# LES ERISES

MÉDIAS 9.août.2018 // Les Crises

## L'étude de l'influence russophile sur le Benallagate : d'énormes failles méthodologiques



Suite au buzz déclenché par Nicolas Vanderbiest de EU DisinfoLab sur le "gonflage numérique" de l'affaire Benalla sur Twitter, nous avons publié cet article mardi dernier – regrettant, comme beaucoup, que l'étude complète n'ait pas été disponible au moment des déclarations publiques de l'auteur (comme ici).

- I. Nos 13 questions, et les réactions de EU DisinfoLab
- II. Analyse de l'étude EU DisinfoLab
- III. Partie 2 Méthodologie
- IV. Partie 3 Analyse
- V. Partie 1 Enseignements
- VI. Parties 4 et 5 Le EU DisinfoLab et FAQ

#### I. Nos 13 questions, et les réactions de EU DisinfoLab

Nous posions 13 questions à l'auteur, en lien avec ses déclarations – nous avons eu depuis quelques réponses :

- 1. comment sont identifiés les tweets sur l'affaire Benalla ? Quels sont les mots clefs retenus ?
- comment Nicolas Vanderbiest définit-il scientifiquement un "bot" et du "militantisme"
   ?
- 3. comment Nicolas Vanderbiest définit-il scientifiquement un "troll" ?
- 4. comment sont traités les Retweets dans l'étude ?
- 5. comment Nicolas Vanderbiest définit-il scientifiquement un "gonflage numérique" ?
- 6. pourquoi Nicolas Vanderbiest ne donne-t-il pas une dizaine d'autres exemples de même type pour analyser la façon dont se propage l'information sur Twitter ?
- 7. comment Nicolas Vanderbiest définit-il scientifiquement un "écosystème russophile" ?
- 8. pourquoi avoir choisi de définir spécifiquement un "écosystème russophile" ? Est-il prévu, par déontologie élémentaire, de définir et mener des études avec un "écosystème américanophile", "écosystème europhile", "écosystème sinophile" etc...
- 9. comment Nicolas Vanderbiest définit-il scientifiquement une personne "russophile"
  ? Et comment qualifie-t-il une personne qui n'est pas "russophile" ce qui semble être son cas ?
- 10. Nicolas Vanderbiest a-t-il bien ainsi intégré ainsi les comptes dans cet écosystème ? A-t-il tenu compte de volonté de désinformer ? Si oui, comment, quels en sont les critères ? Y a t-il des critères de diffusion de désinformation en lien avec "l'intérêt gouvernemental" ou "l'intérêt américain" ? Sinon, pourquoi ?
- 11. Nicolas Vanderbiest a-t-il donc compté les retweets de ce compte PoteRusse comme des tweets ?
- 12. pourquoi Nicolas Vanderbiest n'a-t-il pas pris la peine de contacter ce compte avant de le clouer publiquement au pilori, afin de valider son analyse ?
- 13. pourquoi analyser simplement le nombre de tweets et non pas leur portée ?

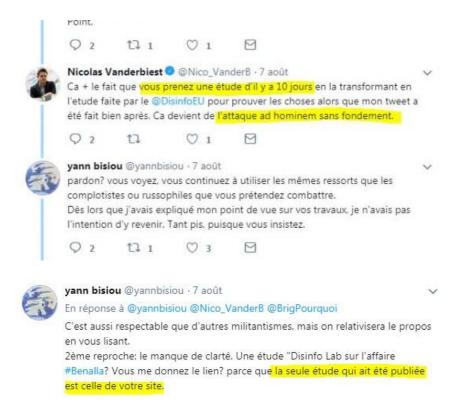
Pour celles et ceux qui n'ont pas Twitter, lorsque Thomas Guenolé a retweeté notre étude, Nicolas Vanderbiest lui a répondu publiquement ceci (source : son tweet) :





Si je comprends bien, il ne répond donc pas à la plupart des questions, et nous reproche d'avoir commenté son étude à lui (qui parlait du volume de l'affaire Benalla sur Twitter), et non pas celle concernant le "gonflage numérique" de l'affaire Benalla – étude qui hélas n'était pas disponible quand le billet est sorti, plusieurs jours après ses propres déclarations sur Twitter... Mais il tente régulièrement de faire croire qu'il existait une autre étude avant le 8 aout:





Ceci étant, ce n'est pas la première fois que ce chercheur pose question – comme il y a un an à l'occasion des MacronLeaks :

# Présidentielle 2017: Jack Posobiec, le militant pro-Trump qui a relayé les "MacronLeaks" 06/05/2017

Jack Posobiec, "militant d'extrême droite connu sur la Toile, a été le premier à en parler sur Twitter" à ses 100.000 abonnés, en ajoutant le hashtag #MacronLeaks, selon Nicolas Vanderbiest, spécialiste de la propagation des rumeurs sur Twitter.

Par Pauline Moullot — 6 mai 2017 à 08:31 (mis à jour le 7 mai 2017 à 15:39)

Le chercheur Nicolas Vanderbiest a retracé pour

«Liberation» la propagation sur 1 witter des fuites du MacronLeaks.



Vous avez facilement retrouvé l'origine des premiers tweets qui ont relayé la rumeur?

C'est limpide. C'était le twitto Jack Posobiec, qui tweete le premier à

MACRONLEAKS

### Le premier internaute à avoir relayé les MacronLeaks est en fait William Craddick

Par Pauline Moullot - 8 mai 2017 à 17:39

Contrairement à ce qui a d'abord été écrit, c'est bien cet Américain, qui se décrit comme un «journaliste d'investigation», qui aurait été le premier à avoir relayé les «MacronLeaks».

Hélas, il avait alors démontré une capacité d'écoute limitée...



#### II. Analyse de l'étude EU DisinfoLab

EU DisinfoLab a finalement publié son étude mercredi – on peut la lire ici (mais ce n'est pas nécessaire, nous la reprenons en bonne part ci-après). Nous allons donc l'analyser dans ce billet.



"Hyperactivisme" qui restera à démontrer...

Cela commence bien : elle n'est apparemment disponible que sur le web en format Adobe Spark, et ne peut pas être téléchargée en pdf, ni être imprimée :

#### **Exporting PDFs / Printing**

Spark Support

Unfortunately, we currently don't support printing/PDFing of Spark projects.

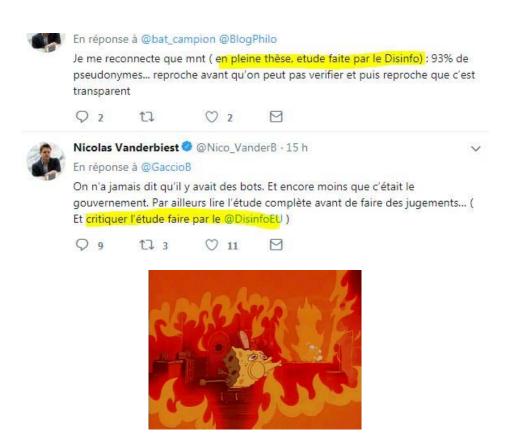
De plus elle ne contient, semble-t-il, aucun nom d'auteur...

Mais on comprend assez bien les nouveaux éléments de langage, dans un contexte où les compétences et l'objectivité de l'équipe sont de plus en plus remises en question : "Il faut sauver le soldat Vanderbiest"

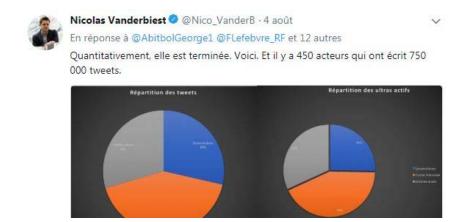


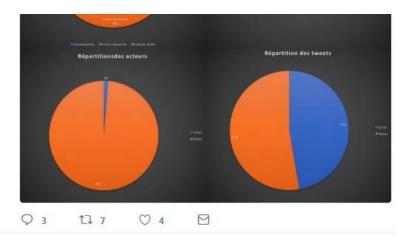
Il s'agit du @DisinfoEU qui a réalisé l'etude, me concentrant sur ma thèse. Toutefois :





Ah bon. C'est tout de même dommage que l'étudiant, qui a écrit tout un article grâce au logiciel Visibrain, qui a communiqué énormément sur la nouvelle étude tout en diffusant des graphiques, ne semble soudain plus tellement intéressé par cette dernière...





Question : cela serait trop demander de connaître le nom du ou des auteurs de l'étude ? Et leurs références en Data Science ?

Car a priori, il n'y a plus que 4 possibilités :



Gary Machado

Gary is our co-founder, with over a decade of EU affairs experience under his belt. He brings his vast EU bubble expertise to EU DisinfoLab, sharing his time with a highly successful international association he has been leading for the past 11 years. Gary holds an MA in European Affairs and BA in Economics from universities in France and Scotland. Despite being French. he's a true Europetrotter, now residing in Brussels, but having also studied or worked in Ireland, Spain, Poland and in the UK. Besides his native French he's also fluent in English, spiced up with basic Spanish and Polish.



#### Our Team







Alexandre Alaphilippe

**Executive Director** 

After a degree in Communications in his homet of Clermont-Ferrand, France, Alexandre has wor for 4 years as Chief Digital Officer at the Clermo Ferrand city hall. Hungry for challenges, he decide to move to Brussels, just "because he liked the common to Brussels, just "because he liked the common to Brussels, just because he liked the common to Brussels, just because he liked the common to Brussels, and the Standard Control of Course). English good as a French man can), Dutch, Spanish (less less) and Italian (more and more).





Martin Mycielski

Special Advisor

Martin serves as Brussels Correspondent for leading Polish daily Gazeta Wyborcza. Having also worked in EU institutions, Polish Permanent Representation to the EU and international NGOs since 2008 – and running political campaigns prior to that – he brings his journalism, EU bubble & political communications experience to EU DisinfoLab. He holds a Master's in Diplomacy & International Relations from Madrid, having also studied English Philology, Iberian Culture and Management at Warsaw University. Raised in the UK and being a dual Polish/Belgian citizen, he is Polish/English bi-lingual with a deceivingly good French accent.



Léa Charlet

Director of Studies

Léa is bringing a wide range of experience and a versatile profile to EU DisinfoLab. She worked in different NGOs, think tanks, small companies and European Institutions, always bringing her knowledge of organisational governance. While graduating in comparative public law and European studies, she has been leading a wide trans-European campaign aimed at raising the turnout of young people during the European Parliament elections of 2014 with her student NGO. In EU DisinfoLab, she is the process queen. She speaks French, English without so much of a French accent, Spanish, Italian and a tiny bit of Czech and German.

L'étude prétendant analyser les interactions de 4 millions de tweets de 250 000 personnes a donc-t-elle été réalisée par :

Réponse A : Gary, avec son Master d'Affaires Européennes et de Coopération Franco-Polonaise ?



Institut d'Etudes politiques de Strasbourg Master of European Affairs, European Affairs and French-Polish Cooperation 2005 – 2006

Il est vrai qu'il a aussi une énorme expérience de "Big Boss"® de l'ONG du numéro d'urgence 112 (c'est probablement le moment d'appeler...)

#### Gary Machado • 2e

Big Boss at EENA; President at EU DisinfoLab; Co-Founder of Saper Vedere; Strategic Advisor at EU Venue Finders





#### **Big Boss**

European Emergency Number Association - EENA112

oct. 2006 – Aujourd'hui • 11 ans 11 mois Brussels Area, Belgium

EENA, the European Emergency Number Association, is a Brussels-based NGO set up in 1999 dedicated to promoting high-quality emergency services reached by the number 112 throughout the EU. EENA serves as a discussion platform for emergency services, public authorities, decision makers, researchers, associations and solution providers with a view to improving the emergency response in accordance with citizens' requirements. EENA is also promoting the establishment of an efficient system for alerting citizens about imminent or developing emergencies.



#### President

EU DisinfoLab

déc. 2017 - Aujourd'hui • 9 mois

Brussels

#### Réponse B : Alexandre, avec son Master en Communication ?



Université Blaise Pascal (Clermont-II) - Clermont-Ferrand

Master's degree, Corporate and administration communications

#### Réponse C : Martin, avec son Master de Management et d'études Ibériques ?



#### Uniwersytet Warszawski

MA and MBA studies, English Philology, Iberian Studies, Management 2003 – 2008

Activités et associations : President of the Institute of Iberian Studies Students' Union 2 terms (3 terms member); Member of the Faculty of Neophilology Council, 2 terms

#### Réponse D : Léa, avec son Master de Droit public comparé européen ?



#### Réponse E : Le Stagiaire ?



Léa Charlet a partagé une publication dans le groupe Communauté européenne du stage.

mar. à 13:28 · Et

We are looking for an "events and fundraising intern".

Apply by 16th of August 4PM Brussels time. Don't hesitate to ask if you have any questions. (We also have a job opening for a social media researcher

http://disinfo.eu/jobs/)



#### **EU DisinfoLab**

6 août à 19:13 · 3

Join our awesome team and become a counter-#disinformation crusader!

We're looking for a smart, curious Events and Fundraising Intern for our Brussels office.

Excellent EN&FR, events / project / fundraising experience an asset: http://disinfo.eu/2018/08/06/06-08-2018 -events-and-fundraising-intern/

"URGENT MERCI!" (source ici et là)

On appréciera la vision de "Militant de la Contre-Désinformation" – qui ne s'exercera à l'évidence pas sur celle de l'Union européenne...

Réponse F: La réponse F?

12/08/2018 10:46 11 of 49

En fait, le seul chapeau de l'étude suffit à en relativiser fortement la crédibilité :

"L'Affaire Benalla, révélée par le Monde, suivie de plusieurs rebondissements de la part de l'exécutif français, existe sans les réseaux sociaux. Dans le même temps, l'affaire Benalla a occupé Twitter en ce mois de juillet dans des volumes que nous n'avions presque jamais croisés auparavant. Ce volume a également été fortement amplifié par un petit nombre de comptes. Nous nous posons alors la question : comment s'exerce cette hyperactivité ? S'agit-il de partages militants ou est-ce également un gonflage numérique contaminé par de la désinformation ?"

Le "Ce volume a également été fortement amplifié par un petit nombre de comptes" n'a guère de sens, puisque par principe Twitter est animé par un petit nombre de comptes…

Après une enquête de longue haleine, l'ONG Les-Crises.fr<sup>©</sup> est heureuse de vous communiquer la source originelle du "gonflage numérique" de l'affaire Benalla, puisqu'avant ce tweet, le volume était quasiment nul :





« Le Monde » identifie, sur une vidéo, un collaborateur de Macron fra... Equipé d'un casque à visière des forces de l'ordre. Alexandre Benalla.

On notera aussi le fait de laisser croire qu'il n'y a qu'une alternative fermée entre "partages militants" et "un gonflage numérique contaminé par de la désinformation".

On voit que l'officine défend son existence, en mettant en avant "la désinformation", alors que l'Affaire Benalla est probablement un très bel exemple d'affaire mobilisant les citoyens sans création de beaucoup de désinformation.

Bien entendu, nous ne disons pas qu'il n'y a pas eu, ici ou là, des rumeurs, erreurs, mensonges ou malveillance – on parle de 4 millions de Tweets! Mais macroscopiquement, il n'y a pas eu de manipulation ayant largement trompés Français et ayant perduré longtemps dans leur esprit. Bref, à ce stade, on voit déjà que l'étude semble très douteuse...

#### III. Partie 2 – Méthodologie

L'étude commençant par les enseignements (sic !), nous sautons cette partie, et y reviendrons, comme il se doit, à la fin. Les premiers mots sont :

Sachant que le mot désinformation que nous utilisons à travers ce rapport est défini comme "la création et le partage délibéré d'informations fausses". (Wardle, 2017)

Définition fort juste – mais on voit mal comment ils vont prouver que le partage était délibéré, et qu'il ne s'agissait pas d'erreurs de bonne foi...

Nous obtenons ensuite la réponse à une question importante : comment ont été sélectionnés les tweets ?

(AlexandreBenalla OR "alexandre benalla" OR BenallaGate OR affairebenalla OR "Vincent Crase" lang:fr) OR (("conseiller en sécurité" OR "adjoint au cabinet" OR collaborateur OR "chargé de mission") (macron OR emmanuelmacron OR "président français" OR élysée OR elysee) lang:fr) OR (CONTREESCARPE lang:fr) OR (benalla lang:fr) OR (crase lang:fr) OR (contains:benalla lang:fr) OR (contains:benalla lang:fr) OR (contains:benalla lang:fr)

macron ("autile viennant me charcher" OR "le equi reconceble d'act mai").

quilsviennentmechercher
allonscherchermacron

Notre seule réserve sera peut-être sur le fait d'avoir pris toute référence à "Macron", ce qui a pu ramener un volume non négligeable d'informations sans lien avec l'affaire.

C'est ensuite que sont amenées certaines définitions, fort surprenantes et critiquables :

Comment les comparatifs entre écosystèmes sont-ils menés ?

Il y a différents types de comparatifs :

 Russia Today et Sputnik: constitué des comptes (les 5000 + actifs) partageants le plus Russia Today et Sputnik en amont de la présidentielle française.

D'accord, mais... pourquoi ? Ou plutôt pourquoi seulement ça ? Pourquoi ne pas faire aussi un écosystème de médias mainstream, d'ailleurs scindé entre gauche et droite... (RTL-Europe 1 – TF1 – BFM et Arte – France Inter etc). Sinon, cela induit un biais colossal... Et encore, c'est une proposition simpliste. Il faudrait mener une étude statistique approfondie, afin d'analyser l'échantillon en composantes principales, pour trouver les regroupements statistiquement logiques.

C'est une telle analyse qui devrait aboutir à prouver que cela a un sens de s'intéresser à RT. Ici la logique est inversée, et donc faussée : on part de la conclusion, et on bâtit l'étude pour prouver qu'on a raison. Illustrons par un exemple : ce choix aboutit à retenir 5 000 comptes "diffuseurs de RT". Imaginons qu'ils soient particulièrement hyperactifs. L'étude va conclure "les diffuseurs de RT ont été hyperactifs dans l'affaire Benalla – suivez mon regard". MAIS imaginons, en fait, que ces comptes partagent systématiquement le contenu de TOUS les médias, RT comme TF1 et le Monde. On voit alors que cela n'a pas de sens de les identifier par la diffusion de RT, la conclusion sera la même si on regrouper les gros diffuseurs du Monde ; on montrait en fait simplement que les "hyperactifs de tout" ont été hyperactifs sur Benalla...

 Réseau des désinformations : nous avions identifié les acteurs partageant de la désinformation durant la présidentielle de 2017. Parmi ceux-là, certains avaient partagé au moins 3 rumeurs dentifiées comme telles par crosscheck.

#### Celles-ci étaient :

- François Fillon, blanchi par la justice. (Et les médias n'en parlent pas)
- · Macron va installer la Charia à Mayotte.
- Macron va instaurer une taxe sur les loyers des propriétaires.
- Jean-Luc Mélenchon s'affiche avec une Rolex.
- Emmanuel Macron est financé par l'Arabie Saoudite.
- "Emmanuel Hollande". (Avant qu'il ne soit repris par François Fillon)
- · "Ali Juppé"
- · Macron serait homosexuel
- "Macron Cahuzac"

On se demande bien comment une identification aussi fine a été faite, à partir de quelles identifications (ce n'est pas précisé), avec quelles requêtes (qui dit qu'elles sont justes et non biaisées ?).

Pour revenir à leur propre définition, comment les auteurs savent-ils que c'est de la désinformation volontaire ? Et pas des erreurs. Ou surtout, de l'humour pour faire rigoler ses amis. "Emmanuel Hollande" et "Macron Cahuzac", ça compte comme "désinformation" – sérieusement ???

3. Macron leaks : il s'agit d'une liste des 1653 premiers comptes à partager le leaks. Les Macron leaks sont la diffusion de documents issus d'un piratage de données (emails notamment) de salariés de la campagne de Emmanuel Macron, mélangeant vrais et faux documents. Ce n'est pas une fake news, mais leur propagation très rapide via des réseaux identifiés comme américains, ayant propagé préalablement de nombreuses fausses informations comme le PizzaGate, était suspicieuse.

Il y a donc un réseau "Macron Leaks" en France – défini à partir d'une infirmation venant de "réseaux identifiés comme américains" (?).

Mais là encore, le document était en grande partie vrai ; c'est logique qu'une telle information se diffuse rapidement.

4. Ecosystème de désinformation russe : identifié par rapport à la propagation active de fausses informations comme #SyriaHoax (pas d'attaque chimique par le gouvernement syrien) ou l'empoisonnement de l'agent Skripal par les services secrets anglais.

Là encore, pourquoi pas, mais pourquoi ne pas avoir défini un "Écosystème de désinformation En Marche" (et il existe!), un "Écosystème de désinformation pro-UE" et un "Écosystème de désinformation américain" ?

#### Comment identifie-t-on un russophile?

Nous avons mis en place une méthodologie qui peut avoir ses limites, voire des contradicteurs. Sur base de deux mois d'activité en février et mars 2017, nous avons identifié les 5000 personnes qui partageaient le plus le contenu de Russia Today et Sputnik en France. Cette identification a été réalisée avant la campagne de la présidentielle française de 2017 et avant les polémiques qui ont fait parler de Russia Today en France, par exemple :

- les accusations d'Emmanuel Macron de désinformation en septembre 2017 ;
- la pétition contre l'autorisation d'émettre de Russia Today en octobre 2017 ;
- le démarrage de l'émission de Russia Today en France en décembre 2017.

Là encore, erreur méthodologique, toujours la même : qui dit que ceux qui partagent RT ne partagent pas tout de la même façon ? Cela-il été vérifié ? Où est l'étude ?

On appréciera particulièrement ceci :

#### Qu'est-ce que l'on entend par "Russophile" ?

Russophile : "Qui est favorable aux Russes ou (anciennement) à l'URSS" - Larousse. Une partie de notre travail porte sur la compréhension des sphères d'influence dans la

diffusion et la manipulation de l'information sur les réseaux sociaux.

Un russophile sur Twitter, c'est qui?

- Partager publiquement et régulièrement des articles de RT et Sputnik, qui sont des médias financés par l'État russe
- C'est donc une sensibilité aux narratifs pro-russes
- Ce n'est pas un jugement de valeur, mais un fait quantifiable selon une méthodologie

"Russophile c'est un fait quantifiable" – comme américanophile?

Un russophile sur Twitter n'est pas

- un espion à la solde de Vladimir Poutine
- un troll payé en rouble dans une ferme à trolls à Saint-Pétersbourg
- une personne passant ses vacances à Moscou ou un lecteur de Dostoïevski

Malgré les soupçons d'ingérence de l'État russe dans plusieurs processus électoraux, nous rappelons qu'il est impossible sur base d'une activité en ligne de qualifier quelqu'un comme étant une personne malveillante à la solde d'Intérêts étrangers. Ceci relève du travail des services de renseignement, ce qui n'a strictement rien à voir avec notre mission.

Et on sent à quel point l'auteur aimerait bien faire partie desdits "services de renseignement"...

La partie "méthodologie" se termine ainsi, sans qu'y figure (pas plus que dans le reste de l'étude!) une information sur la prise en compte des RETWEETS (c'est quand on rediffuse un tweet à sa communauté. C'est comme la différence entre "écrire un mail" et "faire suivre un mail reçu" à tous ses contacts). C'est pourtant fondamental pour la compréhension!

Là encore, la réponse a été obtenue sur Twitter (source);



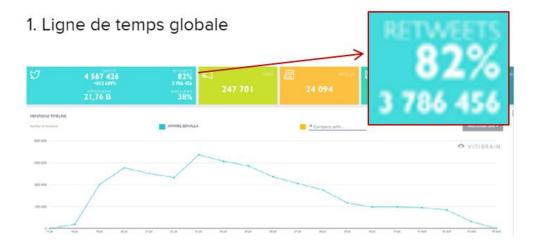
C'est parce que c'est marqué dans la capture

## d'écran du volume. On aurait peut-être du le mettre par écrit également

09:30 - 8 août 2018

#OnAuraitDûLeMettreParÉcrit – oui, sans doute...

Pour illustrer, c'est en fait marqué là, dans la capture en effet :



Du 19 juillet au 3 août, l'affaire a capitalisé 4 567 426 tweets pour 247 701 acteurs. Il s'agit d'un volume de publications exceptionnel. En revanche, le volume d'acteurs est similaire à celui observé lors d'autres phénomènes. À titre de comparaison, voici les volumes pour BalanceTonPorc en français :

Quand les auteurs parlent de 4,5 millions de tweets, il s'agit en fait de 700 000 tweets, retwittés 3,8 millions de fois... (source)



tweet. Par ailleurs pour participer enormement et rapidement, il faut analyser les retweets. Les 3 profils avec automatisations identifiés ne faisaient quasiment que des retweets.

09:29 - 8 août 2018

Question : pourquoi ne pas écrire clairement que les retweets sont comptés comme des tweets ? Pourquoi ne pas le dire systématiquement quand ils communiquent à l'extérieur des résultats ? Cela change tout...

Question : pourquoi ne pas réaliser quelques analyses sur les tweets seuls – vu qu'on parlerait alors ici des seuls créateurs de contenus

D'ailleurs, de même, l'étude ne fait aucune différence entre des tweets et de simples réponses lors de discussions.

EU DisinfoLab a même précisé que (source) :



09:34 - 8 août 2018

C'est là que saute aux yeux le manque de maitrise statistique et méthodologique.

Ils veulent regarder l'activité et la désinformation, mais ne s'intéressent pas à l'influence! Ils vont donc se passionner pour un compte à 2 abonnés qui aurait diffusé 8000 tweets sur Benalla, en expliquant qu'il aura été très actif dans la diffusion du Benallagate?

Mais une telle vision est logique pour eux, car il est certain que l'influence de la désinformation a été négligeable dans cette affaire – et donc que cette officine ne sert à rien sur ce sujet...

Question : pourquoi les auteurs ont ils simplement analysé le nombre de tweets et retweets et non pas leur portée/influence ? À quoi cela sert-il de tenir compte d'une masse de tweets s'ils ne sont lus par presque personne ?

