Measure: no

Proximity Measure: distributional

Edges filtering advanced settings:

Number of top edges to consider: 9999

Number of top neighbours to consider: 4

**Question pour les encadrants:**

Doit on spécifiquement parler de ces mouvements à l’échelle de la France sachant que notre extraction ne prend pas en compte le paramètre zone géographique?

**Réponse SM:**

1. Filtrer la base de données pour ne retenir que les hashtags qui nous intéressent
   1. Cluster 1 (rouge): #tousmobilisés #hôpital #crise #travail #économie #soignants #emploi #entreprise #solidarité #ehpad #crisesanitaire
   2. Cluster 2 (jaune): #municipales #sante #sondage #veran #masque #edouardphilippe #lrem #municipales2020 #raoult #gouvernement #politique #hydroxychloroquine #chloroquine #macron #onnoublierapas
   3. Cluster 3 (vert): #giletsjaunes #stopcovid #fakenews #santé #tests
   4. *Aborder les clusters avec thématique africaine seulement si on peut les contextualiser*
2. Refaire une visualisation à partir de cette base de données filtrée
3. Procéder à l’analyse qualitative des tweets avec une granulartié temporelle plus fine (semaine par semaine)

**Réalisation de la présentation pour accompagner le rapport (Martina)**

**Rédaction du rapport (voir lien ci-dessus) (Maya)**

**Tache de fond: générer un csv à partir de la requête et nettoyer la base de données (toutes)**

**Pour une récolte propre:**

import pandas as pd

import csv

import nest\_asyncio

import twint

nest\_asyncio.apply()

c = twint.Config()

c.Search = "#coronavirus OR #covid\_19 OR #covid19"

c.Hashtags = True

c.Format = "Tweet id: {id} | Hashtags: {hashtags} | Tweet: {tweet}"

c.Since = "2020-04-01"

c.Until = "2020-04-30"

c.Lang = "fr"

c.Pandas = True

c.Output = “tweets.csv”

c.Store\_csv = True

twint.run.Search(c)

Tweets\_df = twint.storage.panda.Tweets\_df

**Ressources/tutos:**

*Utilisation de Twint:*

<https://medium.com/@michael45684568/using-twint-for-twitter-data-gathering-d7197a3d4ce1>

*Création de subsets avec Pandas:*

<https://medium.com/dunder-data/selecting-subsets-of-data-in-pandas-6fcd0170be9c>

# 23/06/2020 : Doc de travail équipe

**Point sur la collecte des données:**

* **mars :** Un premier fichier comporte les donnée relatives à la période du 25 au 31 mars. Les autres données relatives les autres jours du mois de mars seront collectés par la suite.
* **avril** : Un premier fichier comportant 418 651 tweets relatifs à la période du 12 au 30 a été collecté. Les données du 1er au 11 seront collectées par la suite.
* **mai** : Fichier complet de 357 389 observations allant du 1er au 31 mai.

**Principaux obstacles rencontrés pendant la collecte:**

* connections Internet intermittentes
* mémoire ordi saturée > besoin de reformater le disque dure

**Mise en ligne des csv nettoyés:**

Un dossier dédié a été créé sur Google Drive : csv\_nettoyés

<https://drive.google.com/drive/folders/15W1Nd2a9RMNHL6XY1l_WnZxX0p6pLxqE>

# 23/06/2020 : Point équipe, 15h30

**Présents: SM, AS, MAG**

**Consignes:**

1. **générer un premier réseau sur cortext pour chaque mois (comme celui du mois de mai ci-dessus)**
2. **choisir les hashtags en lien avec les mouvements sociaux**
3. **faire la liste des hashtags d’intérêt et les insérer dans la boucle suivante:**

df["hashtags"] = df["hashtags"].astype(str)

hashtag\_a\_garder = ["#tousmobilisés","#hôpital", "#crise", "#travail" ,"#économie", "#soignants", "#emploi", "#entreprise","#solidarité", "#ehpad","#crisesanitaire"