#### IV. Partie 3 – Analyse

Les auteurs ont donc identifié 4 567 426 tweets et retweets émis par 247 701 acteurs en 10 jours – soit 2 tweets et retweets par jour en moyenne.

Ce seul chiffre, rapporté aux 22 millions d'utilisateurs mensuels et aux plus de 4 millions d'utilisateurs par jour de Twitter (source) montre bien l'extrême concentration de l'utilisation de Twitter : la grande majorité des utilisateurs ne twittent pas. (source)

#### Nombre de tweets par compte Twitter

0 tweet : 56%

1 tweet: 11%

2 tweets: 6%

3-5 tweets: 3%

• 6-10 tweets: 3%

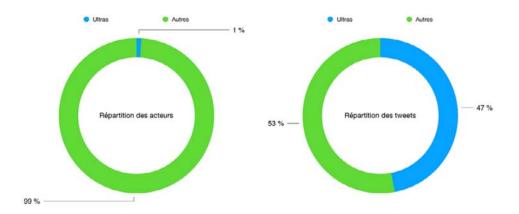
• 11-20 tweets: 7%

21-50 tweets: 6%

• Plus de 50 tweets: 8%

Source : L'utilisateur type de Twitter, en France et dans le monde

Dès lors, le fait qui semble étonner les auteurs – à savoir que "1 % des twitteurs ont publié plus de 47 % des tweets et retweets sur l'affaire Benalla" – n'est pas très étonnant : on sait bien que **la grande majorité des twitteurs twittent peu.** 



#### C'est en fait une loi d'Internet :

#### 1% de créateurs, 9% d'éditeurs et 90% d'audience.

La réalité est que les chiffres sont :

- 1% de créateurs
   C'est à dire, des personnes qui vont créer du contenu pour ou sur vous
- 9% d'éditeurs
   C'est à dire, des commentateurs
- et 90% d'audience.

Prenons comme exemple, les chiffres de Wikipedia, dans le post intitulé Règle du 1%.

- 2006: 1 à 2% des visiteurs de Wikipedia contribuent
- 2005: 50% de toutes les contributions étaient faites par 0,7% des contributeurs et seuls 1,8% des contributeurs avaient écrit plus de 72% de tous les articles

Wikipédia appelle même ça "la règle du 1 %" sur Internet – par exemple en 2005, 1,8 % des contributeurs avaient écrit plus de 72 % de tous les articles de Wikipédia.

Il est donc étonnant que les auteurs aient qualifié "d'hyperactifs" (laissant à penser à une anomalie) les 3 378 comptes ayant parlé de l'affaire Benalla et qui ont twitté **ou retwitté** 

**plus de 200 fois** sur ce sujet en 7 jours . Surtout qu'ils ne mènent aucune étude de concentration sur d'autres sujets !

Question : Pourquoi les auteurs n'ont-ils réalisé aucune étude de concentration sur d'autres sujets du même type ? Comment dès lors peuvent-ils soutenir qu'il y aurait une particularité dans la diffusion de l'information Benalla, et un" gonflage numérique" ? C'est peut-être le fonctionnement normal de Twitter.

Autre point central, sur lequel les auteurs passent très vite :

Dans cette étude, il convient dès lors d'analyser le phénomène de manière quantitative, dans les auteurs et dans les tweets.

Nous avons décidé de retenir les acteurs qui ont publié plus de 200 tweets dans la première semaine. Ce critère volumique peut être considéré comme bas, mais nous avons constaté qu'un tiers d'acteurs émergent à plus de 700 tweets sur la première semaine.

Pour l'analyse, nous avons exclu les médias et les comptes que l'algorithme plaçait entre deux ponts ainsi que la communauté de La République En Marche. Cette dernière était constituée de seulement 27 acteurs (0,7 % des comptes hyperactifs) qui avaient tweeté 11 564 tweets. (0,53 % des tweets), soit une part très marginale des auteurs et du volume.

Vous avez bien lu : "Pour l'analyse, nous avons exclu les médias"!

Pourquoi ? Parce que!

On imagine que les auteurs estiment que, si on s'intéresse à la diffusion de l'affaire Benalla sur Twitter, le rôle de ces médias a été négligeable :

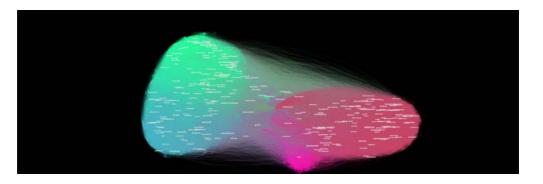
| screen name  | display name | followers | Tweets et<br>Retweets<br>sur Benalla |  |
|--------------|--------------|-----------|--------------------------------------|--|
| BFMTV        | BFMTV        | 2 325 395 | 775                                  |  |
| Le_Figaro    | Le Figaro    | 2 909 220 | 627                                  |  |
| RTenfrancais | RT France    | 94 911    | 473                                  |  |
| LCP          | LCP          | 508 949   | 456                                  |  |
| le_Parisien  | Le Parisien  | 2 061 500 | 380                                  |  |
| 20Minutes    | 20 Minutes   | 2 264 199 | 372                                  |  |

| lemondefr      | Le Monde             | 7 802 768 | 356 |
|----------------|----------------------|-----------|-----|
| Valeurs        | Valeurs actuelles    | 119 101   | 356 |
| LCI            | LCI                  | 125 966   | 340 |
| RMCinfo        | RMC                  | 158 195   | 338 |
| leJDD          | Le JDD               | 354 616   | 317 |
| Europe1        | Europe 1             | 1 367 044 | 307 |
| franceinfo     | franceinfo           | 1 261 743 | 270 |
| infosActus     | Infos Actus          | 75 905    | 251 |
| morandiniblog  | Jean Marc Morandini  | 678 841   | 249 |
| lemonde_pol    | Le Monde Politique   | 321 788   | 236 |
| hangen_claude  | hangen claude        | 68 706    | 235 |
| libe           | Libération           | 2 808 077 | 231 |
| franceinfoplus | franceinfo plus      | 254 942   | 229 |
| LePoint        | Le Point             | 701 114   | 225 |
| Limportant_fr  | L'important          | 105 854   | 221 |
| LeHuffPost     | Le HuffPost []]]     | 1 056 684 | 210 |
| edwyplenel     | Edwy Plenel          | 865 854   | 209 |
| BrunoMasure    | bruno masure         | 52 938    | 206 |
| France         | Actualités en France | 60 002    | 205 |

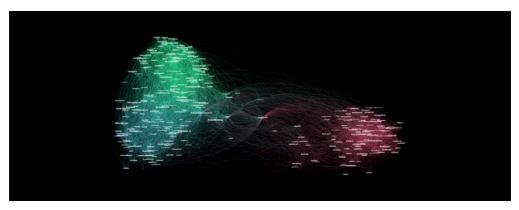
C'est dommage, ils auraient dû directement se limiter aux twittos qui ont le mot "russe" dans leur biographie, on aurait gagné du temps d'analyse...

Question : mais pourquoi diable les auteurs ont-ils exclu les médias de leur étude ?

Les auteurs ont alors réalisé ce graphique via un algorithme d'interactions entre comptes – ici avec plus de 200 tweets et retweets :



#### et là avec plus de 1 000 tweets et retweets en une semaine :



Quand les auteurs redécouvrent l'opposition...

#### Et de là ils définissent 4 communautés politiques :



La communauté constituée par la France insoumise avec des porosités de lien avec les autres communautés qui occasionne quelques faux positifs.\*(2)

La communauté constituée des médias et des porosités avec les autres communautés, et LREM.

La statistique est cruelle, car elle regroupe en une seule communauté "les médias" avec la communauté "En Marche"...:

La communauté constituée des médias et des porosités avec les autres communautés, et LREM.

Rappelons le problème d'attribuer ainsi à un compte une opinion politique, en se vantant de sa fiabilité :

\*(2) Parfois, l'algorithme peut placer un compte dans une communauté ou une autre. Pour cette raison, il peut y avoir quelques faux positifs négligeables.

mais c'est l'objet d'un autre article dédié sur ce site.

Question : pourquoi les auteurs n'ont-ils réalisé aucune analyse statistique sérieuse (ACP etc.) ? Pourquoi n'indiquent-ils jamais les médianes en plus des moyennes – sachant la forte hétérogénéité des contributeurs ?

Les auteurs s'intéressent ensuite aux ultra-actifs (plus de 1000 tweets et retweets) :

Ces <u>ultra-actifs</u> pratiquent allègrement le mass following qui consiste à suivre un grand nombre de comptes pour espérer être suivis en retour. Ces comptes suivent en moyenne 1500 personnes pour être suivis sensiblement par le même nombre. Enfin, la moyenne de tweets par auteur est exceptionnellement haute : entre 46 000 et 63 000 tweets de moyenne selon les clusters.

|          | Moyenne de tweets | Moyenne de following | Moyenne de followers |
|----------|-------------------|----------------------|----------------------|
| Classe 1 | 63880             | 1697                 | 1746                 |
| Classe 3 | 46555             | 1322                 | 1313                 |
| Classe 2 | 50822             | 1560                 | 1895                 |

En se concentrant sur les 25 acteurs les plus prolifiques, on observe même des chiffres allant de 1 149 560 à 343 221 tweets par auteur tandis que le plus gros posteur culmine à **plus de 953** tweets par jour.

Et c'est qu'on retrouve la faille délibérée de l'étude – ne pas analyser l'influence. Car on est sûr d'une chose : un compte (re)twittant 953 fois par jour n'a AUCUNE influence. Car bien évidemment, aucun de ses abonnés (qui ne peuvent qu'être en nombre assez limité) ne lit une telle production !

Voici le compte qui a publié le plus sur Benalla : 5430 tweets

| Tweets | Abonnements | Abonnés | J'aime |
|--------|-------------|---------|--------|
| 726 k  | 2 608       | 5 263   | 184 k  |
|        |             |         |        |

Il n'a que 5300 abonnés, dont le fil est donc noyé... Idem pour un des comptes les plus actifs hors Benalla :

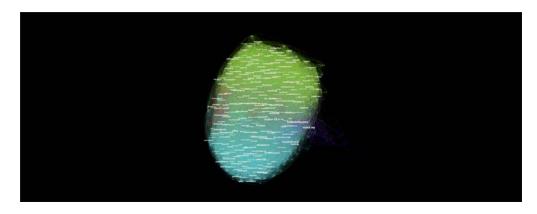
| Tweets | Abonnements | Abonnés | J'aime |
|--------|-------------|---------|--------|
| 1,26 M | 3 629       | 2 259   | 3 714  |

Peu probable que beaucoup des 2 300 abonnés aient lu les 1 260 000 (re)tweets du compte.

Mais pour les auteurs de l'étude, ces comptes sont donc au cœur de leurs préoccupations.

Reviennent alors les "Russophiles" :

En filtrant désormais sur l'écosystème que nous avions identifié en prenant les 5000 profils les plus propices à partager Russia Today et Sputnik, nous obtenons alors une concordance de 27%. Nous rajoutons les personnes identifiées comme présents dans de multiples opérations russes de désinformation (SyriaHoax, rumeurs pendant la présidentielle, etc.), nous avons 86 individus concordants. L'ensemble de cet écosystème est alors le suivant :



Nous constatons ici que le partage d'articles de Russia Today et de Sputnik, ainsi que le partage de fausses informations pendant la présidentielle est corrélé avec une hyperactivité de la part de communautés "souverainistes" et "d'extrême-droite" pendant l'affaire Benalla.

On retrouve ainsi la problématique évoquée précédemment (avoir choisi par principe au hasard un sous-ensemble sans analyser statistiquement la pertinence de ce choix – peut-être que ceux qui partagent RT partagent aussi tout autant LCP). On note également que de nouveau, aucune analyse statistique sérieuse sur les corrélations n'est menée – les auteurs sortent juste un "27 %" sans même qu'on puisse l'analyser (c'est normal, pas normal ?). Et puis il est certain que comme ils enlèvent les médias, ils vont

finir par suffisamment resserrer l'étude pour arriver à ce genre de conclusion fortement biaisée. Enfin, il y a le problème classique de prendre une corrélation pour une causalité : peut-être que "partager RT" est surtout un indicateur très fortement corrélé à "être un opposant actif à Macron"...

#### Ensuite, la plaisanterie continue :

#### 3. Agentivité - Désinformation dans ces communautés

Nous nous sommes ensuite posés la question suivante : est-il tout à fait normal pour certains militants de tweeter 200 fois sur un même sujet ? L'analyse quantitative présentée avait pour

"Nous nous sommes ensuite posés la question suivante : est-il tout à fait normal pour certains militants de tweeter 200 fois sur un même sujet ?"

La réponse est "Oui, en Démocratie".

En revanche, on pourrait se poser la question suivante : est-il normal qu'une microofficine belge partenaire de l'Atlantic Council a priori sans salarié vraiment compétent en statistique se voir qualifiée "d'ONG" et voient ses analyses biaisées reprises par l'AFP et la presse ? (source)





#### Vient alors cette partie cocasse :

objectif d'étudier la part de production de contenu par une minorité d'acteurs. Pour compléter notre étude, il nous fallait évaluer la part d'informations fausses qui aurait pu être propagée par ce réseau. Pour cela, nous avons identifié neuf désinformations qui ont circulé durant l'affaire Benalla : (tout en notant que certaines d'entre elles ont été démarrées par les médias).

- Benalla aurait des gyrophares sur sa voiture de fonction (6881 tweets par 4829 auteurs) -Présence médiatique
- Benalla s'appellerait en fait Lahcene Benalla (13 118 tweets pour 5199 auteurs) Nonprésence médiatique
- Benalla serait un espion marocain (206 tweets pour 158 auteurs) Non présence médiatique
- Benalla pratiquerait des écoutes téléphoniques sur les députés (1716 tweets pour 1506 auteurs) - Présence médiatique via une interview d'EELV
- Benalla allait créer un service secret spécial dont il serait à la tête (4157 tweets pour 2911 auteurs) - Présence médiatique
- Alexandre Benalla est le voisin de Mamadou Gassama (1699 tweets pour 1423 auteurs) -Non présence médiatique
- Benalla était chargé de récupérer l'ADN d'opposants politiques (1082 tweets pour 891 auteurs) - Non présence médiatique
- Macron prend de la cocaïne et Benalla pourrait être son dealer (2220 tweets pour 1657 auteurs) - Non présence médiatique
- Benalla serait l'amant de Macron avec quelques photographies: (3928 tweets pour 2892 auteurs) (pour des questions méthodologiques, notamment du fait que certains le partagent

pour rire, les visuels comme ceux que nous presentons ci-dessous n'ont pas ete consideres pour cette étude) - Non présence médiatique

"Nous avons identifié neuf désinformations qui ont circulé durant l'affaire Benalla": dont AUCUNE désinformation pro-Macron, c'est vrai qu'aucun LREM n'a menti ni désinformé dans cette affaire. Cela ne va donc pas du tout nuire à la fiabilité de l'analyse...

On se rend compte d'abord de la faible portée de ces "désinformations": le volume de (re)tweets est faible. Tellement qu'on se demande comment une information (re)tweetée a été considérée comme pouvant figurer dans cette étude... On parle ici de 35 000 (re)tweets, soit 0,8 % de la base, ce qui montre bien que 99 % des tweets étaient de l'information, ce qui semble déjà un très joli score...

Pour la "désinformation" à propos du gyrophare de la voiture de fonction de Benalla (6881 tweets) – comment dire... Le problème, ce n'est pas tant que c'est RTL qui l'a dit :

# Affaire Benalla : ce que l'on sait de l'affaire qui embarrasse l'Élysée

L'homme avait accompagné Emmanuel Macron lors de sa campagne présidentielle, et faisait partie de sa garde rapprochée. Une fidélité récompensée par de nombreux avantages lorsqu'il deviendra adjoint au chef de cabinet du Président, notamment un logement de fonction à deux pas de la tour Eiffel, selon le journal Le Monde. Il avait également obtenu une voiture de fonction équipée d'un gyrophare et une autorisation de port d'arme.

c'est quand même surtout que ça a été confirmé par le Directeur général de la Police Nationale Eric Morvan devant la commission d'enquête du Sénat :

LE FIGARO • fr

25/07/2018 Le récap' des auditions de ce matin

Eric Morvan, directeur général de la police nationale (DGPN)

 Eric Morvan a déclaré que la voiture de service de Benalla, équipée d'un gyrophare, ne provenait pas du ministère de l'Intérieur et «n'a pas davantage été équipée par les garages du SDLP».

Et quand bien même cela aurait été faux, on voit mal la gravité de la chose pour quelqu'un qui a port d'armes, badges d'accès à l'hémicycle de l'Assemblée et passeport diplomatique...

Et ce genre de bévue a des conséquences...



Quant au "Benalla s'appellerait en fait Lahcene Benalla" (dont on voit mal encore la gravité) "sans présence médiatique", comment dire (source) :

## Affaire Benalla : « Il rêvait d'être indispensable aux stars »

🝙 > Faits divers | Jean-Marc Ducos, envoyé spécial dans l'Eure, avec J.M. Dé | 21 juillet 2018, 20h29 | MAJ : 21 juillet 2018, 22h25 | f 🔰 🧠 88

Originaire du Maroc, lui qui aurait modifié son prénom pour le franciser, n'a pas laissé le souvenir du solide gaillard aux épaules larges qu'on lui connaît après les images de l'agression commise sur un manifestant le 1er mai dernier à Paris sur la place de la Contrescarpe.

Et pour le recueil de l'ADN des opposants politiques (qui sonne quand même plus Gorafi) "sans présence médiatique", cela a été entendu sur France Info :



Bref, si on enlève le gyrophare et le prénom sans intérêt, on arrive à 15 000 soit 0,3 % du total (avec une portée forcément encore plus limitée) – bref, le phénomène est imperceptible...

exacte (sic) ».

Conclusion: Nous pouvons dire que si le nombre de tweets sur l'affaire Benalla est le même à travers les communautés hyperactives, celui de la communauté de la France Insoumise contient drastiquement moins de désinformation

S'en suit alors une partie "Police judiciaire", s'intéressant à certains comptes "suspicieux"

:

Premièrement, nous avons réussi à isoler un écosystème suspicieux présent dans l'entièreté de nos bases de données (Macron leaks, désinformations russes, rumeurs) pendant la présidentielle :

- StopMariageHomo
- Patricia691503
- ThierryRouby
- Nouveaux\_poppys
- Drazelrapido
- Patty69
- 20RueDuCirque
- Azi\_azip
- Paoegilles
- Patriosous

Ces 11 profils ont publié sur 13 mois pas moins de 531 000 tweets soit en moyenne plus de 120 tweets par jour et par profil, principalement pour parler de Trump, des migrants ou de Macron :

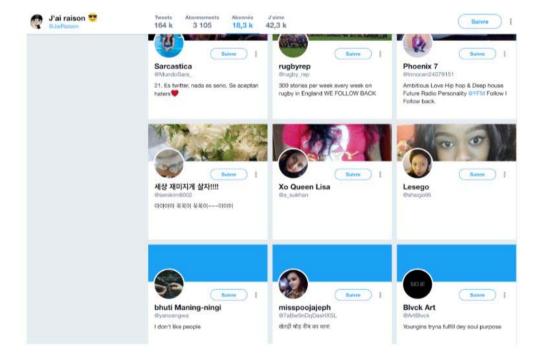
Là encore, voici l'audience de certains de ces comptes à 120 (re)tweets par jour :

| Tweet 53,7   |                 |                  |                  | me<br>55    |
|--------------|-----------------|------------------|------------------|-------------|
| Tweet        |                 | 100 000 PM       |                  | ime<br>6 k  |
| Tweets 136 k | Abonnements 975 | Abonnés<br>1 442 | J'aime<br>81,7 k | Listes<br>9 |

Tweets Abonnements Abonnés J'aime Listes 67,3 k 2 349 1 678 12,1 k 2

On voit le danger... En plus, ils achètent des fans les affreux ! (notez la solidité scientifique de l'analyse) :

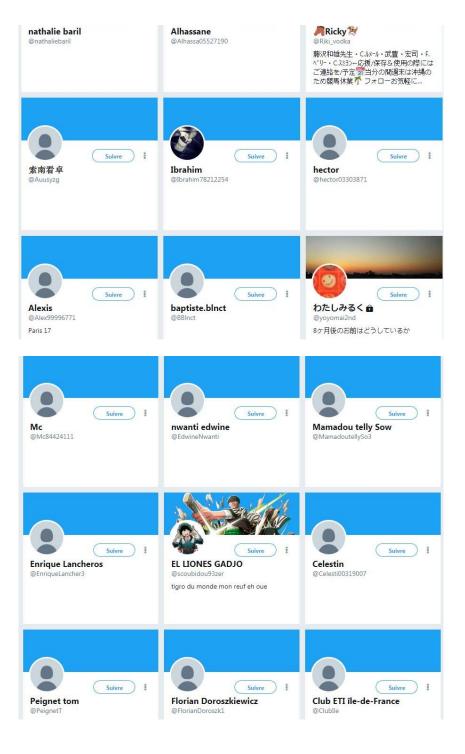
En outre, ils ont réalisé de l'achat de fans :



Soulignons que, fort heureusement, ceci n'arrive jamais sur les comptes sérieux...



https://www.les-crises.fr/l-etude-eu-disinfolab-de-l-influence-russophile-sur-le-benallagate-d-enormes...



Quelques nouveaux abonnés du compte Twitter d'Emmanuel Macron...

### Bref, le jugement est clair :



Nous concluons que ces trois comptes forment donc un réseau automatisé qui utilise la même application enregistrée sous des noms différents. Nous avons ici affaire à des gonflages artificiels de volume puisque les comptes proposent exactement les mêmes retweets.

OK, bravo, bien joué! Mais bon, quel est le problème qu'ils automatisent 3 comptes pour diffuser leur idées? En quoi est-ce un "gonflage artificiel" de volume? C'est quoi un "gonflage naturel" de volume? Twitter est un réseau, le but de chaque personne ayant un message à diffuser est de toucher le plus de personnes possible pour les informer... Le JT de 20h00 ou les journaux des milliardaires participent-ils à des "gonflages artificiels" de volume d'information (par rapport au fait naturel de parler dans un bistro?)?

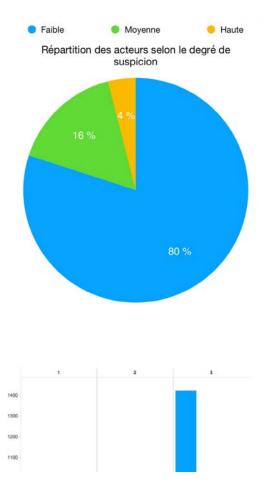
Enfin arrive la conclusion – où les auteurs définissent (allez savoir comment... #Pifomètre) – un "degré de suspicion" de désinformation :

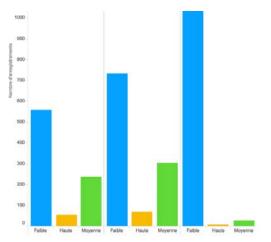
Sur base de l'ensemble de ces informations, et en regardant chacun des profils un à un, nous avons défini un degré de suspicion sur l'activité d'un utilisateur comme acteur récurrent de désinformation :

Hauta Movenne Faible



Donc oui, on risque d'être un compte qui désinforme si on a partagé plus de 5 des 9 âneries précédentes... Mais cela ne concerne que moins de 700 comptes (sur 55 000 relativement actifs, soit 1 %...)





## L'étude se termine sur un nouveau coup sur les "russophiles" :

Nous pouvons également constater d'autres corrélations avec les profils à suspicion moyenne et haute :

- 50 % des profils avaient également propagé des rumeurs durant la présidentielle française. Seuls 4% des comptes avaient propagé dès les premiers instants les Macron leaks.
- 43,6 % des personnes étant présentes dans les comptes partageant le plus Russia Today et Sputnik, et qui ont un item supplémentaire (désinformation russe, Macron Leaks, rumeurs présidentielles) sont présents dans cet échantillon de suspicion moyenne ou haute.

alors qu'à l'évidence ce critère n'est qu'indirect, et que la variable explicative est probablement principalement d'être un compte d'extrême-droite malhonnête...

Mais en conclusion de cette analyse, nous citerons EU DisinfoLab (source) :





En réponse à @Vergnaud

Il y a 5 conclusions au-delà de cela : disinfo.eu/2018/08/08/08- ...

On attendrait quand meme un peu plus de rigueur de votre part.

08:08 - 8 août 2018

## V. Partie 1 – Enseignements

L'étude commence en fait ainsi :

Les personnes hyperactives qui ont publié 1 % des 47 % de tweets sont séparées en deux parties :

- D'une part, des personnes militantes qui publient énormément, ceci sans le moindre but de désinformation. Ce segment est clairement occupé par des comptes proches de la France insoumise
- D'autre part, des personnes qui partagent de la désinformation et que nous pouvons corréler à à l'écosystème autour de

  Puscia Todaviot Sputalis C'ast coux d'aut pous ent.

particulièrement intéressés et qui ont fait l'objet d'une attention plus soutenue.

Elle dresse 5 conclusions:

#### "1. Un volume d'activité extraordinaire" :

"en sélectionnant les plus actifs ayant publié au moins 1000 tweets, on obtient même 445 acteurs qui ont publié 768 281 tweets." Soit 250 (re) tweets par jour : quelle en est l'influence réelle à ce stade ?

### "2. Des hyperactifs avec des caractéristiques hors-norme" :

"Un haut taux de tweets : Les individus qui ont propagé beaucoup de désinformation ont publié en moyenne 173 318 tweets contre 46 000 tweets pour les comptes qui ne diffusent presque pas de désinformation. Ce nombre est en relation directe avec la probabilité qu'un individu diffuse une désinformation."

Ils ne se rendent même pas compte que, en plus de ne sciemment pas avoir mesuré l'influence, ils sont passés à côté d'un autre problème : ils n'ont pas mesuré le "taux de désinformation". Bref, pour eux, il est plus grave d'avoir partagé 5 désinformations sur 1000 (re)tweets à 1 000 abonnés que 4 désinformations sur 15 tweets à 20 000 abonnés... Ils l'affichent eux-mêmes (source) :





7 désinformations sur 10 000 (re) tweets pour un compte de 1 500 abonnés...

Question : Pourquoi ne pas pondérer la "désinformation" par l'activité et l'audience ?

"Habitant nulle part : Un autre fait intéressant est le fait que 41,9 % de ces comptes n'ont aucune information à propos de leur localisation".

On appellerait ça un compte anonyme?

### "3. Un écosystème relié à d'autres réseaux identifiés précédemment";

Parmi ces hyperactifs, on retrouve différents réseaux identifiés pendant la présidentielle française :

- 27 % font partie de l'écosystème qui relayait les médias
   Russia Today et Sputnik en février et mars 2017 (pendant la
   campagne présidentielle française de 2017 et avant la diffusion
   de la chaîne RT en France qui date de décembre 2017).
- 9,43 % faisaient partie des 840 comptes ayant partagé au moins trois désinformations durant la présidentielle qui ont été identifiées comme telles par Crosscheck, un projet de FirstDraft, organisme référent de fact-checking.
- 2,8 % faisaient partie des quelque 1653 acteurs qui ont amplifié les #Macronleaks 2 à 3 heures après leur publication.

Là encore, l'étude cherche à confirmer ses propres biais et obsession- mention spéciale pour les 3 % de comptes qui ont parlé des MacronLeaks... Là encore aucune analyse sur l'écosystème LREM, ou sur les écosystèmes américanophile ou europhile...

### "4. Comptes semi-automatisés"

"Nous avons identifié un réseau de trois personnes utilisant un système automatique de retweet. Ces trois comptes passent par un service web utilisant l'API Twitter afin de publier toute la journée et ont vraisemblablement acheté des followers."



## "5. Une corrélation de désinformation"

Il est impossible de déterminer si une vraie personne se trouve derrière chaque compte; et encore moins d'identifier si un compte est est à la solde d'un réseau de désinformation étranger. Ce constat est sans équivoque.

Cependant, force est de constater au vu des différents cas de désinformation que le taux de russophilie d'un écosystème est corrélé avec sa probabilité de partages de désinformation.

Corrélation n'est pas causalité, dans le sens où cela ne voudra jamais dire que l'on peut prouver une influence étrangère ou tout autre causalité. Par contre, cela peut constituer un bon indicateur pour identifier s'il y a davantage de probabilité de circulation de désinformations.

### #OnNePeutPasLeProuver...

Par ailleurs on ne savait pas que c'était mal d'automatiser un compte Twitter – on ne nous avait pas dit que c'était réservé aux seuls grands médias a priori...

Quant aux "comptes à la solde d'un réseau de désinformation étranger"...



Mais loin de nous l'idée de nier le fait que la Russie a une stratégie d'influence en Occident. Simplement, force est de constater que primo, on lui attribue beaucoup de choses sans preuves solides (voire de manière de plus en plus ridicule, ce qui prend un tour réellement parodique) et que, secundo, quoi que fasse la Russie, cela restera en termes d'efficacité d'influence sur notre pays sans commune mesure avec les interventions permanentes du *soft power* des États-Unis, sujet tabou qui n'est presque jamais abordé.

Et on voit que, face à cette menace insignifiante, de plus en plus d'individus travaillent à la limitation de Libertés fondamentales (loi sur les *fake news*, suivi et fichage des internautes, sanctions et contre sanctions...)

Rappelons également cette vision du professeur Kevin Limonier :





## VI. Parties 4 et 5 – Le EU DisinfoLab et FAQ

Dans cette partie, l'officine donne plus d'informations sur elle par transparence :

Le EU DisinfoLab est une association à but non lucratif (A.S.B.L) enregistrée en décembre 2017 en Belgique. Sa mission est de lutter contre la désinformation par une méthodologie innovante et un soutien scientifique à la communauté de la contredésinformation.

Méthodologie "innovante" que nous venons de voir à l'œuvre. Quant au "soutien scientifique", ce serait bien qu'ils aient de vrais scientifiques...

#### Financement:

Twitter: à la suite des manipulations survenues pendant la campagne américaine de 2016, et qui impliquait l'achat de publicités par Russia Today et Sputnik, Twitter s'est engagé le 26 octobre 2017\*(6) à redistribuer les 1,9 millions de dollars vers des organisations luttant contre la désinformation. À ce titre, le EU DisinfoLab a requ 125 000 \$ en janvier 2018.

Ok – ils ont donc touché 125 000 \$ de Twitter – qui viennent des 1,9 M\$ payés par RT et Sputnik à Twitter (soit 7 % de la somme), que Twitter a choisi de reverser "à des organisations luttant contre la désinformation". On se rend que le trio qui a fondé leur agence de com a eu le nez creux de justement créer une officine présentée comme "luttant contre la désinformation"...

Open Society Foundations: dans le cadre d'un projet conçu par l'équipe du EU DisinfoLab au mois de septembre 2017, le EU DisinfoLab a reçu 25 000\$ pour observer les phénomènes de désinformation pendant la campagne législative italienne de mars 2018. Le rapport de cette étude est disponible sur le site du EU DisinfoLab.

Nous ne sommes pas une agence de l'Union européenne.

Vous pouvez en apprécier la teneur ici (on serait Soros, on demanderait cependant un sérieux rabais sur les 25 000 \$) – surprise, ils parlent encore de "l'écosystème russe" :

Following these promising results, we wanted to test the reversal of the methodology. If we were able to find a correlation between disinformation spreading and the Russian ecosystem, would it work to firstly monitor the Russian ecosystem in order to detect disinformation in real time? Furthermore, we wanted to test if this monitoring technique would contribute to raising awareness amongst the population about the potential issue.

mais au moins elle est signée et transparente sur le financement :

#### Authors:

Alaphilippe, Alexandre Ceccarelli, Chiara Charlet, Léa Mycielski, Martin

First published: 01/06/2018 in Brussels

Project funded by Open Society Foundations

Ah, il y a eu aussi 25 000 \$ de la fondation Soros pour suivre les élections italiennes – on imagine qu'il n'y a aucune structure en Italie avec les compétences de cette officine...

## Comment vérifier vos conclusions?

Nous mettons à disposition des journalistes accrédités et

d'organismes de recherche les datasets suivants :

- Le classement des plus actifs qui attestent du fait que 1 % des auteurs a publié 47 % des tweets
- Le fichier avec le nombre de rumeurs par individus classés dans les plus actifs

Télécharger

Comme on l'a vu dans cet autre billet, ils ont mis le fichier des comptes à disposition...

## Et ils terminent par :

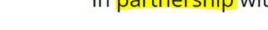
## Disinfolab.com et EUVSDisinfo

Nous n'avons aucun lien avec DisinfoLab.com.

EUVSDisinfo est un compte Twitter géré par le service East Stratcom du Service Européen d'Action Extérieure (Commission européenne). Nous avons rencontré lors d'événements certains de leurs membres, mais n'avons aucun lien de collaboration.

Bon, c'est juste bête que ceci figure en bas de leur site :























Annexe : On lire avec intérêt cette étude complémentaire de Damien Liccia sur Médium, et cet article de synthèse plein de bon sens des Décodeurs du Monde (il faudrait juste arrêter d'appeler cette officine "ONG" qui a un sens différent en français).

Twitter et Soros ne nous donnant pas 150 000 euros, nous rappelons que ce site ne vit que par vos dons. **Vous pouvez nous soutenir ici. Merci d'avance!** 

Dossier: #FichagePolitique - EU Disinfo Lab

#3 #4 #5

Tous fichés?
Comment Twitter
permet le fichage
politique de la

L'étude de l'influence russophile sur le Benallagate : d'énormes failles Le EU DisinfoLab à l'origine d'une intox : l'affaire Benalla, produit de la «

https://www.les-crises.fr/l-etude-eu-disinfolab-de-l-influence-russophile-sur-le-benallagate-d-enormes...

» L'étude de l'influence russophile sur le Benallagate : d'énormes failles méthodologiques

population méthodologiques russosphère » ?

LIBERTÉS PUBLIQUES 9.août.2018 // Les Crises

## Tous fichés ? Comment Twitter permet le fichage politique de la population



Libertés publiques, Propagande

Nous avons largement démonté hier l'étude préliminaire, puis aujourd'hui l'étude finale comportant d'énormes failles méthodologiques du EU DisinfoLab sur l'affaire Benalla.

Beaucoup de personnes ont parlé de cette étude depuis plusieurs jours, discutant son bien-fondé ou non, mais bien peu se sont arrêtés sur un fait extrêmement problématique pour les libertés publiques.

En effet, l'EU DisinfoLab a, par le biais de Nicolas Vanderbiest, réalisé une étude sur les tweets liés à l'affaire Benalla. Nous laisserons aujourd'hui les conclusions de côté pour nous intéresser à la question **des outils utilisés**.

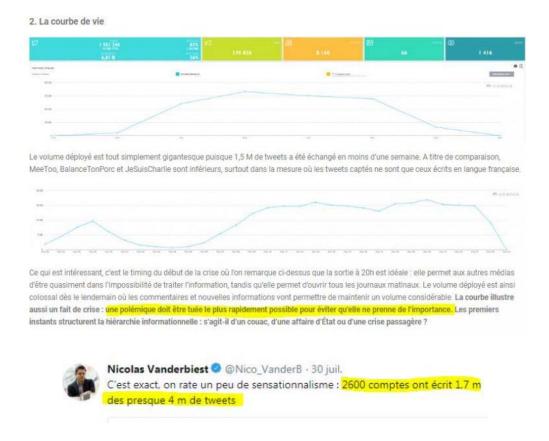
- I. Visibrain, ou Big Brother à la maison
- II. Que prévoient les conditions d'utilisation Twitter?
- III. Quand EU DisinfoLab joue avec Big Brother
- IV. "Et c'est le drame..." : la diffusion des données personnelles
- V. Le fichage politique (mais pas que) par EU DisinfoLab
- VI. Communication de EU DisinfoLab

1 of 73  $12/08/2018 \ 10:45$ 

- VII. Position de EU DisinfoLab quant aux données sensibles
- VIII. Le CNRS aurait aussi fiché les opinions politiques de près de 200 000 personnes !
- IX. Réglementation et Discussion
- X. Plainte

## I. Visibrain, ou Big Brother à la maison

Comme on l'a vu, dès le 23 juillet, Nicolas Vanderbiest a sorti une première analyse sur son blog – plutôt orientée communication : Affaire Benalla sur les réseaux sociaux : où la résurrection des partis de l'opposition (archive). On y voit quelques chiffres, qui rejoignent ses tweets :



Plus précisément on voit ici qu'il s'intéresse à 4 567 426 tweets, venant de 247 701

2 of 73 12/08/2018 10:45

### utilisateurs:



Il utilise pour ce faire un des outils d'analyse de Twitter – il s'agit ici du logiciel Visibrain , qui sert normalement à des analyses marketing pour les grandes entreprises :



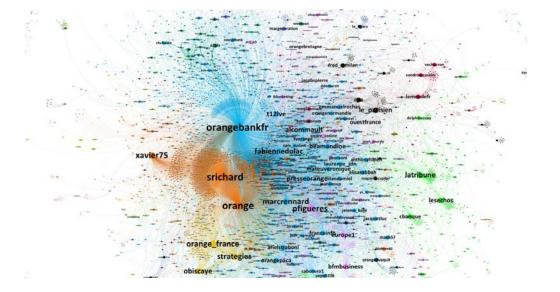


 $3 ext{ of } 73$  12/08/2018 10:45

et permet de réaliser en fait automatiquement les graphiques que présente Nicolas Vanderniest :

## visibrain Influenceurs Cartographie

La plateforme Visibrain possède un format d'export unique sur le marché : l'export GEXF, le format de données du logiciel de cartographie open-source Gephi.
En combinant la puissance de Gephi et Visibrain, vous pouvez réaliser en quelques clics des cartographies complètes des relations entre utilisateurs, en particulier sur des sujets à forte volumétrie.
Grâce aux cartographies Visibrain + Gephi, vous identifiez les différentes communautés autour d'un sujet et comprenez les relations d'influence.



En fait, ce logiciel (dont nous allons détailler les caractéristiques, car c'est très important) donne accès... à tous les tweets de Twitter, en temps réel (sources : 1, 2, 3) :

# visibrain PLATEFORME > FONCTIONNALITÉS 100% des tweets avec le firehose Twitter

Partenaire de Twitter depuis 2011, Visibrain dispose d'un accès privilégié aux données Twitter au travers de leur API professionnelle (anciennement connue sous le nom de Gnip).

Visibrain accède ainsi au firehose l'ensemble des 500 millions de tweets publics publiés chaque jour — de manière fiable, exhaustive, et temps-réel.

## ◆ VISIBRAIN PLATEFORME > FONCTIONNALITÉS Veille des réseaux sociaux

La plateforme de veille Visibrain vous permet de veiller l'ensemble des réseaux sociaux et plus particulièrement :

- **Twitter**: Visibrain dispose d'un accès au *firehose* qui garantit la récupération de 100% des tweets publics, en temps-réel
- Facebook : Visibrain dispose d'un module dédié à la veille des pages publiques Facebook

 Instagram & YouTube: Visibrain permet également de veiller les réseaux sociaux visuels comme Instagram ou YouTube.



Le temps-réel est dans l'ADN de la plateforme Visibrain :

- dès qu'un contenu est publié, il est immédiatement analysé et disponible dans Visibrain, ce qui vous permet d'avoir en permanence des données fraîches
- le système d'alerte de Visibrain vous garantit une notification en moins d'une minute
- l'interface de la plateforme s'actualise instantanément

On peut largement filtrer, en particulier par "métiers et centres d'intérêts" (source) :

visibrain

PLATEFORME > FONCTIONNALITÉS

Système de filtres

La plateforme Visibrain dispose d'un système de filtrage reconnu comme le plus performant du marché.

Plus de 100 filtres et options de tri sont disponibles :

• filtrage sur le contenu : mots-clés,

langues, expressions-exactes, filtrage partiel, sensibilité à la casse, titre de l'article...

- filtrage sur la démographie : localisation, métiers, centres d'intérêt, utilisateurs spécifiques, sources spécifiques...
- filtrage sur la temporalité: granularité à l'année, au mois, à la semaine, au jour ou à la minute
- le tout combinable à l'infini avec l'ensemble des opérateurs booléens ET, OU, EXCLUSION, et PARENTHESES

On compte explicitement dans les Métiers "journaliste" ou... "militants politiques" (source) .

ANALYSE DÉMOGRAPHIQUE Métiers

O VISIBRAIN

La plateforme Visibrain infère le métier des utilisateurs des réseaux sociaux grâce à différents indices, ce qui vous permet de voir en un clin d'oeil quelle typologie de population (journalistes, étudiants, militants politiques...) concerne votre sujet.

Comme tous les indicateurs de Visibrain, il est possible de cliquer sur un métier pour avoir une analyse ciblée sur cette population.

journaliste en writer rédacteur en chef animateur journaliste musician rédactrice en chef alliancy\_lemag avocat journaliste altereco\_ etudiant

 $7 ext{ of } 73$  12/08/2018 10:45

responsable du poie photographe directeur de la bloqueur étudiant student militant rédactrice en chef rédacteur en chef adjoint directeur numérique de **auteur** consultant juriste journalist figurations

Ah, quel est le métier le plus visé...?

De même pour les centres d'intérêts (source) :

# ANALYSE DÉMOGRAPHIQUE Centres d'intérêts

La plateforme Visibrain vous affiche une synthèse des centres d'intérêt des utilisateurs qui ont posté sur votre sujet de veille. Cela vous permet de mieux comprendre quelle est la population qui intéragit avec votre marque (ou n'importe quel sujet de veille), et de découvrir de nouvelles opportunités de communication.

Comme tous les indicateurs de Visibrain, il est possible de cliquer sur un centre d'intérêt pour avoir une analyse ciblée sur cette population.

Comme on n'arrête pas le progrès, on peut même analyser les émojis (source) :

Emojis

La plateforme Visibrain fait ressortir les emojis les plus utilisés dans les conversations sociales.

Vous identifiez ainsi en un clin d'oeil la tonalité d'un sujet.

 $8 ext{ of } 73$  12/08/2018 10:45

Comme tous les indicateurs de Visibrain, il est possible de cliquer sur n'importe quel emoji pour comprendre quels posts sociaux se cachent derrière.



On peut aussi chercher dans les biographies – en particulier pour trouver les "journalistes, militants et activistes" (source) :

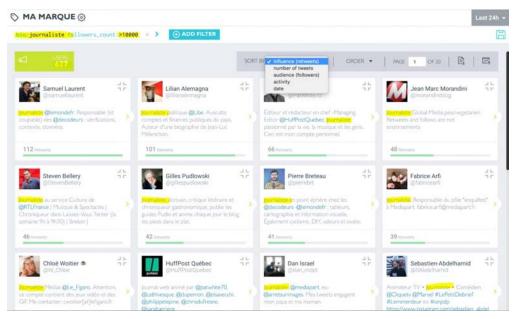
# PLATEFORME > FONCTIONNALITÉS Détection d'influenceurs Twitter

Grâce à Visibrain, vous détectez de nouveaux influenceurs sur vos thématiques, et vous identifiez avec précision les comptes Twitter les plus influents sur un sujet donné. Visibrain vous permet de croiser plusieurs options de tri et de filtrages pour obtenir des listes d'influenceurs sur-mesure :

• Tri par audience, influence ou activité : pour classer les comptes Twitter en

9 of 73

- impact en terme d'engagement, ou leur nombre de tweets sur un sujet donné
- Recherche dans les biographies: pour identifier les personnes qui se déclarent comme "journaliste", "militant", "activiste", "blogueur", "youtubeur"...
- Localisation: pour cibler une zone géographique précise (ville, région ou pays)
- Informations de profils: pour filtrer les comptes qui correspondent à certains critères (compte vérifié, ratio following/followers, date de création du compte...)



Oh des journalistes avec plus de 10 000 abonnés...

10 of 73 12/08/2018 10:45

On pourra alors en faire des panels pour les suivre de très près (source) :

## O VISIBRAIN PLATEFORME > INFLUENCEURS Panel d'influenceurs

Visibrain vous permet d'aller au delà de la veille par mots-clés en mettant en place une écoute d'une liste d'influenceurs (ou panel d'influenceurs).

Nous vous aidons à constituer une liste de

comptes Twitter référents sur une thématique donnée (par exemple 200 comptes de blogueurs, journalistes et passionnés d'automobile), et captons l'ensemble de leur activité (tweets, retweets, articles partagés, photos et vidéos partagées).

De même, le logiciel déterminera généralement votre localisation et votre genre (sources : 1 et 2):

## O VISIBRAIN ANALYSE DÉMOGRAPHIQUE Localisation

La plateforme Visibrain infère la localisation des utilisateurs des réseaux sociaux grâce à différents indices (informations de profil, contenu posté...), et délivre ainsi une information géographique sur plus de 60% des comptes!

Vous pouvez visualiser la propagation d'un sujet sur une carte, et filtrer les

12/08/2018 10:45 11 of 73

### informations par pays, regions ou ville.



La plateforme Visibrain infère le genre des utilisateurs des réseaux sociaux grâce à différents indices (prénom, éléments du profil, contenu posté...), et délivre ainsi une information du genre (homme/femme) de l'utilisateur sur plus de 70% des comptes!

Ce logiciel Big Brother vous facilite la vie en vous permettant d'exporter toutes les données qui vous intéressent sous Excel (source) :



La plateforme Visibrain vous permet d'exporter la totalité des indicateurs de la plateforme au format CSV (Excel).

Les exports Visibrain sont complets, propres et directement exploitables pour retraiter les données ou les inclure dans un rapport.

|    | A            | В           | C        | D             | E                 | F              | G    |       | Н        | 1            | J       |   | K            |         | L | M          |
|----|--------------|-------------|----------|---------------|-------------------|----------------|------|-------|----------|--------------|---------|---|--------------|---------|---|------------|
| 1  | screenNa 🔻   | followers 🔻 | verified | bio 🔻         | first seen date ▼ | client =       | lang | link  | ♥        | text 🔻       | country | V | region       | city    |   | retweets 🔻 |
| 5  | visibrain    | 8989        | TRUE     | La plateform  | 24/10/2017 16:18  | Twitter Web    | fr   | https | ://twitt | Comment ex   | france  | - | île-de-franc | e paris |   | 12         |
| 6  | CCIcotedazui | 13575       | TRUE     | RETROUVEZ     | 26/10/2017 07:09  | Twitter for A  | fr   | https | ://twitt | ©CClcot      | france  |   | provence-al  | r nice  |   | 11         |
| 7  | visibrain    | 8993        | TRUE     | La plateform  | 30/10/2017 14:36  | Twitter Web    | fr   | https | ://twitt | Quand peut-  | france  |   | île-de-franc | e paris |   | 10         |
| 8  | fadouce      | 29999       | FALSE    | Com, stratég  | 30/10/2017 07:35  | TweetDeck      | fr   | https | ://twitt | L'étonnante  | france  |   | île-de-franc | e paris |   | 9          |
| 9  | CabonVro1    | 408         | FALSE    | La Fibre Oran | 03/11/2017 08:52  | Twitter for il | Ffr  | https | ://twitt | Les @amb_c   | france  |   | île-de-franc | e paris |   | 8          |
| 10 | visibrain    | 8989        | TRUE     | La plateform  | 24/10/2017 16:04  | Twitter Web    | fr   | https | ://twitt | #LivreBlanc: | france  |   | île-de-franc | e paris |   | 6          |
| 11 | blendwebmi   | 6222        | FALSE    | #BlendWebN    | 26/10/2017 14:45  | Twitter Web    | fr   | https | ://twitt | Déjà 3169 tv | france  |   | rhône-alpes  | lyon    |   | 6          |
| 12 | visibrain    | 8989        | TRUE     | La plateform  | 24/10/2017 16:04  | Twitter Web    | fr   | https | ://twitt | Téléchargez  | france  |   | île-de-franc | e paris |   | 6          |
| 13 | blendwebmi   | 6238        | FALSE    | #BlendWebN    | 26/10/2017 18:23  | Twitter Web    | fr   | https | ://twitt | Les 5000 me  | france  |   | rhône-alpes  | lyon    |   | 5          |
| 14 | visibrain    | 8993        | TRUE     | La plateform  | 30/10/2017 14:36  | Twitter Web    | fr   | https | ://twitt | Qu'est-ce-qu | france  |   | île-de-franc | e paris |   | 5          |
| 15 | visibrain    | 8989        | TRUE     | La plateform  | 27/10/2017 14:40  | Buffer         | fr   | https | ://twitt | [BAD BUZZ]   | france  |   | île-de-franc | e paris |   | 4          |

| 16 | marc_cogrel  | 2661  | FALSE | Responsable  | 31/10/2017 17:25 | Twitter Web    | fr | https://twitt | Voici le top # france   | île-de-france  | paris           | 4 |
|----|--------------|-------|-------|--------------|------------------|----------------|----|---------------|-------------------------|----------------|-----------------|---|
| 17 | visibrain    | 8998  | TRUE  | La plateform | 25/10/2017 11:53 | Twitter Web    | fr | https://twitt | Comment #Facebook dé    | ermine les pul | olications à af | 4 |
| 18 | visibrain    | 8992  | TRUE  | La plateform | 31/10/2017 13:51 | TweetDeck      | fr | https://twitt | Les twittos si france   | île-de-france  | paris           | 4 |
| 19 | AgenceADeq   | 8976  | FALSE | Développem   | 30/10/2017 09:12 | Twitter Web    | fr | https://twitt | "Les 580 twe france     | midi-pyréné    | toulouse        | 4 |
| 20 | agence636    | 1026  | FALSE | WE V #SOC    | 30/10/2017 14:04 | Twitter Web    | fr | https://twitt | #Apple twee france      | midi-pyréné    | toulouse        | 4 |
| 21 | visibrain    | 8992  | TRUE  | La plateform | 31/10/2017 11:00 | Twitter Web    | fr | https://twitt | 10 comptes v france     | île-de-france  | paris           | 4 |
| 22 | suebecks     | 10414 | FALSE | National Tea | 07/11/2017 11:38 | Twitter Web    | en | https://twitt | Gaining Pow united king | england        | sheffield       | 3 |
| 23 | Alexleserveu | 6019  | TRUE  | #SocialMedia | 31/10/2017 11:04 | TweetDeck      | fr | https://twitt | Quand je voi france     | île-de-france  | paris           | 3 |
| 24 | GuillaumePe  | 112   | FALSE | Chargé de dé | 03/11/2017 11:25 | Twitter Web    | fr | https://twitt | @AmDalmas france        | île-de-france  | paris           | 3 |
| 25 | LettreAudio  | 4479  | FALSE | Premier jour | 27/10/2017 16:25 | Twitter for if | fr | https://twitt | https://t.co/: france   | île-de-france  | paris           | 3 |
| 26 | visibrain    | 8993  | TRUE  | La plateform | 05/11/2017 18:08 | Buffer         | fr | https://twitt | A quel mome france      | île-de-france  | paris           | 3 |

Vous pouvez aussi exporter tous les followers, en les croisant (source) :

## PLATEFORME > INFLUENCEURS Extraction de followers

Avec Visibrain, vous pouvez exporter

l'exhaustivité des followers de n'importe quel

compte au format Excel : les followers de votre

marque, ceux de vos concurrents, ceux d'un

influenceur...