"#municipales", "#sante","#sondage"," #veran", "#masque"," #edouardphilippe","#lrem"," #municipales2020"," #raoult","#gouvernement","#politique", "#hydroxychloroquine", "#chloroquine" ,"#macron", "#onnoublierapas", "#giletsjaunes ","#stopcovid ","#fakenews ","#santé" ,"#tests"]

to\_keep = []

for i in range(df.shape[0]):

row = df.loc[i,"hashtags"]

for y in hashtag\_a\_garder:

if y in row:

to\_keep.append(i)

df = df.loc[to\_keep,:]

1. **générer les réseaux correspondant dans cortext**
2. **procéder à l’analyse qualitative**

# 29/06/2020 : Doc de travail équipe

**Focus sur le mois de mai**

* Filtre de la base avec les clusters 1, 2 et 3
* Base de données: 65 906
* Préparation pour Cortext

**Boucle pour prendre en compte uniquement les hashtags sélectionnés:**

df["hashtags"] = df["hashtags"].astype(str)

hashtag\_a\_garder = ["#tousmobilisés","#hôpital", "#crise", "#travail" ,"#économie", "#soignants", "#emploi", "#entreprise","#solidarité", "#ehpad","#crisesanitaire"

"#municipales", "#sante","#sondage"," #veran", "#masque"," #edouardphilippe","#lrem"," #municipales2020"," #raoult","#gouvernement","#politique", "#hydroxychloroquine", "#chloroquine" ,"#macron", "#onnoublierapas", "#giletsjaunes ","#stopcovid ","#fakenews ","#santé" ,"#tests"]

to\_keep = []

for i in range(df.shape[0]):

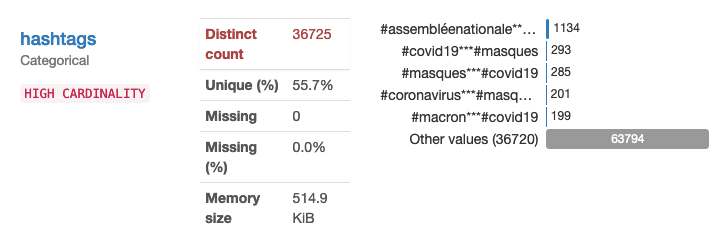
row = df.loc[i,"hashtags"]

for y in hashtag\_a\_garder:

if y in row:

to\_keep.append(i)

df = df.loc[to\_keep,:]

****

**<<<Visualisations réseau de hashtags:**

**Paramètres:**

Number of nodes: 150

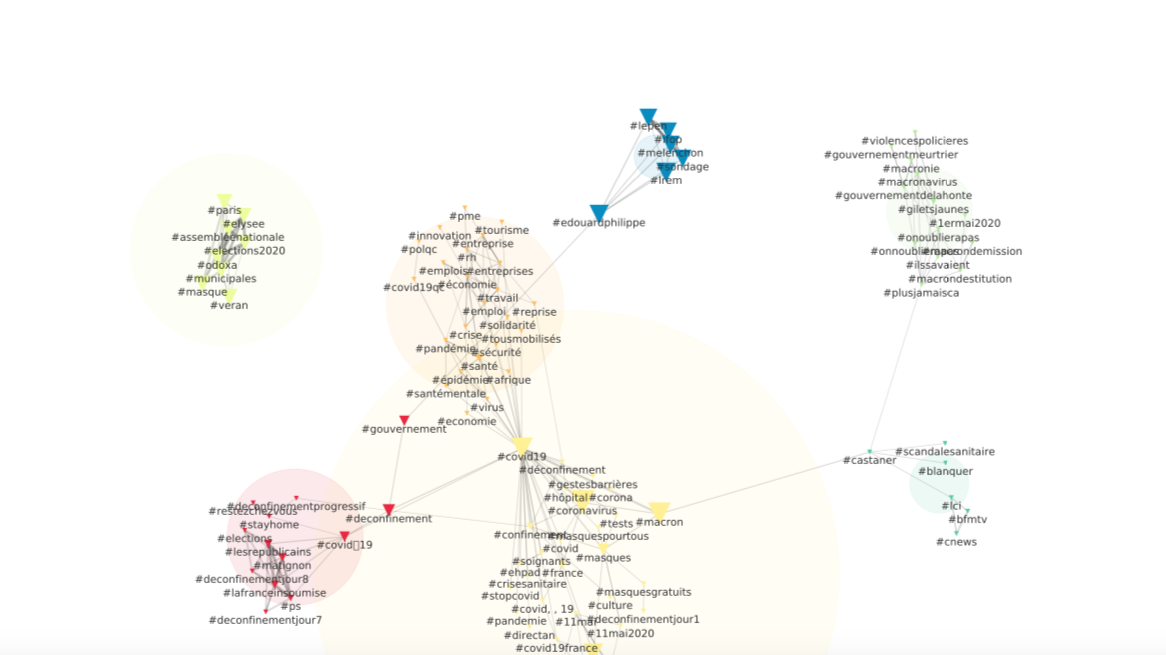
Automatically define the Proximity Measure: no

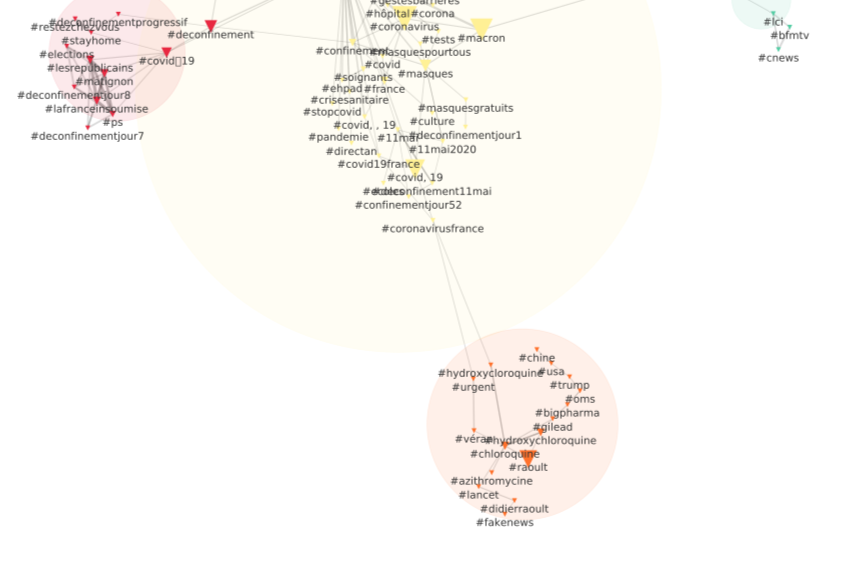
Proximity Measure: distributional

Edges filtering advanced settings:

Number of top edges to consider: 9999

Number of top neighbours to consider: 4

****

****

Mois de mai**:**

Au mois de mai, les événements tournent autour du déconfinement progressif prévu le 11 mai. Les questions tournaient autour des mesures envisagées par le gouvernement et de leur déploiement à l’échelle territoriale.