Visibrain vous permet également d'effectuer

des croisements de liste de followers pour

identifier quels followers vous avez en

commun avec vos concurrents, les comptes

qui suivent vos concurrents sans vous suivre,

ou bien les comptes qui suivent l'ensemble

des influenceurs de votre secteur.

| В                 | E   | F                      | G                  | H              | 1              | 1           |       |
|-------------------|---|------------------------|--------------------|----------------|----------------|-------------|-------|
| Nom d'utilisateur | ▼ Biographie ▼                                    | Site Web 🔻             | Date de création 🔻 | Nb followers - | Nb following - | Nb listed - | Nombr |
| C8TV              | ps émissions : #TPMP #WAM #SLT #LTD #Fami         | http://www.c8.fr       | 20/05/2009 09:53   | 767064         | 1469           | 983         |       |
| TopAchat          | icier du #Geek depuis 1999 ! Grosse comman        | http://www.topachat.   | 19/01/2010 09:42   | 410177         | 9393           | 647         |       |
| bouyguestelecom   | uygues Telecom partage votre passion de la t      | http://www.bouygues    | 02/10/2008 12:02   | 143808         | 65237          | 790         |       |
| SNCF              | ut l'univers SNCF à portée de clic : les infos en | http://www.sncf.com    | 22/11/2010 23:53   | 108644         | 957            | 725         |       |
| Le_Figaro         | Actualités, économie, politique, culture et div   | http://www.lefigaro.fr | 22/08/2007 09:16   | 2800055        | 537            | 14510       |       |
| Europe1           | ute l'information d'@Europe1 et nos émission      | http://www.europe1.f   | 23/04/2009 10:53   | 1311255        | 3542           | 5271        |       |
| 20Minutes         | formation en continu. Actualités, Politique, Sp   | http://www.20minute    | 28/03/2007 14:14   | 2308909        | 1469           | 9180        |       |
| LesEchos          | ctualité économique, financière et boursière      | http://www.lesechos.   | 25/11/2009 18:01   | 1066877        | 500            | 7197        |       |
| LCI               | envenue sur LCI, sur le canal 26 de la TNT. Le r  | http://lci.fr/         | 23/03/2009 23:45   | 109113         | 964            | 1457        |       |
| BFMTV             | re chaîne info de France. Retrouvez-nous sur      | http://www.bfmtv.cor   | 16/04/2010 11:55   | 2267084        | 488            | 8310        |       |
| OuestFrance       | mpte Twitter du quotidien Ouest-France. Rec       | http://www.ouest-frag  | 27/04/2009 17:59   | 518358         | 3518           | 3124        |       |
| romandie          | mpte officiel de http://Romandie.com, média       | http://www.romandie    | 27/03/2009 13:43   | 25117          | 2              | 591         |       |
| Niska_Officiel    | puvel album #Commando maintenant disponil         | http://www.charo-shc   | 02/11/2011 01:37   | 383839         | 36             | 247         |       |
| lalibrebe         | mpte officiel de Lalibre.be Rejoignez-nous ég:    | http://www.lalibre.be  | 04/02/2009 11:37   | 203785         | 2896           | 2155        |       |
| lemondefr         | ctualité de référence par la rédaction du Mor     | http://www.lemonde.    | 16/03/2009 19:44   | 7511211        | 504            | 33063       |       |
| AmazonHelp        | e answer Amazon support questions 7 days a        | http://www.amazon.c    | 28/10/2009 05:17   | 124674         | 52             | 741         |       |
| Orange_conseil    | sez vos questions du lundi au samedi, de 8h à     | 20h dans les domaines  | 13/10/2010 11:16   | 88857          | 44182          | 431         |       |
| afpfr             | orceaux choisis des fils d'actu, coulisses & RT ( | http://www.afp.com     | 20/07/2011 14:07   | 2855687        | 1348           | 12990       |       |
| AmazonFR          | mpte Officiel Twitter http://www.Amazon.fr.       | http://amazon.fr       | 30/09/2015 14:07   | 21722          | 1              | 61          |       |
|                   |   | and the second second  | 727 222 222 222    |                | 0.2            |             |       |

O VISIBRAIN

(sur demande).

| Z.3 | dc311_dcresi2 | pidents & entraves sur le reseau autoroutier denttp://quebecbilintq    | 31/03/2014 15:37 | 3001    | 15    | 29   |    |
|-----|---------------|--|------------------|---------|-------|------|----|
| 24  | Sosh_fr       | sh, la marque sans engagement par Orange, r http://www.sosh.fr/        | 19/07/2011 11:35 | 138831  | 29549 | 578  | Ų. |
| 25  | tvanouvelles  | a télé, sur votre ordinateur ou votre appareil http://tvanouvelles.ca  | 26/10/2009 15:59 | 343378  | 538   | 2174 |    |
| 26  | PSchydlowski  | rtains mettent à jour leur statut @facebook toutes les 5 secondes, ils | 09/01/2011 11:42 | 41003   | 20691 | 280  |    |
| 27  | le_Parisien   | Parisien-Aujourd'hui en France, 1er quotidien http://www.leparisien    | 06/05/2009 10:50 | 1958055 | 910   | 7770 |    |

Et, si vous le souhaitez (et payez), vous pouvez même avoir accès à l'historique Twitter depuis 2006 (source)!



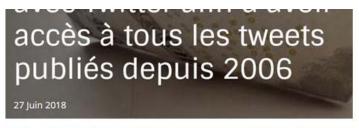
archives complètes de Twitter depuis 2006

Car bien sûr le logiciel est payant – comptez quelques milliers d'euros par mois pour récolter quelques centaines de milliers de tweets...

Comme vous le savez peut-être, le 25 mai 2018, le règlement européen général sur la protection des données (RGPD) est entré en application pour augmenter la protection des données des citoyens.

Eh bien c'est à peine le 27 juin dernier que Twitter a décidé de donner accès à ses archives à Visibrain (source) :

Nouveauté! Visibrain renforce son partenariat avec Twitter afin d'avoir



Suite à un renforcement de notre partenariat avec Twitter, nous avons pu améliorer significativement notre technologie « instaback » de récupération des tweets dans le passé : la plateforme Visibrain permet désormais de récupérer rétroactivement les données Twitter sur n'importe quel sujet de votre choix depuis mars 2006, jusqu'au tout premier tweet fait par @jack Dorsey.

## II. Que prévoient les conditions d'utilisation Twitter?

À ce stade, vous devez commencer à être effrayé de la masse d'informations mise à disposition – et de ce qu'on peut en faire...

Si vous vous demandez ce que prévoit Twitter à ce sujet, voici une sélection des conditions (source {3}) :



# Vos droits et Concession de droits sur le Contenu

Vous conservez vos droits sur tout Contenu que vous soumettez, publiez ou affichez sur ou via les Services. Ce qui est à vous vous appartient. Vous êtes le propriétaire de votre Contenu (ce qui inclut vos sons, photos et vidéos intégrés).

Bonne nouvelle vous conservez vos droits sur tout votre contenu!

Mais bon...:

En soumettant, en publiant ou en affichant un Contenu sur ou via les Services, vous nous accordez une licence mondiale, non exclusive et libre de redevances (incluant le droit de sous-licencier), nous autorisant à utiliser, copier, reproduire, traiter, adapter, modifier, publier, transmettre, afficher et distribuer ce Contenu sur tout support et selon toute méthode de distribution (actuellement connus ou développés dans le futur). Cette licence nous autorise à mettre votre Contenu à disposition du reste du monde et autorise les autres à en faire de même. Vous convenez que cette licence comprend le droit pour Twitter de fournir, promouvoir et améliorer les Services et de mettre le Contenu soumis sur ou via les Services à disposition d'autres sociétés, organisations ou personnes privées, aux fins de syndication, diffusion, distribution, promotion ou publication de ce Contenu sur d'autres supports et services, sous réserve de respecter nos conditions régissant l'utilisation de ce Contenu. Twitter, ou ces autres sociétés, organisations ou personnes privées, pourront utiliser ainsi le Contenu que vous aurez soumis, publié, transmis ou de quelque autre façon mis à disposition via les Services sans que vous puissiez prétendre à une quelconque rémunération au titre de ce Contenu.

Twitter et ses clients en font ce qu'ils en veulent et vous ne pouvez vous y opposer !

Car tout est public (source):

## Informations publiques

Sur Twitter, la plupart des activités sont publiques, y compris les informations de votre profil, votre fuseau horaire et votre langue, la date de création de votre compte, vos Tweets et certaines informations sur vos Tweets, telles que la date, l'heure. l'application et la version de Twitter à partir

desquelles vous avez Tweeté. Vous pouvez également décider de publier votre emplacement dans vos Tweets ou sur votre profil Twitter. Les listes que vous créez, les personnes que vous suivez et celles qui vous suivent, les Tweets que vous Aimez ou Retweetez sont également des informations publiques. Les diffusions Periscope que vous

### Et donc ils vous conseillent :

Vous êtes responsable de vos propres Tweets et des autres informations que vous fournissez par le biais de nos services, et vous devez réfléchir avec soin à ce que vous décidez de publier ①, surtout s'il s'agit d'informations sensibles. Si vous

### Car en plus de l'utilisation du site Twitter :

En plus de fournir directement vos informations publiques au monde sur Twitter, nous avons également recours à des technologies, telles que les interfaces de programmation d'application (API) et les éléments incorporés, pour mettre ces informations à la disposition de sites Web, d'applications et d'autres utilisateurs, par exemple en affichant des Tweets sur un site Web d'informations ou en analysant ce que disent les gens sur Twitter. En général, nous mettons ce contenu à disposition gratuitement s'il s'agit de quantités limitées et nous faisons payer des droits pour tout accès à grande échelle. Nous avons élaboré des conditions standard régissant la façon dont ces données peuvent être utilisées et

ils utilisent d'autres interfaces (API) pour diffuser en masse les contenus, moyennant rémunération. C'est le cas avec Visibrain.

Twitter insiste (source):

Twitter est public et les Tweets peuvent

recherches par qui que ce soit dans le monde entier. Nous mettons également à votre disposition

d'autant plus que c'est Twitter qui vend l'accès aux API qu'il a développées pour permettre de telles recherches... (source) :

## À propos des API Twitter

Twitter reflète ce qui se passe dans le monde et ce dont les gens parlent en temps réel. Vous pouvez accéder à Twitter via le Web ou sur votre appareil mobile. Pour partager les informations sur Twitter aussi largement que possible, nous fournissons aussi aux entreprises, aux développeurs et aux utilisateurs un accès programmatique aux données Twitter via nos API (interfaces de programmation d'applications). Cet article explique ce que sont les API Twitter et

### Tweets et réponses

Nous mettons les Tweets et réponses publics à la disposition des développeurs, et nous les autorisons à publier des Tweets via notre API. Les développeurs peuvent accéder aux Tweets en recherchant des mots-clés donnés, ou en demandant un échantillon de Tweets de comptes spécifiques.

Ces points de terminaison sont utilisés par des ONG et des organisations comme les Nations Unies pour identifier, comprendre et contrer les fausses informations relatives aux initiatives de santé publique. Par exemple, en Indonésie, des rumeurs persistantes faisaient courir le bruit que les vaccins contenaient du porc ou entraînaient la stérilité. Comprendre la façon dont ces rumeurs ont été lancées et propagées a permis aux Nations Unies d'envoyer une équipe sur le terrain pour dissiper ces mythes, à l'origine de préoccupations importantes dans cette nation à majorité musulmane. De la même façon, nous aidons les chercheurs à repérer les premières indications

de l'université Northeastern a récemment développé une nouvelle technique de suivi de la grippe en utilisant des données Twitter qui prévoient les épidémies jusqu'à six semaines à l'avance, bien avant de nombreux autres modèles, sans toutefois sacrifier la précision.

Donc Twitter indique qu'il vend vos contenus, par exemple "aux ONG", aux "Nations-Unies", par exemple pour lutter contre le "complotisme anti-vaccins" (dans les pays "à majorité musulmane") ou lutter contre "les épidémies" telles "la grippe".

C'est beau...

Bon, ok, dans l'écrasante majorité des cas, ce sera des entreprises telles Boeing pour savoir ce que les twittos disent d'elles (voire des "chercheurs" voulant connaître vos opinions politiques) – mais on imagine qu'il ne faut pas se plaindre, puisqu'on vous dit que ça permet à des gens de ne pas mourir... Et Twitter est sympa (source) :

### Messages Privés

Nos points de terminaison de Messages Privés fournissent un accès aux conversations par MP des utilisateurs qui ont explicitement accordé une autorisation à une application spécifique. Nous ne vendons pas les Messages Privés.

Il ne vend pas vos messages privés! (les messages publics étant bien entendu vendus, eux). Car attention, Twitter a des "valeurs fondamentales" (source):

## Défendre et respecter les droits de nos utilisateurs

Défendre et respecter la voix des utilisateurs est l'une de nos valeurs fondamentales chez Twitter. Avec cette valeur, nous nous engageons en faveur de la liberté d'expression et de la confidentialité. La transparence constitue également une part importante de cet engagement. Dans cet esprit,

Twitter souhaite rendre publique la manière dont nous traduisons cette valeur clé dans nos prises de décisions.

Et la "confidentialité" en fait partie. Ils font donc très attention dans leurs "prises de décision" – comme, par exemple, quand ils autorisent Visibrain à accéder à l'historique Twitter depuis 2006... Mais c'était prévu (source) :

## Partager vos données avec les partenaires commerciaux de Twitter

Lorsque nous recevons des données vous concernant, nous utilisons ces informations pour améliorer nos services et personnaliser votre expérience Twitter, ainsi qu'à d'autres fins décrites dans notre Politique de confidentialité. Dans certains cas, cela implique de partager des informations non publiques avec nos partenaires.

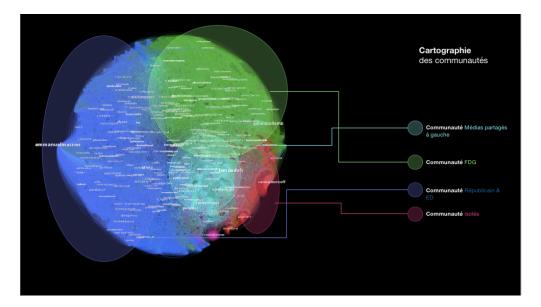
En conclusions, certains d'entre vous se diront peut-être : "Mais où le mal, ce sont des données publiques ?". Eh bien nous illustrons le problème dans la partie suivante...

## III. Quand EU DisinfoLab joue avec Big Brother

Nicolas Vanderbiest a donc utilisé Visibrain pour réaliser ses études. Dans la première (diffusée rapidement, plus orientée communication), au-delà des chiffres de volume (que fournit automatiquement le logiciel), il indique avoir réalisé ceci :

"4 communautés sont identifiées : une autour de la droite, des républicains et de l'extrême droite. (Plus on est à gauche, plus on est proche de la communauté d'extrême droite, en bas, on est plutôt républicains, et au-dessus, ce sont des sources médiatiques traditionnelles) ; une autour de la France Insoumise, qui vise particulièrement Gerard Collomb, une communauté médiatique, plus partagée par les personnes de gauche et enfin une communauté d'isolés qui par leurs tweets, ont élargi l'audience."

Il illustre ceci avec ce graphique:



### qui contient des noms de comptes :





## Il indique aussi ceci:



"Un certain nombre de profils d'extrême droite sont quelque peu tendancieux pour ne pas dire autre chose. En effet, leur activité cumule plusieurs centaines de tweets sur le sujet pour atteindre plus de 1000 tweets à propos de la polémique : " et il ajoute donc ce graphique :



Quand on lui fait remarquer que certains comptes ne sont pas d'extrême droite, Nicolas Vanderbiest répond avec sa rigueur académique (source – archive) :

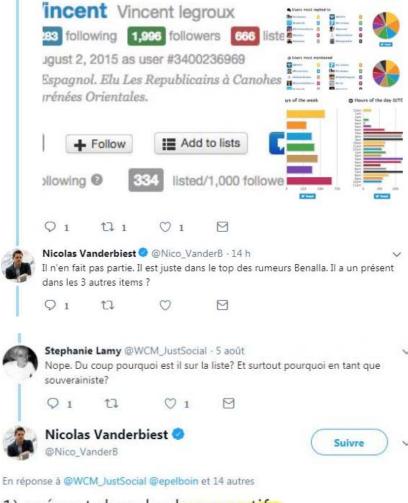
MILITER VEHICLES AND VILLES OF A



#FautApprendreÀLire...

Quand la professionnelle des médias sociaux Stéphanie Lamy le questionne sur le cas d'un twittos, Vincent Legroux, il répond ceci (source – archive) :





- 1) présent dans les hyper actifs
- 2) classés par nombre de rumeurs sur Benalla abordées
- 3) Souverainiste via ses following/followers

12:15 - 5 août 2018

Tiens donc, il en sait des choses Nicolas Vanderbiest...

Et pour les insoumis ? (source – archive)

12/08/2018 10:45 24 of 73



Ou'environ 70 000

Qu'environ 70 000 tweets provenants d'insoumis (le signe dans le pseudo ou insoumis dans la bio) dans ce classement des superactifs. C'est surtout l'ED à l'oeuvre je pense. Mais encore une fois, besoin de plus d'infos pour en parler sereinement

13:43 - 30 juil. 2018

Ne rigolez pas, c'est scientifique!



#Évidemment... Car c'est super précis comme méthode :



Ça se saurait si des insoumis avaient déjà relayé des infos de RT ou Sputnik!

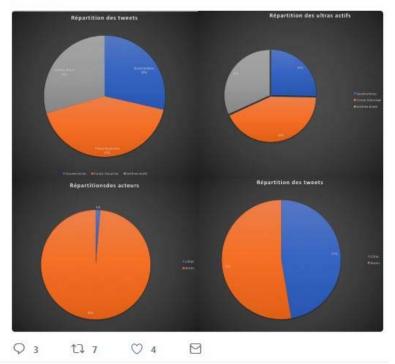
### Et sur les macronistes ?

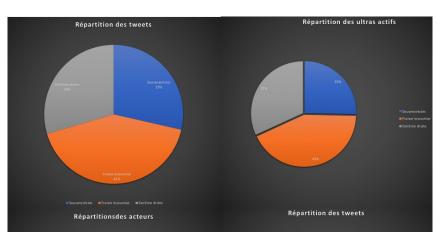


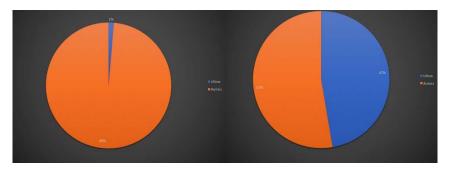
Tiens donc, il en sait des choses Nicolas Vanderbiest...

### Il indique même :

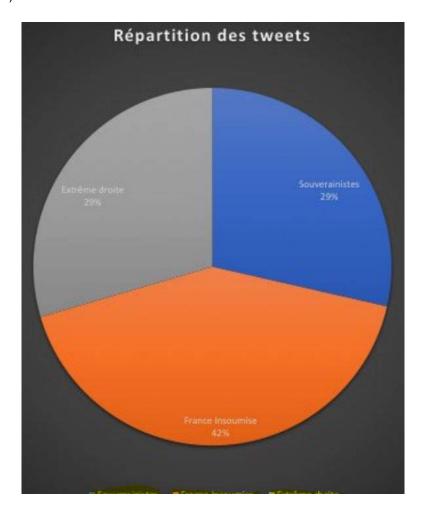








On parle apparemment bien ici des 4 millions de tweets (les hyperactifs sont dans l'autre camembert) :



Tiens donc, il en sait des choses Nicolas Vanderbiest...

Il ne peut faire son étude que s'il a un algorithme pour classer politiquement les comptes Twitter – ce qui n'a rien de très difficile en fait...

L'avantage c'est qu'on le sait, car il n'est pas très cachotier...

# IV. "Et c'est le drame..." : la diffusion des données personnelles

En effet, le chercheur d'EU DisinfoLab a indiqué (source - archive) :



C'est un habitué de la chose, notez – si vous voulez la liste des "russophiles" anglais aussi, pour une étude sur le "Brexit" (sic.) :

Nicolas Vanderbiest @ @Nico\_VanderB - 3 nov. 2017





14:51 - 3 nov. 2017



Et pour les chercheurs/consultants/journalistes/Whatever : les données brutes en Excel :

## docs.google.com/spreadsheets/d ...

23:31 - 31 janv. 2015

Ainsi, il indiquait ici à Samuel Laurent qu'il allait partager la base de données quand il aurait fini :



En réponse à @samuellaurent

Ils ont les mêmes types de RT, retweete les memes personnes, interagissent énormément. Avec le fichier du nb de rumeurs presidentielles croisés avec rumeurs sur Benalla, on a une bonne idée des profils impliqués. Je partagerai le fichier MaJ en DM quand on aura fini

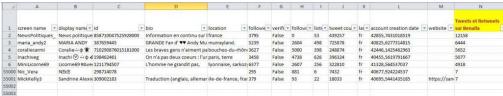
12:23 - 6 août 2018

### Et donc... Il l'a fait ! (source) (Preuve ici) :



Il a d'abord partagé publiquement via Dropbox le fichier des 55 000 plus gros diffuseurs de Tweets et Retweets (ayant diffusé entre 7 et 12 158 tweets), avec toutes leurs informations personnelles :





Cliquez pour agrandir

Rappelons de nouveau qu'1 retweet est compté comme un tweet, et que 82 % des "tweets" annoncés dans l'étude sont des retweets (source) :

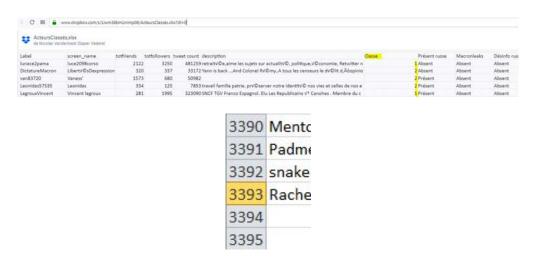




Puis ensuite, Nicolas Vanderbiest a fait beaucoup mieux : il a diffusé publiquement ceci (source – archive) :



à savoir un fichier avec les 3 392 plus gros twitteurs sur Benalla (qui sont donc surtout des retwitteurs), auxquels il a attribué une couleur politique via un algorithme qu'il a créé : 1 pour souverainistes (sic. – c'est la droite Républicaine), 2 pour Extrêmedroite et 3 pour la France insoumise! Et son tweet a été liké par EU DisinfoLab...



Analysons ce fichier dans la partie suivante.

# V. Le fichage politique (mais pas que) par EU DisinfoLab

Le fichier des 3 392 plus gros twitteurs sur Benalla se scinde ainsi par couleur politique (selon la définition et l'algorithme EU DisinfoLab) :

| Classe | Politique              | Nombre |
|--------|------------------------|--------|
| 1      | Proches Souverainistes | 855    |
| 2      | Proche Extrême-droite  | 1 073  |
| 3      | France Insoumise       | 1 447  |
| Isolés | 5                      | 17     |
| Total  | Total                  | 3 392  |

Bien entendu, nous ne diffuserons pas le fichier sur ce site. Mais nous vous montrerons un court extrait sur quelques très gros comptes, afin que chacun perçoive bien l'étendue

du problème. (mais le mal est fait, car la liste a été largement diffusée sur Twitter)

Voici donc pour ces quelques gros comptes "Classe 3 : France Insoumise" :

| Label           | ▼ screen_name ▼       |  |  |  |  |
|-----------------|-----------------------|--|--|--|--|
| cat_nole        | Cat_Love @üê±@üíñ     |  |  |  |  |
| DomJDom         | JAMNET EL BISMUST     |  |  |  |  |
| petimiti        | Christian B           |  |  |  |  |
| NTM_FN          | NTM le FN ⊡üîª⊡üåø    |  |  |  |  |
| NardineMorano   | Nardine Morano        |  |  |  |  |
| France Insoumis | se Helene Franco      |  |  |  |  |
| edwyntweet      | Edwyn                 |  |  |  |  |
| Dalipas         | Dalipas ф @üîª@üåø@   |  |  |  |  |
| Le_Ptrolls      | Le pisteur de Trolls  |  |  |  |  |
| CitoyenAtterre  | Olivier o Düåø, úä Dü |  |  |  |  |
| TwittNTIN       | Sami Abid             |  |  |  |  |
| Cabinet_Noir    | Le Cab. ©             |  |  |  |  |
| chgrand666      | christian666 φ ,úäΒἰ  |  |  |  |  |
| PSchydlowski    | Pierre Schydlowski    |  |  |  |  |
| IUMPmatuer      | Macron m'a tuer       |  |  |  |  |
| Bassounov       | Bassounov             |  |  |  |  |
| NeoResistant    | Jean Gautier          |  |  |  |  |
| tropical_boy    | 2üå∫ Tropical Boy 2ü  |  |  |  |  |
| yvanauet        | yvan auet             |  |  |  |  |
| Rollandcathare  | etienne               |  |  |  |  |
| MFrippon        | Duc de Saint-Frippor  |  |  |  |  |
| Vogelsong       | Vogelsong             |  |  |  |  |