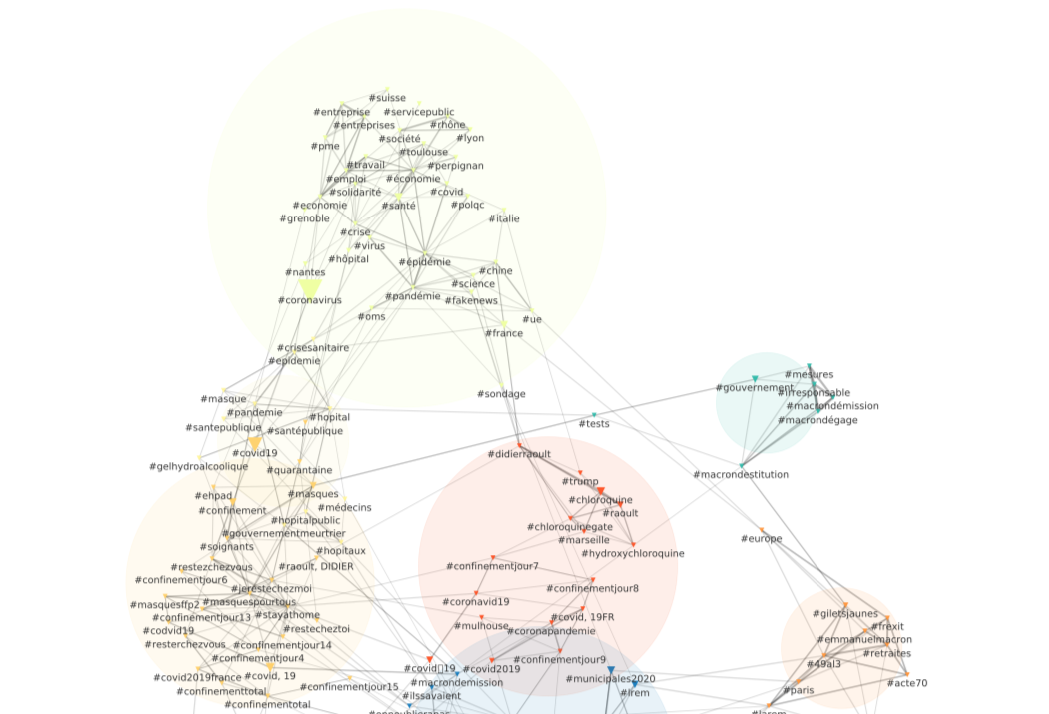
Dans le milieu sanitaire, les personnels soignants en ligne de front dans la crise du Covid, contestent leur conditions de travail et restent sceptique sur la question du déconfinement et de sa gestion.

Des contestations de grande ampleur vont avoir lieu en public malgré l’interdiction de rassemblement le 1er mai. Pour ces nombreux français sortis dans la rue il était vital de dénoncer les bavures policières qui existaient déjà mais ont été plus mis en lumière par le contexte de Covid. Le hashtags #giletsjaunes fait son retour et est associé au #1ermai, #gouvernementmeurtrier, #gouvernementdelahonte, #violencespolicières

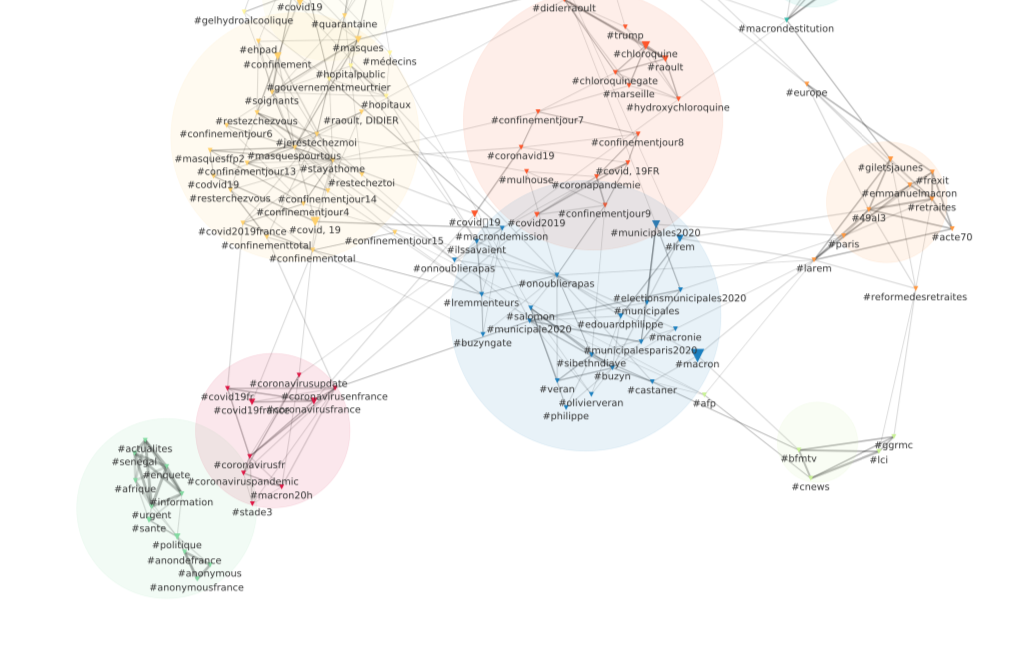
Le maintien des élections municipales au 28 juin a suscité aussi de nombreuses réactions. Des sondages politiques pour questionner sur l’avenir des élections sont faits. L’opposition donne son avis sur la gestion de la crise. Tous ces hashtags mènent vers une controverse sanitaire, sociale, écologique, économique qui va bouleverser toute la sphère politique française.