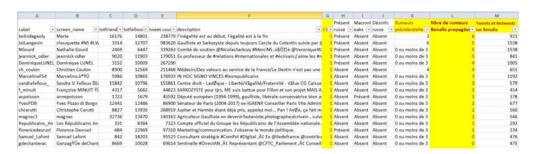
| A                | 8                      | C           | D           | E            | F   | G   | н       | . 1     |          | K                  | L.                | M             |        |
|------------------|------------------------|-------------|-------------|--------------|---|-----|---------|---------|----------|--------------------|-------------------|---------------|--------|
|                  |                        |             |             |              |   |     | Présent | Macroni | Désinfo  | Rumeurs            | Nore de rumeurs   | Tweets et Ret | tweets |
| Label -          | screen_name            | totfriend * | totfollow * | tweet cour * | description   | CI. | russe 1 | eaks =  | russe    | · présidentielle · | Benalla propagéei | sur Benalla   |        |
| cat_nole         | Cat_Love S06±S0iñ      | 420         | 8299        | 398013       | Insoumise à vie   | 1   | Absent  | Absent  | Absent   | . 3                |                   | 5             | 1865   |
| Dom/Dom          | JAMNET EL BISMUS       | 1 12096     | 11102       | 161412       | Ingénieur des #ArtsetMétiers Soutien de l'Association de Défense des    |     | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | 1             | 2362   |
| petimiti         | Christian B            | 7312        | 14417       | 135340       | Physicien #CNRS, melomane/collectionneur de blues, grand voyageur.      | 3   | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | 4             | 823    |
| NTM_FN           | NTM le FN B0PB0åe      | 2118        | 10310       | 24516        | Ni #FN ni #LR ni #LREM ,ûñ exténué par les cons ,ûñ mes tweets n'enr    | 1   | Présent | Absent  | Absent   | 0 ou moins de 3    |                   | 3             | 877    |
| NardineMorano    | Nardine Morano         | 965         | 21632       | 71545        | Compte parodique de Nadine Morano. Je tape plus vite que mes doigt      |     | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | 1             | 406    |
| France Insoumise | e Helene Franco        | 6607        | 11947       | 145421       | Co-animatrice du Livret justice de l'Avenir en commun #FranceInsoum     |     | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | 2             | 805    |
| edwyntweet       | Edwyn                  | 798         | 15702       | 16047        | Vâtudiant en science politique. Engagé.                                 | 1   | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | 2             | 758    |
| Dalipas          | Dalipas & Sül 130åel   | 850         | 5863        | 126656       | La vie est faite pour être vécue pas pour amasser " Partagé, tout est n | 3   | Présent | Absent  | Absent   | 0 ou moins de 3    |                   | 2             | 477    |
| Le_Ptrolls       | Le pisteur de Trolls   | 2970        | 10093       | 14356        | La France n'a pas de pétrole, mais qu'est ce qu'elle a comme trolls !!! | 3   | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | L             | 727    |
| CitoyenAtterre   | Olivier ф 2008 a, ú820 | 5475        | 5231        | 42349        | #FranceInsoumise FBPE - le hashtag des nuisibles - passez votre cher    | 3   | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | L Comment     | 605    |
| TwittNTIN        | Sami Abid              | 5701        | 6653        | 65601        | Psychiatre-Psychothérapeute-Addictologue / Homme de Gauche & #lir       | 1   | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | t.            | 435    |
| Cabinet_Noir     | Le Cab0                | 1039        | 29874       | 105422       | Responsable du cabinet noir de l'@Elysee.                               |     | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | 1             | 365    |
| chgrand666       | christian666 φ ,úli3   | 29007       | 31540       | 220447       | Fan de #Métal@úÄ De #Gauche ,úä -Plus jamais je ne voterais #Ps - #     | 1   | Absent  | Absent  | Absent   | 0 ou moins de 3    | 1                 | E.            | 358    |
| PSchydlowski     | Pierre Schydlowski     | 12796       | 18321       | 258456       | Vous pouvez reprendre une activité normale.                             |     | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | ı.            | 303    |
| IUMPmatuer       | Macron m'a tuer        | 10286       | 10982       | 191725       | L'UMP m'a tuer et le P5 m'a enterré ! #PlusJamaisP5 #TransitionEnerg    | . 1 | Absent  | Absent  | Absent   | 0 ou moins de 3    | 1                 | L.            | 221    |
| Bassounov        | Bassounov              | 1943        | 14596       | 59402        | Maréchal Léonid Bassounov. Grand Héros National de l'Union Soviétiq     | 1 3 | Présent | Absent  | Absent   | 0 ou moins de 3    |                   | t.            | 200    |
| NeoResistant     | Jean Gautier           | 15176       | 14887       | 308812       | Partageux et écolo Bûêù Vâlectron libre de gauche #Insoumis Athée       |     | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | 5             | 641    |
| tropical_boy     | Bull Tropical Boy B    | 1142        | 14381       | 37169        | ,ñðÔ∏è 1er Compte Twitter d'#humour et d'#INFO consacré en priorit      | 3   | Absent  | Absent  | Présent. | 0 ou moins de 3    |                   | 2             | 554    |
| yvanauet         | yvan auet              | 6208        | 6973        | 190801       | Avoir l'esprit RUGBY   #Marié; électron libre de la gauche #insoumis    | 1   | Absent  | Absent  | Absent   | 0 ou moins de 3    |                   | 9             | 393    |
| Rollandcathare   | etienne                | 5783        | 5753        | 46052        | INSOUMIS Conducteur de travaux ESTP bipolaire en souffrance Monite      |     | Absent  | Absent  | Absent   | Nouveau compte     |                   | 3             | 355    |
| MFrippon         | Duc de Saint-Frippo    | 43          | 12190       | 1994         | Duc et mémorialiste du règne d'Emmanuel 1er Roy des riches. Mémol       |     | Absent  | Absent  | Absent   | Nouveau compte     |                   | 3             | 275    |
| Vogelsong        | Vogelsong              | 511         | 17551       | 168574       | Chicaneur et non violent  | 1   | Absent  | Absent  | Absent   | 0 ou moins de 3    | - 1               | 9             | 233    |

### Cliquez pour agrandir

Vous voyez en colonne G la classification réalisée par Nicolas Vanderbiest, en colonne K le nombre de rumeurs/désinformations diffusées durant la présidentielle 2017 (voir ci-après pour une précision sur cette colonne), en colonne M le nombre de tweets et retweets en lien avec l'affaire Benalla et en colonne L le nombre de rumeurs/désinformations liées à l'affaire Benalla...

Voici un deuxième exemple pour ces quelques gros comptes "Classe 1 : Proche Souverainistes" (sic.) :

| Label           | screen_name          |  |  |  |  |
|-----------------|----------------------|--|--|--|--|
| belindageoly    | Marie                |  |  |  |  |
| JstLangevin     | chouquette #NS #LW   |  |  |  |  |
| NGourd          | Nathalie Gourd       |  |  |  |  |
| jeannick_odier  | jeannick odier       |  |  |  |  |
| DominiqueLUNEL  | Dominique LUNEL      |  |  |  |  |
| ch_coulon       | Christian Coulon     |  |  |  |  |
| MarcelinoF54    | Marcelino,ò®ŸÜ       |  |  |  |  |
| sandrafellous_  | Sandra V. Fellous ⊠ü |  |  |  |  |
| f_minuit        | Françoise MINUIT ŸÜ  |  |  |  |  |
| acpoisson       | annepoisson          |  |  |  |  |
| YvesPDB         | Yves Pozzo di Borgo  |  |  |  |  |
| chcerutti       | Christophe Cerutti   |  |  |  |  |
| magnac3         | magnac               |  |  |  |  |
| Republicains_An | Les Républicains An  |  |  |  |  |
| florencedesruol | Florence Desruol     |  |  |  |  |
| Samuel_Lafont   | Samuel Lafont        |  |  |  |  |
| gdechanterac    | GonzagŸÜe deChant    |  |  |  |  |
|                 |                      |  |  |  |  |



Voici un troisième exemple pour ces quelques gros comptes "Classe 2 : Proche Extrêmedroite" :

| Label           | screen_name 💌       |
|-----------------|---------------------|
| VotezPoisson    | Votez Poisson ,úùÔŢ |
| _Stalker_69_    | Stalker             |
| Valeurs         | Valeurs actuelles   |
| David_vanH      | David van Hemelrych |
| Dormez_Dormez   | Dormez braves gens  |
| ChLECHEVALIER   | Ch.LECHEVALIER      |
| lafrancelibretv | La France Libre TV  |
| Amandagr19      | Anne DeFrance       |
| FrkGuiot        | F.Guiot             |
| hangen_claude   | hangen claude       |
| SJallamion      | Sébastien JALLAMIO  |
| RTenfrancais    | RT France           |
| KimJongUnique   | Kim-Jong-Un ⊡üê∏    |



Mais vous notez que figurent aussi dans ce fichier les informations biographiques, en colonne E. Nous vous présentons donc un extrait avec quelques grands comptes de... journalistes :

screen name Label infos non stop ActusNonStop mansan man Mansan anamashiva France Saint Pierre GaspardGlanz Gaspard Glanz eddiemars212 Herve Couturier Sebastish Sébastien Dubois rsistons eva r-sistons Nassira El Moaddem NassiraELM

| TBMJ2_         | deSperate            |
|----------------|----------------------|
| nournewsnet    | A.Toumi ÿπÿ®ÿØ ÿߟ    |
| ThierryMorvan  | Thierry Morvan       |
| Jules78120     | Jules                |
| Val_Do         | Valérie de St-Do⊠üîª |
| catherinegaste | Catherine Gasté      |



Peut-être que maintenant notre choix de "Big Brother" se justifie-t-il plus clairement ?

Une petite application a été réalisée afin de voir si votre compte a été fiché politiquement par le DisinfoLab, c'est ici.

Pour connaître votre matricule d'agent de la russosphère dans la base DisinfoEU et les informations qu'elle contient à votre sujet, tapez votre @ twitter :



D'ailleurs EU DisinfoLab a également mis le 8 aout à disposition 2 autres fichiers, avec moins de données personnelles "pour être transparents" (source) :



Nous avons mis en ligne 2 fichiers pour être

transparents. Le fait de figurer dans un fichier signifie que vous avez été très actifs pendant l'affaire Benalla. Rien d'autre. Par ailleurs, le fichier dropbox qui circule n'est pas l'officiel. L'officiel : dl.free.fr/hUpdedh0I



### En effet (source):



Le fichier dropbox n'est pas le fichier original. Il y a eu un souci de colonne pour celle-ci. Le seul fichier qui compte est ici :



11:53 - 8 août 2018

### #IIYAEuUnSouciDeColonne

39 of 73

Le premier comprend un classement des 55 000 comptes les plus actifs sur Benalla avec simplement le nombre de tweets :

| 1     | А               | В             | C    |
|-------|-----------------|---------------|------|
| 1     | screen name     | Nombre de twe | ets. |
| 2     | NewsPolitiques_ | 12158         |      |
| 3     | maria_andy2     | 6444          |      |
| 4     | coraliecarmi    | 5652          |      |
| 5     | Inachiveg       | 5077          |      |
| 55000 | Nic_Vera        | 7             |      |
| 55001 | MickKelly3      | 7             |      |
| 55002 |                 |               |      |
| 55003 |                 |               |      |

Et un second fichier comprend des informations sur 3 890 comptes les plus actifs avec le nombre de prétendues désinformations sur l'affaire Benalla :

| _d   | A             | В                       | C              | D                | E                           | F | G               | H                            | 1                      |
|------|---------------|-------------------------|----------------|------------------|-----------------------------|---|-----------------|------------------------------|------------------------|
| 1    | Label         | Intérêt<br>• RT/Sputnik | ■ Macronleak = | Désinfo<br>russe | Rumeurs<br>présidentielle - |   |                 | Additions<br>désinformations | Tweets sur<br>le sujet |
| 2    | Marpa62       | Présent                 | Absent         | Absent           | Présent                     | 8 | 7               | 15                           | 2661                   |
| 3    | MariaBadin18  | Présent                 | Absent         | Absent           | Présent                     | 8 | 6               | 14                           | 3265                   |
| 4    | paoegilles    | Présent                 | Absent         | Présent          | Présent                     | 8 | 6               | 14                           | 2673                   |
| 5    | Philippe_FQSP | Présent                 | Absent         | Absent           | Présent                     | 8 | 6               | 14                           | 2363                   |
| 6    | chris_valy    | Présent                 | Présent        | Absent           | Présent                     | 7 | 6               | 13                           | 2543                   |
| 3890 | zehunter38    | Absent                  | Absent         | Absent           | Absent                      | 0 | 0 ou moins de 3 | 0                            | 316                    |
| 3891 | zigoto01      | Absent                  | Absent         | Absent           | Absent                      | 0 | 0 ou moins de 3 | 0                            | 261                    |
| 3892 |               |                         |                |                  |                             |   |                 |                              |                        |
| 3893 |               |                         |                |                  |                             |   |                 |                              |                        |

(Note : la colonne G détaille le nombre de désinformations lors de la Présidentielle. La colonne K du fichier de Nicolas Vanderbiest étant erronée, nous avons remplacé dans nos captures précédentes cette donnée par celles issues de ce fichier)

Vous pouvez donc regarder si votre compte Twitter a été fiché (mais en fait ils ont fiché 100 % des comptes...) et surtout classé politiquement (c'est le cas au moins si vous figurez dans les 3 800 détaillés) en téléchargeant ces fichiers ici.

L'équipe a d'ailleurs précisé comment avait été faite la limitation du second fichier : ce sont toutes les personnes qui ont fait 200 tweets ET retweets en 10 jours, le 200 étant défini pour arriver à environ 1 % de tous les twitteurs sur le sujet...



En réponse à @Narabro @Chani\_Bayle et 2 autres

Le but était de prendre plus ou moins uniquement 1% de l'échantillon

15:51 - 8 août 2018

Précisons enfin une chose, vu que nous avons ces données ; on obtient facilement ceci sur Excel :

| Classe                  | Nombre | Moyenne<br>de<br>followers | Moyenne de<br>Tweets et<br>Retweets sur<br>Benalla | Moyenne de<br>Nbre de rumeurs<br>Benalla<br>propagées | Taux de<br>Rumeurs<br>pour 1 000<br>tweets |
|-------------------------|--------|----------------------------|--|---|--|
| Drachae Counterainistee | OFF    | 1 746                      | 620  |   |  |

| Total                    | 3 392 | 1 599 | 557 | 2 | 4 |
|--------------------------|-------|-------|-----|---|---|
| Isolés                   | 17    | 143   | 373 | 1 | 3 |
| France Insoumise         | 1 447 | 1 310 | 544 | 1 | 3 |
| Proche Extrême-droite    | 1 073 | 1 893 | 521 | 3 | 7 |
| FIUUITES JUUVETAITIISTES | 033   | 1 /40 | 020 | J | J |

Ainsi, alors que Twitter n'aura eu à l'évidence qu'une influence bien limitée sur le développement de l'affaire Benalla (qui s'étalait dans tous les grands médias), on se rend compte, en plus, que le nombre de "rumeurs" (notion discutable que nous discuterons dans un autre billet) aura été de l'ordre, en moyenne, de 3 à 7 pour 1 000 tweets diffusés à 1 600 comptes en moyenne (seule une petite fraction les ayant vus)!

Ces chercheurs mettent ainsi en place des fichages généralisés de la population pour des phénomènes totalement dérisoires !

### VI. Communication de EU DisinfoLab

La journée du 8 a été riche en réactions de Twittos, ce qui a obligé EU DisinfoLab à réagir. Et ça tombe bien, car la structure a pour mémoire dans son orbite une agence de communication spécialisée sur les crises...

The Saper Vedere team helps you understand the concept of a crisis, its types and spreading methods. We then strategically and operationally support your organisation to solve any issues.

### Our crisis team includes:

~

Nicolas Vanderbiest, an academic researcher specialising in reputation crisis, who has studied and analysed more than 500 online reputation crises and their impact on organisations;

DADT A Madia aviala

# Nowadays, every single stakeholder claims to be a big data player by over-processing datasets. Most of the time, data visualisation simply displays the available information homogeneously, without giving constructive insights on how to benefit from it. Our methodology lets us look at every stakeholder and community involved in an issue separately, putting data in perspective. Our Media crisis methodology can: V Identify active stakeholders; V Understand elements and individuals that trigger the crisis; V Provide rapid and accurate recommendations to limit the effects of the crisis.

### Application (source):



|              | dérange e                   | et surtout les | acusations  | et sous entendu | rattachés.c |        |
|--------------|-----------------------------|----------------|-------------|-----------------|-------------|--------|
|              | Q 2                         | 17 1           | $\bigcirc$  |                 |             |        |
| EJ Diseo(La) | <b>EU Disi</b><br>@Disinfol | <b>nfoLab</b>  |             |                 |             | Suivre |
| En répo      | nse à @Ale                  | x031078 @cc    | olau57 et 2 | autres          |             |        |
| Vou          | s n'ật                      | ac nac         | fichás      | descus          | Lac dor     | nnáac  |

Vous n'êtes pas fichés dessus. Les données sont ici : dl.free.fr/hUpdedh0I

Quand on vous dit qu'ils fichent les gens - mais pas vous Alexandre, ça ira... (source)



En réponse à @leoxymoron @mimidellagioia @DisinfoEU\_FR

Vous y figurez pour dire que vous n'êtes pas quelqu'un qui fait de la désinformation.

... enfin sauf si c'est de la Désinformation Pro-Macron – mais oups, ce n'était pas dans les définitions retenues par le rédacteur...

Mais c'est toujours sympa d'avoir son certificat par la police, surtout sur des bases aussi solides.





Vous étiez parmi les plus actifs (+ de 200 tweets) mais avez partagé moins de 2 rumeurs. Il n'y a donc aucun problème et nous ne vous mettons pas en cause. Pour rappel, voici l'étude :

Jane, pas de souci, moins de 2 rumeurs, ça va... (c'est scientifique). Vous pouvez circuler. (source)



En réponse à @LeBobophobe @Nico\_VanderB

Bonjour, pas du tout. Vous n'êtes présents dans aucun écosystème et avez partagé moins de 2 rumeurs. Il n'y a donc aucun problème et nous ne vous mettons pas en cause. Pour rappel, voici l'étude :

Vous aussi BoboPhobe, vous pouvez circuler – mais pensez à changer de pseudo, ça ne plait pas à Macron, merci. (source)



Rassurez-vous George, tout va bien. Vous êtes un bon citoyen (N.B. : on voit que le "donc" est erroné, le critère pour figurer dans l'autre fichier étant (a priori) d'avoir fait 200 tweets, et non pas d'avoir diffusé au moins une rumeur...) (source)



En réponse à @FredSergent @Le\_Bonnet\_R et 2 autres

Vous y êtes en effet mais vous avez partagé moins de 3 rumeurs. Vous n'êtes donc pas catégorisé.

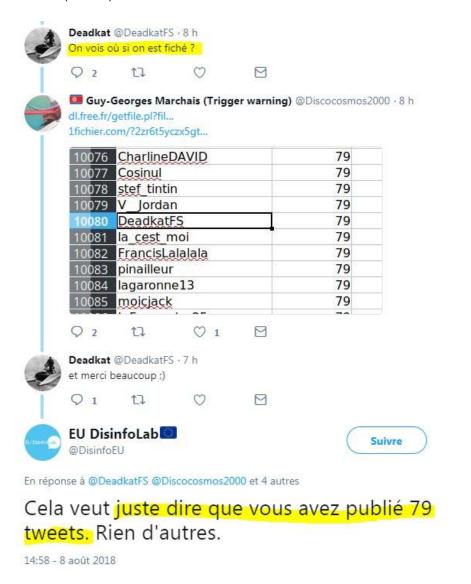


15:28 - 8 août 2018

Tssss Frédéric, tu as trop tweeté sur l'affaire Banalla. La prochaine fois, on te conseille de faire 1 tweet sur Benalla et immédiatement après 1 tweet pro-Macron. Comme ça, plus de problème avec la police! Après, si tu veux être sûr que Benalla ne vienne pas te voir en personne, contente-toi uniquement de faire des tweets pro-Macron, c'est plus sûr... (source)



Allez, Laurent ça ira pour cette fois... Mais change le hashtag immédiatement ou on t'envoie Benalla! (source)



Eh oui, Deadkat, tout va bien, t'es juste fiché! #RienDAutre. Bon ne dépasse pas 13 tweets la prochaine fois, et ça ira! (source)

Franck Noir = Чёрный Франк - Opposant №2461 @Le\_Bonnet\_R · 6 h



Opposant politique? Vous avez moins de 2 rumeurs et absents de tout. C'est tout ce que l'on dit.

15:26 - 8 août 2018

"Opposant politique" – tout de suite les grands mots. Ils n'ont jamais dit ça! (source)

Enfin, sauf dans leur fichier hein – classe 3, alors comme ça on fricote (peut-être ?) avec Mélenchon, Franck ?



En fait, le tout est de ne pas désinformer. Et en l'espèce, le nouveau Code Pénal d'Internet, établi sur des bases scientifiques, est très clair (source) :



Moins de 4 rumeurs et il n'y a aucun souci. Le souci vient quand il y a plus de 5...

15:08 - 8 août 2018

Tu vois Chyld, c'est vraiment terrible 5 rumeurs. Il y en a qui ont essayé, et ils ont eu des problèmes... (source)



Mais n'ayez pas peur, la brigade d'intervention contre la Désinformation est en route et va rapidement maitriser la forcenée.

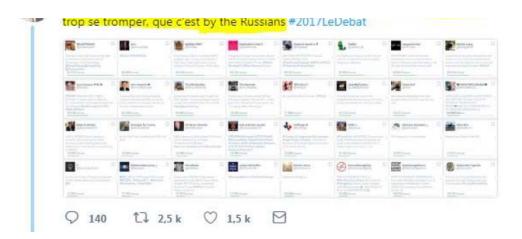


# VII. Position de EU DisinfoLab quant aux données sensibles

Ce n'est pas la première fois que se pose ce problème.

Nicolas Vanderbiest a déjà réagi l'année dernière suite à la diffusion de ceci (ah, c'était encore les Russes...) :





### une personne citée avait réagi :



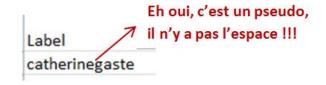


#C'EstPasUnFichage #C'EstUnExport #PasUneBaseDeDonnéesAuSensLégal

Il n'y a apparemment pas eu de mise à jour des connaissances depuis... Car des twittos se sont rapidement plaints à l'auteur (source) :



Eh oui il n'y a que des pseudos!



Mais:





Après, quand on parle directement à DisinfoLab du RGPD, le Règlement européen de protection des données personnelles qui vient d'entrer en vigueur, ils expliquent qu'il ne faut pas s'inquiéter, car ils n'ont fiché que **55 000 personnes**:





Bonjour, vous figurez dans le dossier qui reprend tous les auteurs de tweets sur l'affaire Benalla. (Comme plus de 55 000 personnes) Vous n'êtes pas présents dans l'autre fichier des hyper-actifs. Ce fichier sera supprimé par la suite.

14:01 - 8 août 2018

Et puis le fiché sera bientôt "supprimé" – ce n'est pas comme les gens l'avaient téléchargé et que des copies circulaient hors de tout contrôle hein! (source)



officiel. L'officiel: dl.free.fr/hUpdedh0I

14:02 - 8 août 2018

Et d'ailleurs, EU DisinfoLab explique que c'est à cause du RGPD (donc la règlementation de protection des données) qu'ils sont obligés de publier le fichier de données personnelles et sensibles (source) :



En réponse à @RoseLaloba

Les données ont été publiées pour qu'on ne remette pas en cause les enseignements et pour montrer aux gens comment leurs données ont été traitées. (GDPR) Regardez les comptes qui partagent le plus de rumeurs et dites-nous si nous sommes dans le faux. La liste sera supprimée.

14:38 - 8 août 2018

Voici d'ailleurs la réaction officielle de M. RGPD à la CNIL (source) :

Merci M. RGPD.

# VIII. Un laboratoire du CNRS aurait aussi fiché les opinions politiques de près de 200 000 personnes!

Dans le cadre de nos recherches, nous sommes tombés sur cet article des *Décodeurs* du *Monde* du 4 décembre 2017 :

# Le blog du Décodex

UN BLOG DE LA RÉDACTION . Le Monde

# Comment des chercheurs ont utilisé les données du Décodex pour analyser la présidentielle

Sciences Po a réuni chercheurs, développeurs et journalistes pendant quatre jours pour exploiter différents jeux de données liés à l'élection présidentielle, un événement dont nous étions partenaires.

LE MONDE | 04.12.2017 à 15h32 • Mis à jour le 23.01.2018 à 17h16 |

Par Adrien Sénécat

Adrien Sénécat nous indique qu'ils ont pu croiser "une dizaine de bases de données sur la période électorale", lors de l'évènement Datapol :

Plus d'une dizaine de bases de données sur la période électorale ont été proposées aux participants, dont celle du Décodex, sous licence ODbL. Concrètement, nous avons apporté deux types de données :

- l'annuaire des sources du Décodex, qui fournit une typologie des sources d'informations en fonction de leur fiabilité et donne des informations sur chacune des références étudiées (les données sont accessibles ici au format JSON);
- 2 notre base de données des fausses informations qui ont circulé en ligne et des liens qui les ont fait circuler (les données sont accessibles ici au format JSON).

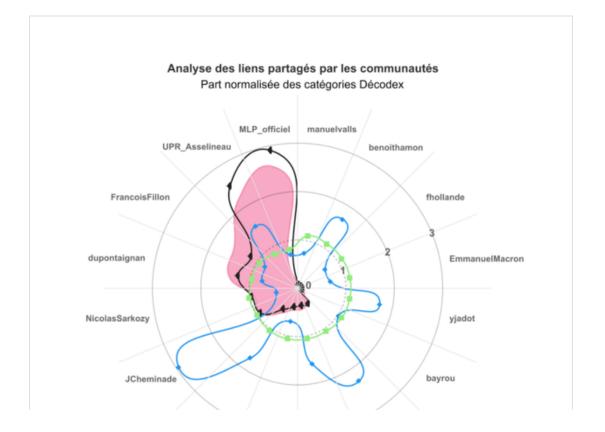
En croisant nos données avec d'autres bases et en faisant appel aux connaissances de participants venus d'horizons différents du nôtre, Datapol nous a permis de faire plusieurs découvertes intéressantes. En voici trois.

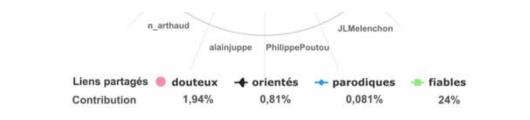
Ils ont réalisé une première étude sur "les fausses informations" (définies selon les critères méthodologiques du... Décodex). Puis une seconde qu'ils titrent "Les partisans de Marine Le Pen partagent plus de sources peu fiables que les autres" :

"Un autre enseignement intéressant est apparu en croisant les données du Décodex avec celles du Politoscope. [...] À partir de cette typologie et des données anonymisées du Politoscope, les participants à Datapol ont donc pu regarder dans quelle mesure les différentes communautés politiques partagent plus ou moins de liens vers les différents types de sources d'information identifiées dans le Décodex."

### L'infographie du Monde est ainsi construite :

- chaque candidat figure sur une tranche;
- plus un point est éloigné du 0, plus la communauté proche du candidat correspondant a partagé le type de sources en question :



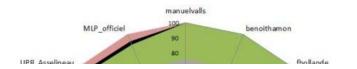


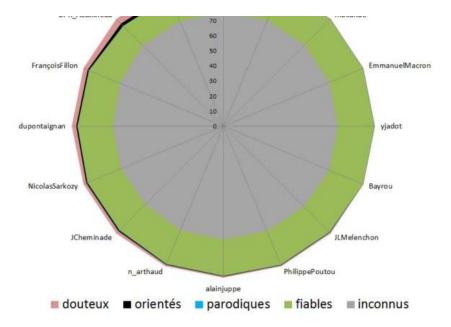
"Il apparaît ainsi que les partisans de Marine Le Pen et François Asselineau sont ceux qui, en moyenne, ont partagé le plus de liens vers des sources considérées comme peu fiables dans le Décodex. Une analyse qui mériterait d'être affinée par un travail approfondi, mais qui a le mérite d'apporter des données pour accréditer un comportement identifié par de nombreux observateurs pendant la campagne."