**Choix final ( à faire ensemble) 1 ou 2 viz**

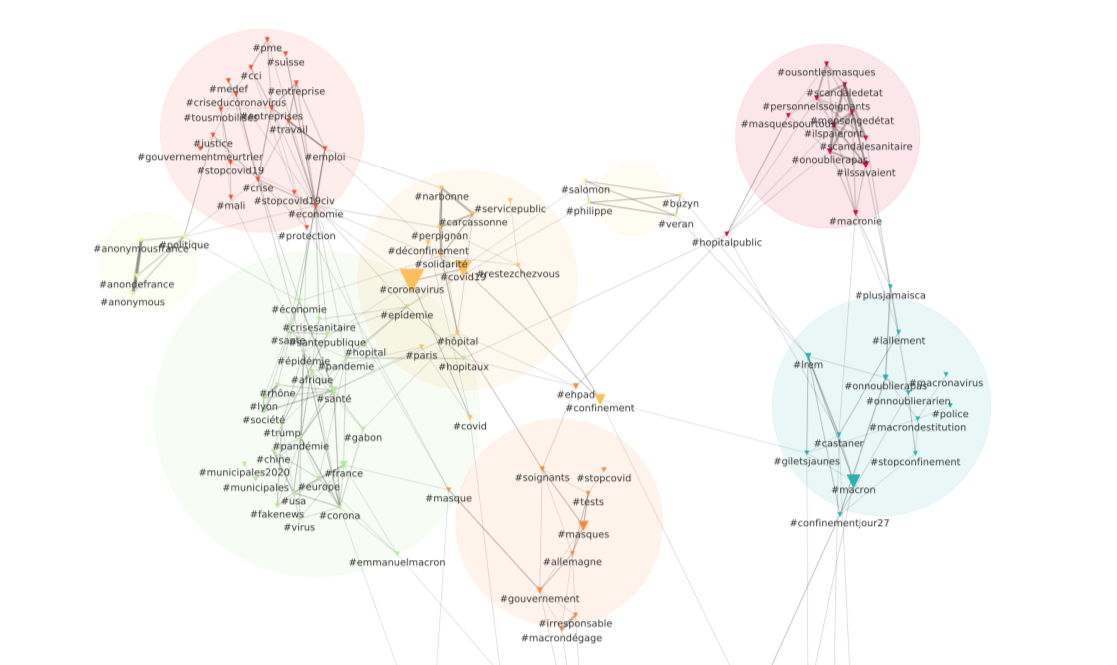
**Df Mars avec les trois clusters**

****

Début mars, des revendications sociales étaient menées sous la bannière des gilets jaunes pour contrer l’applicabilité de la loi 49-3 concernant la réforme des retraites. Les hashtags en lien avec les #giletsjaunes, #macrondémissionne, #macrondégage entre autres servent à exprimer le mécontentement des français. Des manifestations publiques ont eu lieu avant d’être stoppées par la mise en vigueur du confinement.