Avant de continuer sur notre axe, signalons quelques interrogations méthodologiques. D'abord, rappelons que la population des personnes inscrites sur Twitter n'est clairement **pas représentative** de la population générale (avec 49 % de CSP+ chez les 25-49 ans par exemple – source).

En fait le graphique a utilisé une présentation "douteuse". On a l'impression en regardant le graphique que les "communautés pro-Le Pen" et "pro-Asselineau" auraient inondé Twitter de liens vers des sites classés "très peu fiables". Or, les *Décodeurs* indiquent dans l'article que "*lorsqu'un point est proche du cercle de rayon 1, cela veut dire que le type de contenus correspondant a été partagé dans les mêmes proportions que la moyenne. Lorsqu'il est proche du trois, cela veut dire qu'il a été trois fois plus partagé." Ils représentent donc un simple écart à la moyenne, d'un phénomène marginal, d'où les forts écarts observés, qui sont en fait peu représentatifs. Ceci apparaît en fait clairement quand on observe la courbe verte peu visible de la diffusion de sites "fiables" : il y a assez peu d'écarts entre les candidats en réalité.* 

Illustrons. Nous avons reconstitué à la main les données du diagramme des *Décodeurs*. On peut dès lors reconstituer simplement ce que donnerait un graphique plus honnête représentant la diffusion de liens vers les sites classés par le *Décodex*:





Ce graphique donne donc une impression fort différente au lecteur de l'activité des personnes sur Twitter...

Mais revenons au Politoscope de L'Institut des Systèmes Complexes Paris-Île-De-France (ISC-PIF, du CNRS). L'article insiste bien : "Dans le cadre de ce projet, l'équipe de l'Institut des systèmes complexes Paris Ile-de-France, un laboratoire du CNRS, a analysé sur la durée de la campagne les messages de milliers d'internautes sur Twitter. L'un des intérêts du Politoscope est qu'il identifie la proximité d'un utilisateur de la plate-forme avec tel candidat au fil du temps". Voilà ce qu'on trouve sur le site dédié du Politoscope qui explique ceci :

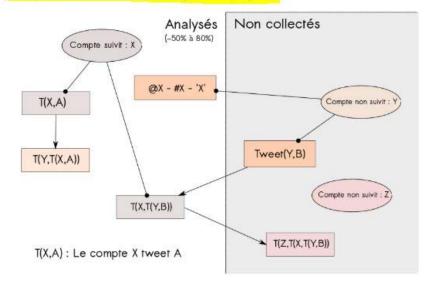
# Pourquoi un politoscope?

Twitter est devenu un nouveau média d'information ainsi qu'un passage obligé pour tout groupe qui cherche à influencer l'opinion publique d'une manière ou d'une autre, y compris pour certains, en publiant de fausses informations (« fake news »).

Le but du *Politoscope* est d'inverser les rôles en permettant au "grand public" de plonger dans les masses de données générées sur les réseaux sociaux à l'approche des présidentielles grâce à des outils et méthodes d'analyses produites par la recherche.

Depuis août 2016, l'ISC-PIF analyse les interactions de plus de 3000 comptes Twitter liés à la campagne présidentielle française 2017 (élus, partis, candidats) : (re)tweets, mentions, citations, etc. Une plateforme collecte et traite ainsi chaque mois plusieurs millions d'événements auxquels ont participé au total plus d'un million d'utilisateurs uniques. Une série d'analyses portant sur la structure des interactions et leur contenu est présentée dans ce dispositif, qui fait partie du *Politoscope*. L'objectif est de resituer dans le temps des événements perçus dans le flux continu de l'actualité. C'est une plongée inédite dans les différentes temporalités de la campagne présidentielle qui est proposée ici.

Le Politoscope suit depuis août 2016 plus de 3000 comptes Twitter appartenant à des médias et des personnalités du monde des médias, ainsi que certains flux ciblés thématiquement à partir de mots-clés et hashtags dans la limite du 1% Twitter (ex. #présidentielle2017). Plusieurs centaines de milliers de tweets politiques sont ainsi récoltés et analysés chaque jour.



Résumé. À l'occasion de l'élection présidentielle française de 2017, nous avons développé le Politoscope, une plateforme ouverte d'analyse des tweets politiques qui a collecté et analysé près de 60 millions d'échanges Twitter entre plus de 2,4 millions d'utilisateurs interagissant avec les personnalités politiques françaises. Dans cet article, nous nous intéressons

En fait, dans cette étude méthodologique, ils expliquent qu'ils sont partis de 3 700 comptes Twitter de figures politiques françaises, dont ils recueillent ainsi les tweets, mais surtout toutes les informations lors de retweets de ceux-ci ou de réponses (cf schéma précédent):

#### 7.1 Collection focused on main political accounts

We established a list of more than 3 700 accounts of French political figures (candidates, deputies, senators, persons with a significant party mandate). We gather informations from these accounts with the follow API of Twitter (see the inset).

The documentation describes exactly the data collected for a followed user account:

From Twitter follow API documentation<sup>a</sup>. For each user specified, the stream will contain:

- · Tweets created by the user.
- · Tweets which are retweeted by the user.
- · Replies to any Tweet created by the user.
- Retweets of any Tweet created by the user.
- Manual replies, created without pressing a reply button (e.g. "@twitterapi I agree").

The stream will not contain:

- Tweets mentioning the user (e.g. "Hello @twitterapi!").
- Manual Retweets created without pressing a Retweet button (e.g. "RT @twitterapi The API is great").
- Tweets by protected users.

ahttps://dev.twitter.com/streaming/overview/request-parameters

Given an account X, a content A and T(X,A) a tweet published by X containing A, we collected, within the Twitter API volume limit, for each followed account X and any account Y the tweets T(X,A), T(Y,T(X,A)) (e.g. retweet), and if A is a response to X(```````````````````, all <math>T(Y,A) (see the Fig. 10).

61 of 73

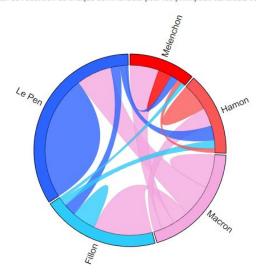
Et partant de là – vous le voyez venir -, quand l'échantillon de tweets est suffisamment important, ils peuvent (eux-aussi) inclure un compte dans une "communauté politique" :

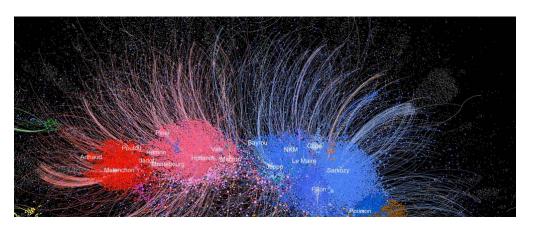
#### Thèmes et communautés

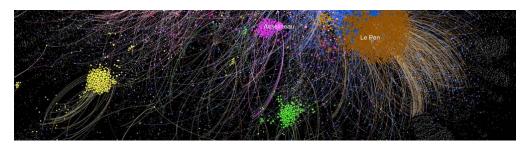
Les communautés politiques autour des candidats ont été définies algorithmiquement sur la base de l'analyse du partage de tweets à l'identique entre comptes Twitter (retweets "secs"). Les thèmes ont été identifiés grâce à des outils de fouille de texte développés à l'ISC-PIF. Voir la méthodologie pour plus de détails.

#### Attention des communautés pour la journée du 18.06.17

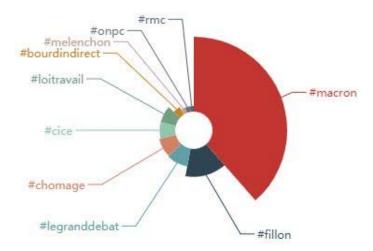
Détail de l'attention de chaque communauté pour les principaux candidats concurrents (survolez pour obtenir les infos)



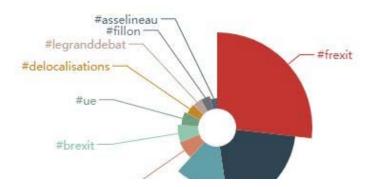




Et ils font alors des statistiques :



Répartition des hashtag les plus fréquents pour ce thème dans la communauté de @JLMelenchon





Et on apprend alors (discrètement) combien de comptes ils ont ainsi classé politiquement durant la présidentielle : 187 619 !

| Name                  | # people present | Average time in days |
|-----------------------|------------------|----------------------|
| Marine Le Pen         | 35 812           | 33.8                 |
| François Fillon       | 30 516           | 30.7                 |
| Jean-Luc Mélenchon    | 42 206           | 21.8                 |
| Emmanuel Macron       | 57 651           | 16.2                 |
| Benoît Hamon          | 37 748           | 17.7                 |
| François Asselineau   | 14 774           | 11.8                 |
| Philippe Poutou       | 51 908           | 11.6                 |
| Nicolas Dupont-Aignan | 30 366           | 5.6                  |
| Nathalie Arthaud      | 45 450           | 4.5                  |
| Jacque Cheminade      | 29 153           | 4.4                  |
| Global                | 187 619          | 55.8                 |

Table 2: Table showing, for each candidate, the number of persons present at least once in his/her community, as well as the average time spent inside the community. The communities of Le Pen and Fillon are clearly different to the others, in the sense that they are the most stable with a very high average time. The fact that the average time spent by an account within a community (last line) is very high indicates that a significant proportion of accounts change from one community to another, without disappearing from Twitter.

Si on cumule chaque ligne, on arrive à un total 375 000 personnes, car leur méthode est assez frustre (ils n'analysent pas le contenu des tweets par exemple). Ainsi une personne peut-elle est comptée comme pro-Poutou une semaine, puis pro-Arthaud la suivante, puis pro-Mélenchon la suivante, selon ses retweets (car cette sensibilité politique a plusieurs candidats possibles proches). Mais on voit que les auteurs indiquent que les communautés définies comme pro-Le Pen et pro-Fillon sont très stables... (sic.)

Et le plus fort est que leur opération ne s'est pas terminée avec la présidentielle – la base a doublé depuis :

64 of 73



Le Politoscope en est à plus de 126 millions de "tweets politiques", émanant de plus de 6 millions de comptes Twitter ! (dont une seule fraction est analysée politiquement comme on l'a vu)

L'avantage de ce laboratoire du CNRS, cependant, est qu'ils ne diffusent pas la base publiquement – mais évidemment, la simple existence d'un tel fichier pose de très lourds problèmes – sachant que n'importe quel gros utilisateur de Visibrain (ou de l'API de Twitter directement) peut faire de même, et largement perfectionner les attributions politiques s'il le souhaite...

Après, certains diront que ce sont juste des chercheurs animés de bonnes intentions. Ce qui est vrai.

Mais on répondra : "mais quelle est la réelle valeur ajoutée de ces travaux ?". Bien sûr on apprend des choses, mais est-ce vraiment si important ? Cela vaut-il le risque que de tels fichiers soient constitués – et donc peut-être utilisés un jour (ou simplement piratés) ? Êtes-vous à l'aise avec le fait qu'un tel fichier existe ?

Big Brother?

On imagine cependant que le laboratoire aura fait montre de prudence que DisinfoLab.

Mais cela pose néanmoins de nombreuses questions : les données sont-elles anonymisées, comment sont elles-stockées, est-il vraiment impossible de lever l'anonymat si la base était piratée ? (cela semble difficile, car il suffit de retrouver certains tweets dans Twitter, etc.)

Espérons en tout cas que ce laboratoire répondra rapidement à nos interrogations, qui restent à prendre au conditionnel...

Espérons en tout cas que ce laboratoire répondra rapidement à nos interrogations, qui restent à prendre au conditionnel...

## IX. Réglementation et Discussion

Rappelons que le Code pénal précise ceci :

Article 226-16

Le fait, y compris par négligence, de procéder ou de faire procéder à des traitements de données à caractère personnel sans qu'aient été respectées les **formalités préalables** à leur mise en oeuvre prévues par la loi est puni de cinq ans d'emprisonnement et de 300 000 euros d'amende.

Article 226-18

Le fait de collecter des données à caractère personnel par un moyen frauduleux, **déloyal** ou illicite est puni de cinq ans d'emprisonnement et de 300 000 euros d'amende.

Art. 226-19

Le fait, hors les cas prévus par la loi, de mettre ou de conserver en mémoire informatisée, sans le consentement exprès de l'intéressé, des données à caractère personnel qui, directement ou indirectement, font apparaître les origines raciales ou ethniques, **les opinions politiques**, philosophiques ou religieuses, ou les appartenances syndicales des personnes, ou qui sont relatives à la santé ou à

66 of 73

l'orientation ou à l'identité sexuelle de celles-ci, est puni de cinq ans d'emprisonnement et de 300 000 € d'amende.

Art. 226-22

Le fait, par toute personne qui a recueilli, à l'occasion de leur enregistrement, de leur classement, de leur transmission ou d'une autre forme de traitement, des données à caractère personnel dont la divulgation aurait pour effet de porter atteinte à la considération de l'intéressé ou à l'intimité de sa vie privée, de porter, sans autorisation de l'intéressé, ces données à la connaissance d'un tiers qui n'a pas qualité pour les recevoir est puni de cinq ans d'emprisonnement et de 300 000 € d'amende.

**La divulgation** prévue à l'alinéa précédent est punie de trois ans d'emprisonnement et de 100 000 € d'amende lorsqu'elle a été commise par imprudence ou négligence.

L'article 8 de la loi n°78-17 du 6 janvier 1978 (CNIL) apporte toutefois un bémol :

- I. Il est interdit de traiter des données à caractère personnel qui révèlent la prétendue origine raciale ou l'origine ethnique, **les opinions politiques**, les convictions religieuses ou philosophiques ou l'appartenance syndicale d'une personne physique ou de traiter des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique.
- II. **Dans la mesure où la finalité du traitement l'exige** pour certaines catégories de données, ne sont pas soumis à l'interdiction prévue au I :
- 1° Les traitements pour lesquels la personne concernée a donné son consentement exprès, sauf dans le cas où la loi prévoit que l'interdiction visée au I ne peut être levée par le consentement de la personne concernée;
- 4° Les traitements portant sur des données à caractère personnel rendues publiques par la personne concernée ;

Il y a bien sûr deux types de données. Celles publiées par l'utilisateur, et celles déduites par des algorithmes à partir des précédentes.

C'est ainsi que, par exemple, Samuel Laurent a pu dire ceci, interpellé à propos des fichiers DisinfoLab :



C'est à dire qu'on est sur un réseau public, et qu'on parle de chercheurs. Donc bon.

13:24 - 6 août 2018

Ce n'est évidemment pas faux – pour la partie publique (et peut-on véritablement parler de "chercheurs" pour des personnes non universitaires ?).

Mais, première objection : quand ce jeune étudiant indique son nom et sa couleur politique, pour agir sur Twitter avec ses... 23 abonnés ;



68 of 73



À t-il bien véritablement donné son accord pour que de nombreux analystes créent des fichiers avec son nom et ses opinions politiques ?

Il l'a certes dit publiquement sur Twitter, mais **si c'est public, cela reste discret**. Et un tweet s'oublie vite dans le fil, et le compte peut s'effacer. Mais ce n'est pas possible si des chercheurs créent des fichiers en permanence, non purgés.

Avec ce raisonnement, on pourrait d'ailleurs ficher plein de personnes : on vous reconnait en train de participer à une manifestation France Insoumise, vous sortez du local Les Républicains, vous avez défendu Macron dans l'affaire Benballa lors d'une discussion au bistro (surtout, ne prenez pas la route!). C'était public – peut-on vous ficher avec vos opinions politiques?

Et par ailleurs, on peut vraiment se demander si "la finalité du traitement exige" une telle utilisation des données.

Au-delà, il faut bien comprendre que **cette notion légale de "données rendues publiques", acceptable en 1976, ne l'est plus aujourd'hui.** En effet, par exemple avec Twitter, il y a différentes façons de connaitre vos opinions politiques :

- vous l'avez dit en dur dans votre biographie ou votre libellé de compte, comme l'étudiant ci-dessus (public) ;
- vous l'avez dit dans un tweet le soir du 1er tour de la présidentielle (public) ;
- vous relayez surtout un dirigeant politique (inférence très probable);
- vous parlez en général de vos opinions de nature politique, sans être partisan "il

faut libérer les énergies, on paie trop d'impôts pour trop de fonctionnaires, etc." (inférence probable)

 vous parlez beaucoup de vous sur Twitter, vous suivez et retwittez beaucoup de personnes. Et c'est là que le Big Data peut inférer vos opinions politiques avec une précision que vous n'imaginez probablement pas. (inférence possible)

Ainsi, désormais, les pouvoirs publics ne doivent pas seulement s'occuper des données non publiques, mais également de la collecte et surtout la diffusion de masses de données publiques, en particulier sur les réseaux sociaux.

Twitter est ainsi gravement fautif – finalement bien plus que DisinfoLab ou le ISC-PIF du CNRS. Après tout, si vous donnez des millions de données à des chercheurs, certains finiront immanquablement par les utiliser ainsi...

Bref c'est bien **les conditions d'utilisation de l'API de Twitter** qu'il faut viser dans le combat pour la protection des données de la population (source) :



L'affaire Benalla enflée par le "gonflage numérique", qui



Quand on pense que la CNIL a été créée en 1978 après une vive émotion dans l'opinion publique suite à un projet gouvernemental visant à identifier chaque citoyen par un numéro et interconnecter, via ce numéro, tous les fichiers de l'administration – on se rend compte de l'énorme baisse de notre vigilance, et des risques pour les Libertés publiques...

#### X. Plainte

Nicolas Vanderbiest a déclaré :



et:



71 of 73



Pour vérifier cela, j'ai porté plainte à la CNIL. – comme beaucoup d'entre vous l'ont fait. Nous soutenons en effet que l'identifiant Twitter (pseudo ou pas) est bien évidemment une donnée personnelle.

J'ai également mandaté mon avocat afin de **prévenir le Procureur de la République** de ces faits particulièrement graves.

Nous allons évidemment saisir également la CNIL belge et la CNIL irlandaise (siège de Twitter), ainsi que les services Protection des données de la Commission européenne.

Si des gens veulent nous aider pour agir à l'international, ou pour prévenir d'autres utilisateurs concernés et afin de grouper les plaintes, ou si vous êtes spécialiste de ce sujet de protection des données et pouvez nous aider, vous pouvez nous écrire ici.





## Dossier: #FichagePolitique - EU Disinfo Lab

#3 #4 #5

Tous fichés?
Comment Twitter
permet le fichage
politique de la
population

L'étude de l'influence russophile sur le Benallagate : d'énormes failles méthodologiques

Le EU DisinfoLab à l'origine d'une intox : l'affaire Benalla, produit de la « russosphère » ?

73 of 73

MÉDIAS 11.août.2018 // Les Crises

# Communiqué de presse : EU DisinfoLab, la diffusion scandaleuse d'un démenti mensonger





Après avoir montré des **lacunes** méthodologiques dans leur étude, puis leur **inconscience** dans le traitement de données personnelles sensibles et la création d'un fichage politique, puis leur **irresponsabilité** dans la diffusion publiques de ces données, voilà maintenant que le DisinfoLab donne dans... **la Désinformation et les Fake News.** 

## I. Rappel des faits sur le fichage de EU DisinfoLab

Rappelons que l'association belge EU DisinfoLab (créée le 27 décembre 2017) a acheté un accès à Visibrain, applicatif qui lui donne accès à TOUTE la base Twitter. Avec, ils ont récupéré les 247 701 comptes qui ont fait au moins un tweet ou un retweet sur l'affaire Benalla, et ont aspiré un certain nombre d'informations sur les dilatateurs figurant dans la base.

Ils ont alors créé un autre fichier comprenant les 55 000 comptes qui ont produit plus de 7 (re)tweets, et comprenant de nombreuses informations personnelles sensibles auxquelles ce fichier permet d'accéder instantanément. La plupart figurent dans la Biographie du profil (rédigée par l'utilisateur) ; il est à noter que ces informations permettent de faire une recherche ciblée sur un indicateur de religion, d'orientation sexuelle, d'opinion politique, etc., et de faire des regroupements très facilement. Bien que les données soient indiquées publiquement par le titulaire du compte, il faut savoir qu'il n'est pas possible de les retrouver avec une utilisation "classique" du moteur de recherche Twitter (essayez donc de trouver 200 personnes ayant indiqué être bouddhistes dans leur biographie... C'est impossible pour 99 % des utilisateurs). Il faut payer Twitter pour cela.

EU DisinfoLab a ensuite extrait plus de 3 000 profils qui ont produit plus de 200 (re)tweets, avec les informations précédentes auxquelles s'ajoute une classification réalisée par un algorithme qu'ils ont créé qui donne une couleur politique à ces comptes d'opposants (« Souverainiste », « Extrême Droite » ou « France Insoumise »).

Le 5 août, Nicolas Vanderbiest, fondateur d'EU DisinfoLab, a alors diffusé sur Twitter les 2 fichiers ("Data brutes.csv" pour les 55 000 et "ActeursClassés.xlsx" pour le fichage politique), qui ont alors circulé sur Twitter – ne tenant pas compte des demandes pressantes de suppression du fichier (exemples ici, ici, ici ou ici). Il a fini par supprimer les fichiers 4 jours après, suite à une masse de plaintes à la CNIL, qui s'est saisie du dossier. EU DisinfoLab a alors diffusé, via dl.free.fr, un fichier zip comprenant 2 fichiers « allégés », lien donné sur Twitter (26 fois...) et dans leur étude Benalla, sans données personnelles ni politiques ("Data brutes 1 – 47.xlsx" et "Rumeurs & items.xlsx"), mais comprenant le nombre de rumeurs supposément diffusées par les plus de 3 000 acomptes très actifs, avant de supprimer également ces fichiers sur dl.free.fr le 9 aout à 18h38 (source : étude Benalla DisinfoLab).

#### Comment vérifier vos conclusions?

Nous mettons à disposition des journalistes accrédités et d'organismes de recherche les datasets suivants :

- Le classement des plus actifs qui attestent du fait que 1 % des auteurs a publié 47 % des tweets
- Le fichier avec le nombre de rumeurs par individus classés dans les plus actifs

#### <u>Télécharger</u>

Fidèles à notre mission, nous encourageons toute initiative de recherche qui permettrait d'approfondir notre méthodologie ou d'en développer de nouvelles.

#### Comment vérifier vos conclusions ? (Edit: 9/08/2018 18h38)

Nous mettons à disposition des journalistes accrédités et d'organismes de recherche les datasets suivants :

- Le classement des plus actifs qui attestent du fait que 1 % des auteurs a publié 47 % des tweets
- Le fichier avec le nombre de rumeurs par individus classés dans les plus actifs

Fidèles à notre mission, nous encourageons toute initiative de recherche qui permettrait d'approfondir notre méthodologie ou d'en développer de nouvelles.

#### II. Nos tweets du 10 août 2018

Afin que chacun comprenne la problématique de ces fichiers 1/ comprenant des masses de données sensibles, certes fournies publiquement (mais de façon discrète, à sa communauté de followers, généralement limitée) et 2/ du fichage politique réalisé par algorithme EU Disinfolab, nous avons réalisé des tweets avec une extraction (que nous avons anonymisée) de certains profils présentant des informations très sensibles :

À ceux qui minimisent le #FichagePolitique de DisinfoLab, voici un extrait anonymisé

 $3 ext{ of } 26$  12/08/2018 10:21

https://www.les-crises.fr/eu-disinfolab-la-diffusion-scandaleuse-dun-dementi-mensonger/

des 2 fichiers qu'ils ont créés et diffusés.

Quelles que soient nos convictions politiques, nous devrions tous nous réunir pour lutter contre ça. C'est ce qui avait permis la création de la @CNIL pic.twitter.com/FWpvHIjCAv

— Olivier Berruyer (@OBerruyer) 10 août 2018

Afin que tout soit bien clair, nous avons réalisé un rappel des faits en thread :

Rappels importants sur l'affaire DisinfoLab. Un dossier détaillé est disponible ici : https://t.co/kqKm6RNiis pic.twitter.com/3ODCJGMLJ6

— Olivier Berruyer (@OBerruyer) 10 août 2018

Puis nous avons alerté les partis politiques et associations concernées :

Que pensez-vous de modifier la loi pour une meilleure protection de nos données personnelles pour éviter tout fichage ? @FranceInsoumise @partisocialiste @enmarchefr @gouvernementFR @lesRepublicains @DLF\_Officiel @RNational\_off @CNIL @\_LICRA\_ @LDH\_Fr @amnestyfrance @laquadrature pic.twitter.com/Cu8DzHiWvq

— Olivier Berruyer (@OBerruyer) 10 août 2018

N'hésitez d'ailleurs pas à les retweeter.

## III. Le Communiqué EU DisinfoLab du 11 août 2018

Nous avons alors eu la surprise de lire ce communiqué de EU Disinfo Lab :

4 of 26 12/08/2018 10:21

https://www.les-crises.fr/eu-disinfolab-la-diffusion-scandaleuse-dun-dementi-mensonger/

## 11/08/2018 : Communiqué de presse – Fausses captures d'écran

Nous remarquons qu'une capture d'écran circule actuellement sur Twitter, Facebook et des sites d'information mettant en cause un des fichiers que nous avons transmis. Sur cette capture d'écran apparaissent des mots surlignés qui feraient penser qu'il y aurait des personnes « gay », « lesbienne », « juifs ». Aucune de ces biographies n'est issue du fichier. Tout un chacun pourra le vérifier simplement en faisant une recherche dans les fichiers. Il pourra constater que :

- Le nombre de tweets affichés sur la capture d'écran est impossible. Seules les personnes avec + de 200 tweets étaient rassemblées.
- Les localisations ne faisaient pas partie des fichiers que nous avons transmis. Par ailleurs, nous nous étonnons d'une localisation « marié gay en prc »

Par ailleurs, nous rappelons que le premier fichier de 55 000 pseudonymes ne comprenait que les pseudonymes des personnes ayant tweeté sur le sujet et le nombre de tweets. Rien d'autre.

Nous regrettons vivement le climat de désinformation qui règne autour de cette affaire et rappelons que, conformément au RGPD, les personnes concernées peuvent faire valoir leur droits d'effacement légitimes en nous contactant par e-mail à droitdacces @disinfo.eu

Source : Disinfo.eu (archive)

Puis celle de lire ces 7 tweets de cette association nous diffamant de façon stupéfiante (sources : 1, 2 et 2 bis, 3, 4, 5, 6, 7 – toutes) :





Attention, de fausses captures d'écran



En réponse à @Mass\_Critic @OBerruyer @gchampeau

Il suffit de verifier pour voir que même dans le fichier acteurs classés, il n'y a pas ce qui est affiché: nul part. C'est un faux.

Le fichier de 50 000 utilisateurs ne comprends aucune autre metadonnée que le nombre de tweets.

11:48 - 11 août 2018

 $6 ext{ of } 26$  12/08/2018 10:21

#### Ce mensonge fait évidemment réagir :



Si, si, la cuisine est nécessaire...

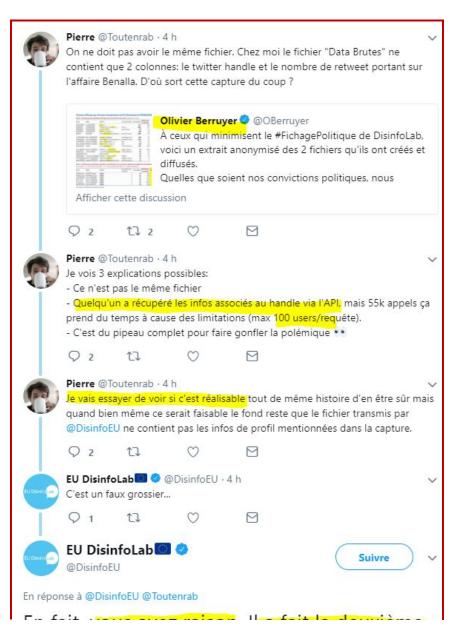


7 of 26 12/08/2018 10:21

## C'est un faux grossier...

12:24 - 11 août 2018

#### Puis vient alors l'instant complotisme délirant :



en Tait, vous avez raison. Il a Tait la deuxieme option

12:40 - 11 août 2018

Ils indiquent donc que nous aurions NOUS-MÊMES créé le fichier, en recopiant à la main via l'API Twitter, 100 par 100, 55 000 profils Twitter (ce qu'elle ne permet pas), et d'ailleurs pour classer correctement et arriver au même fichier qu'eux, ce sont les 4 300 000 (re) tweets qu'il aurait fallu récupérer (et dommage, on ne peut récupérer plus de quelques retweets sans l'API)

Nous plaignons Pierre, ce twittos qui, encouragé par eux, aura passé son samedi après-midi à essayer de voir si ce délire est crédible... Quel manque de respect envers les citoyens vigilants!

#### Et ils insistent:



En réponse à @DisinfoEU @Toutenrab

Nous n'avons jamais eu de fichier brut comme lui. Le fichier d'extraction de Visibrain ne donne que l'username et le mot de passe.

12:43 - 11 août 2018

Comme ils ont fait une erreur sur "mot de passe" (sic.), ils corrigent :



Maintenant, pour eux, Visibrain ne permet pas d'extraire autre chose que "l'username et le mot de passe"; oups, "l'username et le nombre de tweets". Par chance, nous avons exposé le fonctionnement de ce logiciel Big Brother ici dans ce billet (source) :



10 of 26 12/08/2018 10:21

Et on peut aussi exporter tous les followers, en les croisant (source) :

## VISIBRAIN PLATEFORME > INFLUENCEURS Extraction de followers

Avec Visibrain, vous pouvez exporter

l'exhaustivité des followers de n'importe quel compte au format Excel : les followers de votre marque, ceux de vos concurrents, ceux d'un

influenceur...

Visibrain vous permet également d'effectuer des croisements de liste de followers pour identifier quels followers vous avez en commun avec vos concurrents, les comptes qui suivent vos concurrents sans vous suivre, ou bien les comptes qui suivent l'ensemble des influenceurs de votre secteur.

| B B                 | E   | F                         | G                  | н              | 1              | 1           |        |
|---------------------|---|---------------------------|--------------------|----------------|----------------|-------------|--------|
| Nom d'utilisateur - | Biographie  | Site Web 🔻                | Date de création 🔻 | Nb followers - | Nb following - | Nb listed - | Nombre |
| C8TV                | ps émissions : #TPMP #WAM #SLT #LTD #Famil        |                           | 20/05/2009 09:53   | 767064         | 1469           | 983         |        |
| TopAchat            | icier du #Geek depuis 1999 ! Grosse command       | http://www.topachat.      | 19/01/2010 09:42   | 410177         | 9393           | 647         |        |
| bouyguestelecom     | uygues Telecom partage votre passion de la te     | http://www.bouygues       | 02/10/2008 12:02   | 143808         | 65237          | 790         |        |
| SNCF                | ut l'univers SNCF à portée de clic : les infos en | http://www.sncf.com       | 22/11/2010 23:53   | 108644         | 957            | 725         |        |
| Le_Figaro           | Actualités, économie, politique, culture et div   | http://www.lefigaro.fr    | 22/08/2007 09:16   | 2800055        | 537            | 14510       |        |
| Europe1             | ute l'information d'@Europe1 et nos émission      | http://www.europe1.f      | 23/04/2009 10:53   | 1311255        | 3542           | 5271        |        |
| 20Minutes           | formation en continu. Actualités, Politique, Sp   | http://www.20minute       | 28/03/2007 14:14   | 2308909        | 1469           | 9180        |        |
| LesEchos            | ctualité économique, financière et boursière f    | http://www.lesechos.      | 25/11/2009 18:01   | 1066877        | 500            | 7197        |        |
| LCI                 | envenue sur LCI, sur le canal 26 de la TNT. Le n  | http://lci.fr/            | 23/03/2009 23:45   | 109113         | 964            | 1457        |        |
| BFMTV               | re chaîne info de France. Retrouvez-nous sur S    | http://www.bfmtv.cor      | 16/04/2010 11:55   | 2267084        | 488            | 8310        |        |
| OuestFrance         | mpte Twitter du quotidien Ouest-France. Rec       | http://www.ouest-fram     | 27/04/2009 17:59   | 518358         | 3518           | 3124        |        |
| romandie            | mpte officiel de http://Romandie.com, média       | http://www.romandie       | 27/03/2009 13:43   | 25117          | 2              | 591         |        |
| Niska_Officiel      | puvel album #Commando maintenant disponit         | http://www.charo-shc      | 02/11/2011 01:37   | 383839         | 36             | 247         | 5      |
| lalibrebe           | mpte officiel de Lalibre.be Rejoignez-nous éga    | http://www.lalibre.be     | 04/02/2009 11:37   | 203785         | 2896           | 2155        |        |
| lemondefr           | ctualité de référence par la rédaction du Mon     | http://www.lemonde.       | 16/03/2009 19:44   | 7511211        | 504            | 33063       |        |
| AmazonHelp          | e answer Amazon support questions 7 days a        | http://www.amazon.c       | 28/10/2009 05:17   | 124674         | 52             | 741         |        |
| Orange_conseil      | sez vos questions du lundi au samedi, de 8h à     | 20h dans les domaines     | 13/10/2010 11:16   | 88857          | 44182          | 431         |        |
| afpfr               | orceaux choisis des fils d'actu, coulisses & RT o | http://www.afp.com        | 20/07/2011 14:07   | 2855687        | 1348           | 12990       |        |
| AmazonFR            | mpte Officiel Twitter http://www.Amazon.fr.       | http://amazon.fr          | 30/09/2015 14:07   | 21722          | 1              | 61          |        |
| Qc511_QcLevis       | cidents & entraves sur le réseau autoroutier de   | http://quebec511.info     | 31/03/2014 15:37   | 3001           | 15             | 59          |        |
| Sosh_fr             | sh, la marque sans engagement par Orange, re      | http://www.sosh.fr/       | 19/07/2011 11:35   | 138831         | 29549          | 578         |        |
| tvanouvelles        | a télé, sur votre ordinateur ou votre appareil    | http://tvanouvelles.ca    | 26/10/2009 15:59   | 343378         | 538            | 2174        |        |
| PSchydlowski        | rtains mettent à jour leur statut @facebook to    | outes les 5 secondes, ils | 09/01/2011 11:42   | 41003          | 20691          | 280         |        |
| le_Parisien         | Parisien-Aujourd'hui en France, 1er guotidien     | http://www.leparisien     | 06/05/2009 10:50   | 1958055        | 910            | 7770        |        |

Un autre exemple des données Visibrain...

C'est d'ailleurs ce qu'ils expliquent... dans leur propre étude :

https://www.les-crises.fr/eu-disinfolab-la-diffusion-scandaleuse-dun-dementi-mensonger/

Ensuite, nous analysons ces tweets de plusieurs manières, dont la cartographie des acteurs. Pour cette étude particulière, nous avons analysé la manière dont les acteurs se suivaient sur Twitter.

Rappelons que EU DisinfoLab est censé lutter contre la désinformation... Fact-checkons donc ses déclarations !

## IV. Fact-checking du scandaleux démenti de Disinfo : "Pourquoi c'est faux"

Vous vous imaginez ce que doit ressentir face à ceci un citoyen lambda qui aurait réalisé le même genre de travaux que nous... Ou tout simplement les twittos fichés qui ont téléchargé le fichier et ont vu de leurs propre yeux ce que DisinfoLab cherche insidieusement à démentir ?

Précisons également que, selon leur site, l'équipe du DisinfoLab est très limitée, que Nicolas Vanderbiest est le cofondateur de l'association, et la seule "caution scientifique" (bien qu'il soit encore étudiant doctorant, diplômé en information et communication). Il a été pratiquement la seule personne à communiquer des informations techniques sur Twitter sur l'étude en cours :



Our Team



Gary Machado
President

Gary is our co-founder, with over a decade of EU

Nicolas Vanderbiest



Alexandre Alaphilippe

Nicolas Vanderbiest is a researcher and PhD

After a degree in Communications in his hometown

affairs experience under his belt. He brings his vast EU bubble expertise to EU DisinfoLab, sharing his time with a highly successful international association he has been leading for the past 11 years. Gary holds an MA in European Affairs and BA a database of more than 530 social media crises. in Economics from universities in France and Scotland. Despite being French, he's a true Europetrotter, now residing in Brussels, but having also studied or worked in Ireland, Spain, Poland and in the UK. Besides his native French he's also fluent in English, spiced up with basic Spanish and Polish.

candidate at the Université Catholique de Louvain in Belgium. His thesis subject is "Reputation crises of organizations on social media". He is the author of the highly-popular Reputatio Lab blog, which gathers

He also analyzes crises, astroturfing and social media phenomena for the media - previously at Rue 89, France Culture / La Première (radio), currently for RTBF (TV). In EU DisinfoLab he acts as a special advisor, providing his technical expertise and

thodology as well as shaping processes, with his secret skill being bringing structure to Martin's mess. His working language is French, but he also speaks fluent English, albeit with a distractingly strong accent.

of Clermont-Ferrand, France, Alexandre has worked for 4 years as Chief Digital Officer at the Clermont-Ferrand city hall. Hungry for challenges, he decided to move to Brussels, just "because he liked the city" (1). Being very vocal on the crucial role of civil society organisations in maintaining democratic values, in 2017 Alexandre co-founded the EU DisinfoLab. He speaks French (of course), English (as good as a French man can), Dutch, Spanish (less and less) and Italian (more and more).



## Martin Mycielski

Martin serves as Brussels Correspondent for leading EU institutions, Polish Permanent Representation to the EU and international NGOs since 2008 - and running political campaigns prior to that - he brings his journalism, EU bubble & political communications experience to EU DisinfoLab. He holds a Master's in Diplomacy & International Relations from Madrid, having also studied English Philology, Iberian Culture and Management at Warsaw University. Raised in the UK and being a dual Polish/Belgian citizen, he is Polish/English bi-lingual with a deceivingly good French accent.





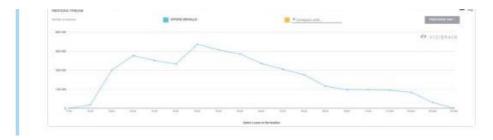
Léa Charlet

Léa is bringing a wide range of experience and a versatile profile to EU DisinfoLab. She worked in different NGOs, think tanks, small companies and European Institutions, always bringing her knowledge of organisational governance. While graduating in comparative public law and European studies, she has been leading a wide trans-European campaign aimed at raising the turnout of young people during the European Parliament elections of 2014 with her student NGO In EU DisinfoLab, she is the process queen. She speaks French, English without so much of a French accent, Spanish, Italian and a tiny bit of Czech and

Comme il l'avait annoncé, le chercheur a donc diffusé sur Twitter les données détaillées de l'étude DisinfoLab le 5 août (source) (Preuve ici) : (TWEET SUPPRIMÉ par Nicolas Vanderbiest):

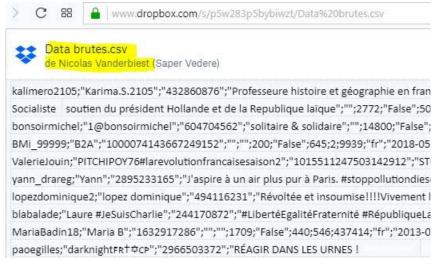


12/08/2018 10:21 13 of 26



Note: Nicolas Vanderbiest a supprimé de très nombreux tweets "pour cacher ses erreurs" – espérons que la CNIL et la Justice demanderont bien à Twitter l'historique complète de toutes ses suppressions des derniers jours.

Il a donc par ce tweet d'abord partagé publiquement via Dropbox "Data brutes.csv", le fichier des 55 000 plus gros diffuseurs de Tweets et Retweets (ayant diffusé plus de 7 (re)tweets sur l'affaire Benalla), avec toutes leurs informations personnelles :



| - 4  | A               | В               | С                  | D                            | Ε                   | F           | G        | Н      | 1       | 1         | K     | L                     | M           | N                                 |
|------|-----------------|-----------------|--------------------|------------------------------|---------------------|-------------|----------|--------|---------|-----------|-------|-----------------------|-------------|-----------------------------------|
| 1    | screen name     | display name *  | id s               | bio                          | location            | • followe • | verifi * | follow | list( * | tweet cou | lat * | account creation date | * website * | Tweets et Retweets<br>sur Benalla |
| 2    | NewsPolitiques_ | News politique  | 858710047525928000 | Information en continu sur   | Ifrance             | 3795        | False    | 0      | 53      | 439257    | fr    | 42855,7431018519      |             | 12158                             |
| 3    | maria_andy2     | MARIA ANDY      | 387659445          | GRANDE Fan d' ♥♥ Andy Mi     | murrayland.         | 5239        | False    | 2604   | 498     | 725878    | fr    | 40825,6277314815      |             | 6444                              |
| 4    | coraliecarmi    | Coralie △ ф 🛊   | 710290876015181000 | Les braves gens n'aiment p   | a bouches-du-rhô    | n 3627      | False    | 5000   | 398     | 248874    | fr    | 42446,1425462963      |             | 5652                              |
| 5    | Inachiveg       | Inachi (Ψ - φ 4 | 198462461          | On n'a pas deux coeurs : l'u | r paris, terre      | 3458        | False    | 4738   | 626     | 396324    | fr    | 40455,5619791667      |             | 5077                              |
| 6    | MiniLicorne69   | Licorne69 #Ave  | 1211794507         | L'homme ne grandit pas,      | lyonnaise, sarko    | oz) 6377    | False    | 2607   | 256     | 322810    | fr    | 41328,564537037       |             | 4918                              |
| 5000 | Nic_Vera        | NEcE            | 298714078          |                              |                     | 295         | False    | 881    | 6       | 7432      | fr    | 40677,924224537       |             | 7                                 |
| 5001 | MickKelly3      | Sandrine Alexie | 309002183          | Traduction (anglais, allema  | r Ile-de-france, fr | rar 379     | False    | 93     | 22      | 18033     | fr    | 40695,5441435185      | https://san | 7                                 |
| 5002 |                 |                 |                    |                              |                     |             |          |        |         |           |       |                       |             |                                   |
| 5003 |                 |                 |                    |                              |                     |             |          |        |         |           |       |                       |             |                                   |

Cliquez pour agrandir. On a le nom de l'auteur et on voit les informations sensibles "que EU DisinfoLab n'a jamais eu"

14 of 26 12/08/2018 10:21

Ce fichier qui n'a jamais existé contient ces 14 données en colonne :

- A. l'username / libellé du compte (le @),
- B. le nom public,
- C. "listed"l'id,
- D. la biographie du profil,
- E. la localisation,
- F. le nombre de followers,
- G. le statut vérifié,
- H. le nombre de comptes suivis,
- I. "listed" (?),
- J. le nombre de tweets du compte,
- K. la langue,
- L. la date de création du compte,
- M. le site internet,
- N. le nombre de Tweets et Retweets sur Benalla

De plus, au vu la gravité la chose, nous avons organisé un réseau de témoins tiers de confiance pour prouver la véracité de ce fichier, en vue de notre plainte à la CNIL, nous doutant bien que les auteurs finiraient par le supprimer, et étaient probablement capables de nier son existence... #PasDeBol

Et nous avons bien fait, vu qu'il est désormais possible au vu de leurs tweets que DisinfoLab ait détruit les preuves de leurs agissements pour faire échec à l'enquête judiciaire et de la CNIL – nous le verrons. Nous tenons si besoin le fichier à la disposition de la CNIL (avant de bien entendu le détruire).

Ensuite, Nicolas Vanderbiest a récidivé : il a diffusé publiquement le fichier "ActeursClassés.xlsx" (source – archive) :

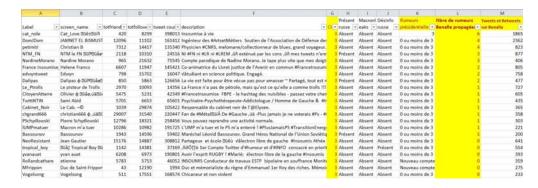


à savoir un fichier avec les 3 392 plus gros twitteurs sur Benalla (qui sont donc surtout des retwitteurs), auxquels il a attribué une couleur politique via un algorithme qu'il a créé : 1 pour souverainistes (sic. – c'est la droite Républicaine), 2 pour Extrêmedroite et 3 pour la France insoumise!

Et son tweet a été "liké" par EU DisinfoLab...

Voici le contenu du fichier :

Et ici plus de très gros comptes classés par DisinfoLab comme "France Insoumise" (qui nous excuseront, nous l'espérons, de cette indication) :



Ce fichier qui n'a jamais existé contient ces 13 données en colonne :

- A. l'username / libellé du compte (le @),
- B. le nom public,
- C. le nombre de comptes suivis,
- D. le nombre de followers,
- E. le nombre de tweets du compte,
- F. la biographie du profil,
- G. la "classe" politique attribuée par Disinfo
- H. "Présent russe" (russophile ?)
- I. Présent dans la diffusion "Macronleaks"
- J. Nombre de "Désinfo russe"
- K. nombre de "Rumeurs présidentielle"
- L. "Nombre de rumeurs Benalla propagées"
- M. le nombre de Tweets et Retweets sur Benalla

Ce sont ces deux fichiers que Nicolas Vanderbiest (fondateur du EU DisinfoLab) a supprimés.

Ce n'est qu'ensuite que le compte EU DisinfoLab a publié les 2 autres fichiers – les

mêmes mais sans les données personnelles de la biographie ni le classement politiques (fichiers "Data brutes 1 – 47.xlsx" et "Rumeurs & items.xlsx"). Avant de le supprimer rapidement – vu qu'ils comprennent les identifiants des comptes (source) :

La publication des données brutes de l'étude a été réalisée afin de répondre aux nombreuses demandes publiques de transparence. Ces demandes ayant été satisfaites, nous avons procédé aux diligences nécessaire auprès de Free (l'hébergeur de ces données) pour en obtenir la suppression.

Ainsi, il y a bien eu **4 fichiers diffusés par cette officine**, **pas 3**. Plus aucun n'est accessible (il faudrait donc qu'ils arrêtent de dire qu'on peut vérifier facilement dans leurs tweets, jouant de la confusion)

Nous avons déjà expliqué tout ceci dans ce billet de jeudi dernier.

Et Disinfo a publiquement reconnu ce problème de fichiers (source – d'autres exemples ici sur les fichiers "officiels") :



Ah, il y a des fichiers "non officiels" – qui sortent d'où...?

Ils avaient même répondu à Aude Lancelin (source) :

Lucas @lucasgautheron · 9 h

Faudra m'expliquer ce que je fiche dans votre colonne « présent - russe » alors.

© 2 17 5 © 28 

EU DisinfoLab

Suivre



Le fichier dropbox n'est pas le fichier original. Il y a eu un souci de colonne pour celle-ci. Le seul fichier qui compte est ici :

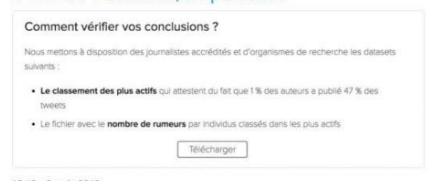


Oui un petit "souci de colonne"... Mais c'est à cause de nous peut-être ?

Et ceci? (source)



Nous avons mis en ligne 2 fichiers pour être transparents. Le fait de figurer dans un fichier signifie que vous avez été très actifs pendant l'affaire Benalla. Rien d'autre. Par ailleurs, le fichier dropbox qui circule n'est pas l'officiel. L'officiel : dl.free.fr/hUpdedh0l





Ah, ce fichier Dropbox donc?



Mais qui donc a bien pu le diffuser ???

Ainsi, en résumé : Nicolas Vanderbiest, le cofondateur et chercheur du DisinfoLab a diffusé 2 fichiers :

• "Data brutes.csv" : contenant de nombreuses informations personnelles

- sensibles récupérées des profils publics de Twitter, dont la biographie, pour **55000** comptes ;
- "ActeursClassés.xlsx": contenant les mêmes informations personnelles sensibles, complétées d'une classification politique qu'il a lui même réalisée (en indiquant un chiffre 1, 2 ou 3), pour plus de 3000 comptes.

Il les a supprimés au bout de plusieurs jours. Le compte DisinfoLab a ensuite diffusé 2 autres fichiers ("Data brutes 1-47.xlsx" et "Rumeurs & items.xlsx"), avec bien moins de données personnelles, avant, également, de les supprimer.

Afin que tout le monde comprenne la gravité de ce genre de fichiers, nous avons donc réalisé notre tweet avec un extrait des deux premiers fichiers, diffusés par Nicolas Vanderbiest.

## V. Fact-checking collaboratif du tableau de notre tweet

Le mieux est de laisser les twittos qui disposent du fichier original "Data brutes.csv" (à ne pas confondre avec le fichier épuré "Data brutes 1 – 47.xlsx") de vérifier notre travail – que nous republions, en rajoutant les numéros de ligne du fichier original pour une vérification aisée – les voici :





Cliquez pour agrandir.

N.B. attention sur quelques comptes, il y a des sauts de ligne dans la biographie, veillez donc à bien aller au bout de la cellule si besoin. Nous avons explicité dans la capture le titre de la colonne sur les Tweets Benalla ("in filtergetvalue tweet count")

Nous mettrons les copies d'écran en réponse ici.



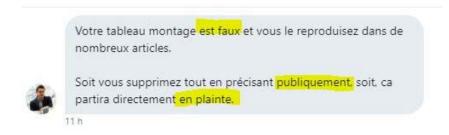
### VI. Plainte

Le 10 aout, Nicolas Vanderbiest présente ses excuses, retweeté par DisinfoLab :



Au vu de tout ceci, on comprendra notre étonnement quand nous avons reçu ce matin, 11 aout à à 9h00, ce message diffamatoire et menaçant de Nicolas Vanderbiest :





En réponse, je lui ai fait part de mon étonnement, et je lui ai demandé ce qu'il considérait comme faux :



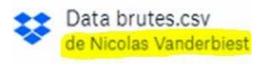
Je n'ai reçu aucune réponse de sa part 8 heures après.

Je reste stupéfait, car, non seulement j'ai vu son fichier "Data Brutes.csv", mais, plus fort, s'il l'a supprimé sur Dropbox (qui est un hébergement en ligne) depuis 3 jours, il se trouve que j'ai toujours dans mon navigateur l'onglet ouvert sur ce même fichier Dropbox (qui n'a pas encore été effacé, n'ayant pas actualisé la page ; et il est impossible de modifier ce fichier) :





On voit que ce fichier est rempli de données personnelles sensibles de 55 000 personnes – et qu'il comprend le nom du titulaire du compte qui l'a mis en ligne :



Bref, comme il me demandait une mise au point publique, elle est donc faite. De rien.

Il est quand même stupéfiant que **Twitter ait donné 125 000** \$ à cette officine pour "lutter contre la désinformation"! Nous allons leur demander des comptes au vu de l'énorme désinformation produite.

### Financement:

Twitter: à la suite des manipulations survenues pendant la campagne américaine de 2016, et qui impliquait l'achat de publicités par Russia Today et Sputnik, Twitter s'est engagé le 26 octobre 2017\*(6) à redistribuer les 1,9 millions de dollars vers des organisations luttant contre la désinformation. À ce titre, le EU DisinfoLab a reçu 125 000 \$ en janvier 2018.

En plus c'était l'argent des publicités achetées à Twitter par RT et Sputnik! #Kafka

Du coup, il est bien évident que nous portons plainte contre l'officine EU DisinfoLab pour diffamation :



Monsieur Olivier Berruyer

Paris, le 11 août 2018

#### Nos Réf : BERRUYER c/ EU DISINFOLAB

Cher Olivier,

Vous m'avez fait suivre les tweets rédigés ce jour par l'association EU DISINFOLAB vous accusant de la fabrication et de la diffusion de faux dans l'affaire des fichiers diffusés par celle-ci.