****

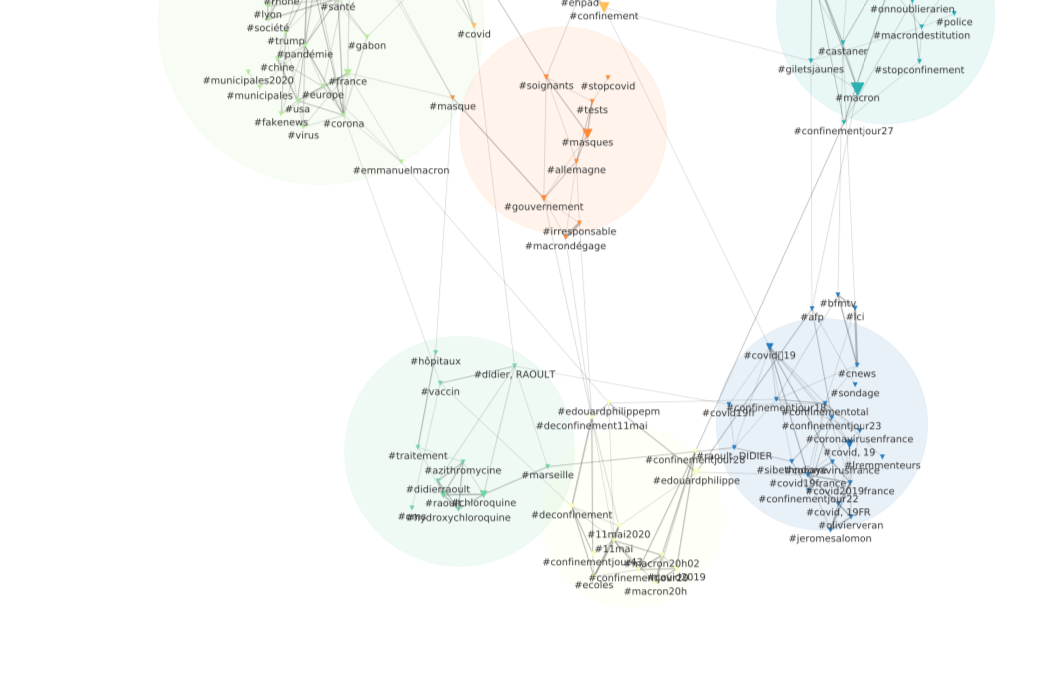
**Df avril avec les trois clusters**

****

En Avril, en plein confinement, les revendications et désaccords s’expriment via les réseaux sociaux et médias. Avec le nombre croissant de morts, la crise sanitaire prend de l’ampleur. Les revendications changent de bord, la réforme des retraites mise en stand-by. On parle de plus en plus de crise sanitaire, un manque criard de matériels à destination des soignants. Ceux-là même qui deviennent acteurs de la contestation. Le nombre de décès ayant augmenté, la population estime que l’état n’a pas été prévoyant et lui impute toute responsabilité sur ces morts.

De cela naît une vague d’indignations avec les hashtags #onnoublierapas, #mensongedétat, #ilspaieront, #ilssavaient, #plusjamaisca.

De manière nette, dans notre réseau apparaît les hashtags #police, #castaner pour faire référence aux bavures policières pendant la crise (amende abusive, interpellation ou contrôle agressif)

****

# 30/06/2020 : Présentation, 14h

**Commentaires:**

Annick:

1/ granularité temps

2/ expurger les mots utilisés dans la requête

Bilel:

1/ collecte itérative à chaque fois qu’on découvre un nouveau hashtag pour être exhaustif

2/ justifier la méthode des hashtags avec une bibliographie

Marianne Noel:

1/ L’utilisation des hashtags est un vrai choix méthodologique

2/ L’extraction des termes doit être mieux mise en valeur

3/ Pousser la question de ce qu’est un mouvement social en ligne

Varelia Ramirez

1/ contrast entre la réalité et ce qu’il se passe en ligne

2/ possibilité d’illustrer l’article avec une/des photos