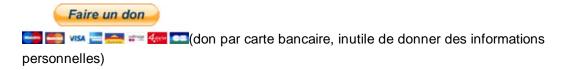
Au vu des éléments complémentaires que vous m'avez communiqués, il semble que les affirmations de cette association sont à mon sens constitutives d'une diffamation, répréhensible au visa de l'article 29 alinéa 1<sup>er</sup> de la Loi du 29 juillet 1881.

Je vous confirme donc que, selon vos instructions, je saisis les juridictions répressives via une plainte avec constitution de partie civile à l'encontre de l'association EU DISINFOLAB, dont je vous soumettrai le projet dans les meilleurs délais.

Veuillez croire, cher Olivier, à mes sentiments cordiaux et respectueux.

Jérémie ASSOUS Avocat à la cour

Vous pouvez participer aux frais de justice par un paiement à notre avocat ici (même modeste). Merci d'avance.



P.S. au vu de ce comportement hautement suspect, **nous continuerons bien entendu nos investigations sur cette officine.** 

(Billet édité)

### Dossier: #FichagePolitique - EU Disinfo Lab

#1 #2 #3 #4 #5

Communiqué de presse : EU DisinfoLab, la diffusion scandaleuse d'un démenti mensonger Un laboratoire du CNRS aurait aussi fiché les opinions politiques de près de 200 000 personnes! #FichagePolitique? Tous fichés ? Comment Twitter permet le fichage politique de la population

L'étude de l'influence russophile sur le Benallagate : d'énormes failles méthodologiques Le EU DisinfoLab à l'origine d'une intox : l'affaire Benalla, produit de la « russosphère » ?

# LES ERISES

LIBERTÉS PUBLIQUES 11.août.2018 // Les Crises

Un laboratoire du CNRS aurait aussi fiché les opinions politiques de près de 200 000 personnes! #FichagePolitique?



Nous avons largement démonté l'étude préliminaire, puis l'étude finale comportant d'énormes failles méthodologiques du EU DisinfoLab sur l'affaire Benalla.

Ce billet reprend une partie de l'étude sur le fichage politique, axée sur le rôle d'un des laboratoires du CNRS, pour une meilleure visibilité (inutile de le relire si vous avez déjà lu l'étude sur le fichage politique de jeudi). Contacté par nos soins, ni lui ni le CNRS ne nous a répondu...

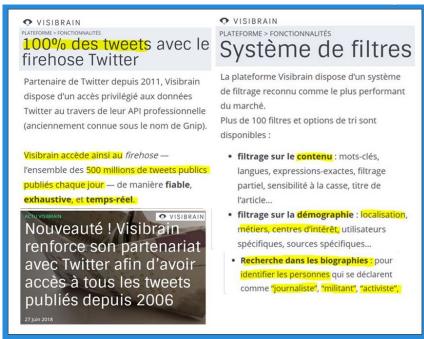
- I. Visibrain, ou Big Brother à la maison
- II. Que prévoient les conditions d'utilisation Twitter?
- III. Un laboratoire du CNRS aurait aussi fiché les opinions politiques de près de 200 000 personnes!
- IV. Réglementation et Discussion
- V. Plainte

# I. Visibrain, ou Big Brother à la maison

12/08/2018 10:46 1 of 38

Eu DisinfoLab a utilisé un outil très puissant (et coûteux) permettant d'analyser Twitter – il s'agit ici du logiciel Visibrain , qui sert normalement à des analyses marketing pour les grandes entreprises :





L'essentiel est dans le graphique précédent. Si vous voulez en savoir plus sur ce logiciel, vous pouvez dérouler ici une analyse détaillée : (elle est affichée par défaut)





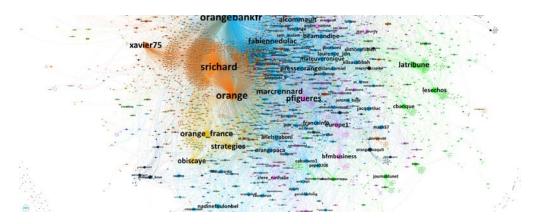
et permet de réaliser en fait automatiquement les graphiques que présente Nicolas Vanderniest:

### VISIBRAIN INFLUENCEURS Cartographie

La plateforme Visibrain possède un **format** d'export unique sur le marché : l'export GEXF, le format de données du logiciel de cartographie open-source Gephi. En combinant la puissance de Gephi et Visibrain, vous pouvez réaliser en quelques clics des cartographies complètes des relations entre utilisateurs, en particulier sui des sujets à forte volumétrie. Grâce aux cartographies Visibrain + Gephi, vous identifiez les différentes communautés autour d'un sujet et comprenez les relations d'influence.



12/08/2018 10:46 3 of 38



En fait, ce logiciel (dont nous allons détailler les caractéristiques, car c'est très important) donne accès... à tous les tweets de Twitter, en temps réel (sources : 1, 2, 3) :

# visibrain PLATEFORME > FONCTIONNALITÉS 100% des tweets avec le firehose Twitter

Partenaire de Twitter depuis 2011, Visibrain dispose d'un accès privilégié aux données Twitter au travers de leur API professionnelle (anciennement connue sous le nom de Gnip).

Visibrain accède ainsi au *firehose* — l'ensemble des 500 millions de tweets publics publiés chaque jour — de manière **fiable**, **exhaustive**, et **temps-réel**.



La plateforme de veille Visibrain vous permet de veiller l'ensemble des réseaux sociaux et

4 of 38 12/08/2018 10:46

### plus particulièrement :

- Twitter : Visibrain dispose d'un accès au firehose qui garantit la récupération de 100% des tweets publics, en temps-réel
- Facebook : Visibrain dispose d'un module dédié à la veille des pages publiques Facebook
- Instagram & YouTube : Visibrain permet également de veiller les réseaux sociaux visuels comme Instagram ou YouTube.



Le temps-réel est dans l'ADN de la plateforme Visibrain:

- · dès qu'un contenu est publié, il est immédiatement analysé et disponible dans Visibrain, ce qui vous permet d'avoir en permanence des données fraîches
- le système d'alerte de Visibrain vous garantit une notification en moins d'une
- l'interface de la plateforme s'actualise instantanément

On peut largement filtrer, en particulier par "métiers et centres d'intérêts" (source) :



La plateforme Visibrain dispose d'un système de filtrage reconnu comme le plus performant du marché.

Plus de 100 filtres et options de tri sont disponibles:

- filtrage sur le contenu : mots-clés, langues, expressions-exactes, filtrage partiel, sensibilité à la casse, titre de l'article...
- filtrage sur la démographie : localisation, métiers, centres d'intérêt, utilisateurs spécifiques, sources spécifiques...
- filtrage sur la temporalité : granularité à l'année, au mois, à la semaine, au jour ou à la minute
- le tout combinable à l'infini avec l'ensemble des opérateurs booléens ET, OU, EXCLUSION, et PARENTHESES

On compte explicitement dans les Métiers "journaliste" ou... "militants politiques" (source)

### VISIBRAIN

ANALYSE DÉMOGRAPHIQUE

# Métiers

La plateforme Visibrain infère le métier des utilisateurs des réseaux sociaux grâce à différents indices, ce qui vous permet de voir en un clin d'oeil quelle typologie de population (journalistes, étudiants, militants politiques...) concorno votro cuist

12/08/2018 10:46 6 of 38

COLICELLIE AOU E PAIET

Comme tous les indicateurs de Visibrain, il est possible de cliquer sur un métier pour avoir une analyse ciblée sur cette population.

**OCCUPATIONS** journaliste en writer rédacteur en chef animateur journaliste musician secretaire général community manager journaliste politique pigiste rédactrice en chef rédacteur en chef adjoint directeur numérique de auteur consultant

Ah, quel est le métier le plus visé...?

De même pour les centres d'intérêts (source) :



La plateforme Visibrain vous affiche une synthèse des centres d'intérêt des utilisateurs qui ont posté sur votre sujet de veille. Cela vous permet de mieux comprendre quelle est la population qui intéragit avec votre marque (ou n'importe quel sujet de veille), et de découvrir de nouvelles opportunités de communication.

Comme tous les indicateurs de Visibrain, il est possible de cliquer sur un centre d'intérêt pour avoir une analyse ciblée sur cette population.

Comme on n'arrête pas le progrès, on peut même analyser les émojis (source) :

PLATEFORME > ANALYSE DU CONTENU

# **Emojis**

La plateforme Visibrain fait ressortir les emojis les plus utilisés dans les conversations sociales.

Vous identifiez ainsi en un clin d'oeil la tonalité

### d'un sujet.

Comme tous les indicateurs de Visibrain, il est possible de cliquer sur n'importe quel emoji pour comprendre quels posts sociaux se cachent derrière.

| emojis   | Number of mentions |       |
|----------|--------------------|-------|
|          |                    | 1 209 |
|          |                    | 1 027 |
|          |                    | 898   |
| <b>→</b> |                    | 756   |
| •        |                    | 729   |
| ÷.       |                    | 667   |
| (m)      |                    | 542   |

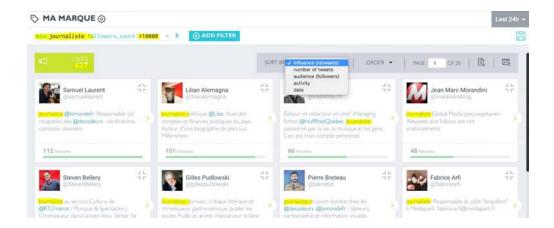
On peut aussi chercher dans les biographies – en particulier pour trouver les "journalistes, militants et activistes" (source):

> O VISIBRAIN PLATEFORME > FONCTIONNALITÉS Détection d'influenceurs Twitter

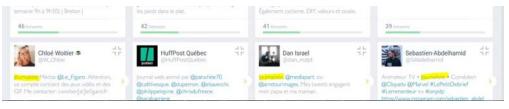
Grâce à Visibrain, vous détectez de nouveaux influenceurs sur vos thématiques, et vous identifiez avec précision les comptes Twitter

les plus influents sur un sujet donné. Visibrain vous permet de croiser plusieurs options de tri et de filtrages pour obtenir des listes d'influenceurs sur-mesure :

- · Tri par audience, influence ou activité : pour classer les comptes Twitter en fonction de leur nombre de followers, leur impact en terme d'engagement, ou leur nombre de tweets sur un sujet donné
- Recherche dans les biographies : pour identifier les personnes qui se déclarent comme "journaliste", "militant", "activiste", "blogueur", "youtubeur"...
- Localisation : pour cibler une zone géographique précise (ville, région ou pays)
- Informations de profils : pour filtrer les comptes qui correspondent à certains critères (compte vérifié, ratio following/followers, date de création du compte...)



12/08/2018 10:46 9 of 38



Oh des journalistes avec plus de 10 000 abonnés...

On pourra alors en faire des panels pour les suivre de très près (source) :



De même, le logiciel déterminera généralement votre localisation et votre genre (sources : 1 et 2):



La plateforme Visibrain infère la localisation des utilisateurs des réseaux sociaux grâce à

différents indices (informations de profil, contenu posté...), et délivre ainsi une information géographique sur plus de 60% des comptes!

Vous pouvez visualiser la propagation d'un sujet sur une carte, et filtrer les informations par pays, régions ou ville.



La plateforme Visibrain infère le genre des utilisateurs des réseaux sociaux grâce à différents indices (prénom, éléments du profil, contenu posté...), et délivre ainsi une information du genre (homme/femme) de l'utilisateur sur plus de 70% des comptes!

Ce logiciel Big Brother vous facilite la vie en vous permettant d'exporter toutes les données qui vous intéressent sous Excel (source) :

## VISIBRAIN **EXPORT** CSV (Excel)

La plateforme Visibrain vous permet d'exporter la totalité des indicateurs de la plateforme au format CSV (Excel). Les exports Visibrain sont complets, propres et directement exploitables pour retraiter les données ou les inclure dans un rapport.

12/08/2018 10:46 11 of 38

|    | A            | В           | С        | D            | E                 | F              | G    |     | Н            | L                | J            | K            | L                | M          |
|----|--------------|-------------|----------|--------------|-------------------|----------------|------|-----|--------------|------------------|--------------|--------------|------------------|------------|
| 1  | screenNa 🔻   | followers ▼ | verified | ▼ bio ▼      | first seen date ▼ | client =       | lang |     | link =       | text 🔻           | country      | region       | city 🔻           | retweets 🔻 |
| 5  | visibrain    | 8989        | TRUE     | La plateform | 24/10/2017 16:18  | Twitter Web    | fr   | 111 | https://twit | t Comment ex     | france       | île-de-fran  | ce paris         | 12         |
| 6  | CCIcotedazui | 13575       | TRUE     | RETROUVEZ    | 26/10/2017 07:09  | Twitter for A  | fr   |     | https://twit | t 🍬 @CClcot      | france       | provence-a   | alp nice         | 11         |
| 7  | visibrain    | 8993        | TRUE     | La plateform | 30/10/2017 14:36  | Twitter Web    | fr   |     | https://twit | t Quand peut-    | france       | île-de-fran  | ce paris         | 10         |
| 8  | fadouce      | 29999       | FALSE    | Com, stratég | 30/10/2017 07:35  | TweetDeck      | fr   |     | https://twit | t L'étonnante    | france       | île-de-fran  | ce paris         | 9          |
| 9  | CabonVro1    | 408         | FALSE    | La Fibre Ora | 03/11/2017 08:52  | Twitter for if | fr   |     | https://twit | t Les @amb_c     | france       | île-de-fran  | ce paris         | 8          |
| 10 | visibrain    | 8989        | TRUE     | La plateform | 24/10/2017 16:04  | Twitter Web    | fr   |     | https://twit | #LivreBlanc:     | france       | île-de-fran  | ce paris         | 6          |
| 11 | blendwebmi   | 6222        | FALSE    | #BlendWeb!   | 26/10/2017 14:45  | Twitter Web    | fr   |     | https://twit | t Déjà 3169 tv   | france       | rhône-alpe   | s lyon           | 6          |
| 12 | visibrain    | 8989        | TRUE     | La plateform | 24/10/2017 16:04  | Twitter Web    | fr   |     | https://twit | t Téléchargez    | france       | île-de-fran  | ce paris         | 6          |
| 13 | blendwebmi   | 6238        | FALSE    | #BlendWeb!   | 26/10/2017 18:23  | Twitter Web    | fr   |     | https://twit | t Les 5000 me    | france       | rhône-alpe   | s lyon           | 5          |
| 14 | visibrain    | 8993        | TRUE     | La plateform | 30/10/2017 14:36  | Twitter Web    | fr   |     | https://twit | t Qu'est-ce-qu   | france       | île-de-fran  | ce paris         | 5          |
| 15 | visibrain    | 8989        | TRUE     | La plateform | 27/10/2017 14:40  | Buffer         | fr   |     | https://twit | [BAD BUZZ]       | france       | île-de-fran  | ce paris         | 4          |
| 16 | marc_cogrel  | 2661        | FALSE    | Responsable  | 31/10/2017 17:25  | Twitter Web    | fr   |     | https://twit | t Voici le top # | france       | île-de-fran  | ce paris         | 4          |
| 17 | visibrain    | 8998        | TRUE     | La plateform | 25/10/2017 11:53  | Twitter Web    | fr   |     | https://twit | t Comment #F     | acebook déte | ermine les p | ublications à af | 4          |
| 18 | visibrain    | 8992        | TRUE     | La plateform | 31/10/2017 13:51  | TweetDeck      | fr   |     | https://twit | Les twittos s    | france       | île-de-fran  | ce paris         | 4          |
| 19 | AgenceADeq   | 8976        | FALSE    | Développem   | 30/10/2017 09:12  | Twitter Web    | fr   |     | https://twit | t "Les 580 twe   | france       | midi-pyrén   | é toulouse       | 4          |
| 20 | agence636    | 1026        | FALSE    | WE V #SOC    | 30/10/2017 14:04  | Twitter Web    | fr   |     | https://twit | t #Apple twee    | france       | midi-pyrén   | é toulouse       | 4          |
| 21 | visibrain    | 8992        | TRUE     | La plateform | 31/10/2017 11:00  | Twitter Web    | fr   |     | https://twit | t 10 comptes     | france       | île-de-fran  | ce paris         | 4          |
| 22 | suebecks     | 10414       | FALSE    | National Tea | 07/11/2017 11:38  | Twitter Web    | en   |     | https://twit | Gaining Pow      | united kingd | england      | sheffield        | 3          |
| 23 | Alexleserveu | 6019        | TRUE     | #SocialMedia | 31/10/2017 11:04  | TweetDeck      | fr   |     | https://twit | Quand je vo      | france       | île-de-fran  | ce paris         | 3          |
| 24 | GuillaumePe  | 112         | FALSE    | Chargé de de | 03/11/2017 11:25  | Twitter Web    | fr   |     | https://twit | @AmDalmas        | france       | île-de-fran  | ce paris         | 3          |
| 25 | LettreAudio  | 4479        | FALSE    | Premier jour | 27/10/2017 16:25  | Twitter for if | fr   |     | https://twit | https://t.co/    | france       | île-de-fran  | ce paris         | 3          |
| 26 | visibrain    | 8993        | TRUE     | La plateform | 05/11/2017 18:08  | Buffer         | fr   |     | https://twit | A quel mom       | france       | île-de-fran  | ce paris         | 3          |

Vous pouvez aussi exporter tous les followers, en les croisant (source) :

# ▼ VISIBRAIN PLATEFORME > INFLUENCEURS Extraction de followers

Avec Visibrain, vous pouvez exporter

l'exhaustivité des followers de n'importe quel

compte au format Excel : les followers de votre marque, ceux de vos concurrents, ceux d'un

influenceur...

Visibrain vous permet également d'effectuer

des croisements de liste de followers pour

identifier quels followers vous avez en

commun avec vos concurrents, les comptes

qui suivent vos concurrents sans vous suivre,

ou bien les comptes qui suivent l'ensemble

des influenceurs de votre secteur.

| В                   | E   | F                      | G                  | H              | 1              | 1           |        |
|---------------------|---|------------------------|--------------------|----------------|----------------|-------------|--------|
| Nom d'utilisateur - | Biographie +                                      | Site Web 🔻             | Date de création 🔻 | Nb followers + | Nb following - | Nb listed + | Nombre |
| emportune_n         | are rimportune serectionine des tweets du me      | mesp://immportune.ii   | 10/03/2013 10:23   | 20022          | 2202           | FFEO        |        |
| C8TV                | ps émissions : #TPMP #WAM #SLT #LTD #Famil        | http://www.c8.fr       | 20/05/2009 09:53   | 767064         | 1469           | 983         |        |
| TopAchat            | icier du #Geek depuis 1999 ! Grosse command       | http://www.topachat.   | 19/01/2010 09:42   | 410177         | 9393           | 647         |        |
| bouyguestelecom     | uygues Telecom partage votre passion de la te     | http://www.bouygues    | 02/10/2008 12:02   | 143808         | 65237          | 790         |        |
| SNCF                | ut l'univers SNCF à portée de clic : les infos en | http://www.sncf.com    | 22/11/2010 23:53   | 108644         | 957            | 725         |        |
| Le_Figaro           | Actualités, économie, politique, culture et div   | http://www.lefigaro.fr | 22/08/2007 09:16   | 2800055        | 537            | 14510       |        |
| Europe1             | ute l'information d'@Europe1 et nos émission      | http://www.europe1.f   | 23/04/2009 10:53   | 1311255        | 3542           | 5271        |        |
|                     |   |                        |                    |                | 2000           | 72722       |        |

| ZUIVIIIIULES   | formation en continu. Actualités, Politique, Spinttp://www.zuminutej   | 28/03/2007 14:14 | 2306909 | 1403  | 3100  |
|----------------|--|------------------|---------|-------|-------|
| LesEchos       | ctualité économique, financière et boursière (http://www.lesechos.     | 25/11/2009 18:01 | 1066877 | 500   | 7197  |
| LCI            | envenue sur LCI, sur le canal 26 de la TNT. Le nhttp://lci.fr/         | 23/03/2009 23:45 | 109113  | 964   | 1457  |
| BFMTV          | re chaîne info de France. Retrouvez-nous sur \$http://www.bfmtv.cor    | 16/04/2010 11:55 | 2267084 | 488   | 8310  |
| OuestFrance    | mpte Twitter du quotidien Ouest-France. Rechttp://www.ouest-france.    | 27/04/2009 17:59 | 518358  | 3518  | 3124  |
| romandie       | mpte officiel de http://Romandie.com, média http://www.romandie        | 27/03/2009 13:43 | 25117   | 2     | 591   |
| Niska_Officiel | puvel album #Commando maintenant disponit http://www.charo-sho         | 02/11/2011 01:37 | 383839  | 36    | 247   |
| lalibrebe      | mpte officiel de Lalibre.be Rejoignez-nous éga http://www.lalibre.be   | 04/02/2009 11:37 | 203785  | 2896  | 2155  |
| lemondefr      | ctualité de référence par la rédaction du Mon http://www.lemonde.      | 16/03/2009 19:44 | 7511211 | 504   | 33063 |
| AmazonHelp     | e answer Amazon support questions 7 days a http://www.amazon.c         | 28/10/2009 05:17 | 124674  | 52    | 741   |
| Orange_conseil | sez vos questions du lundi au samedi, de 8h à 20h dans les domaines    | 13/10/2010 11:16 | 88857   | 44182 | 431   |
| afpfr          | prceaux choisis des fils d'actu, coulisses & RT chttp://www.afp.com    | 20/07/2011 14:07 | 2855687 | 1348  | 12990 |
| AmazonFR       | mpte Officiel Twitter http://www.Amazon.fr. http://amazon.fr           | 30/09/2015 14:07 | 21722   | 1     | 61    |
| Qc511_QcLevis  | cidents & entraves sur le réseau autoroutier de http://quebec511.info  | 31/03/2014 15:37 | 3001    | 15    | 59    |
| Sosh_fr        | sh, la marque sans engagement par Orange, r http://www.sosh.fr/        | 19/07/2011 11:35 | 138831  | 29549 | 578   |
| tvanouvelles   | a télé, sur votre ordinateur ou votre appareil (http://tvanouvelles.ca | 26/10/2009 15:59 | 343378  | 538   | 2174  |
| PSchydlowski   | rtains mettent à jour leur statut @facebook toutes les 5 secondes, ils | 09/01/2011 11:42 | 41003   | 20691 | 280   |
| le Parisien    | Parisien-Aujourd'hui en France, 1er guotidien http://www.leparisien    | 06/05/2009 10:50 | 1958055 | 910   | 7770  |

Et, si vous le souhaitez (et payez), vous pouvez même avoir accès à l'historique Twitter depuis 2006 (source)!

> VISIBRAIN Archives et données passées

La plateforme Visibrain vous permet de récupérer instantanément 30 jours de données passées sur un nouveau sujet ("backdata" / "données rétroactives"). Visibrain dispose également d'un accès aux archives complètes de Twitter depuis 2006 (sur demande).

Car bien sûr le logiciel est payant - comptez quelques milliers d'euros par mois pour récolter quelques centaines de milliers de tweets...

Comme vous le savez peut-être, le 25 mai 2018, le règlement européen général sur la protection des données (RGPD) est entré en application pour augmenter la protection des données des citoyens.

Eh bien c'est à peine le 27 juin dernier que Twitter a décidé de donner accès à ses archives à Visibrain (source) :

12/08/2018 10:46 13 of 38



Suite à un renforcement de notre partenariat avec Twitter, nous avons pu améliorer significativement notre technologie « instaback » de récupération des tweets dans le passé : la plateforme Visibrain permet désormais de récupérer rétroactivement les données Twitter sur n'importe quel sujet de votre choix depuis mars 2006, jusqu'au tout premier tweet fait par @jack Dorsey.

## II. Que prévoient les conditions d'utilisation Twitter?

À ce stade, vous devez commencer à être effrayé de la masse d'informations mise à disposition – et de ce qu'on peut en faire...

Si vous vous demandez ce que prévoit Twitter à ce sujet, voici une sélection des conditions (source {3}) :



Vos droits et Concession

Vous conservez vos droits sur tout Contenu que vous soumettez, publiez ou affichez sur ou via les Services. Ce qui est à vous vous appartient. Vous êtes le propriétaire de votre Contenu (ce qui inclut vos sons, photos et vidéos intégrés).

Bonne nouvelle vous conservez vos droits sur tout votre contenu!

#### Mais bon...:

En soumettant, en publiant ou en affichant un Contenu sur ou via les Services, vous nous accordez une licence mondiale, non exclusive et libre de redevances (incluant le droit de sous-licencier), nous autorisant à utiliser, copier, reproduire, traiter, adapter, modifier, publier, transmettre, afficher et distribuer ce Contenu sur tout support et selon toute méthode de distribution (actuellement connus ou développés dans le futur). Cette licence nous autorise à mettre votre Contenu à disposition du reste du monde et autorise les autres à en faire de même. Vous convenez que cette licence comprend le droit pour Twitter de fournir, promouvoir et améliorer les Services et de mettre le Contenu soumis sur ou via les Services à disposition d'autres sociétés, organisations ou personnes privées, aux fins de syndication, diffusion, distribution, promotion ou publication de ce Contenu sur d'autres supports et services, sous réserve de respecter nos conditions régissant l'utilisation de ce Contenu. Twitter, ou ces autres sociétés, organisations ou personnes privées, pourront utiliser ainsi le Contenu que vous aurez soumis, publié, transmis ou de quelque autre façon mis à disposition via les Services sans que vous puissiez prétendre à une quelconque rémunération au titre de ce Contenu.

Twitter et ses clients en font ce qu'ils en veulent et vous ne pouvez vous y opposer!

### Car tout est public (source):

# Informations publiques

Sur Twitter, la plupart des activités sont publiques, y compris les informations de votre profil, votre fuseau horaire et votre langue, la date de création de votre compte, vos Tweets et certaines informations sur vos Tweets, telles que la date, l'heure, l'application et la version de Twitter à partir desquelles vous avez Tweeté. Vous pouvez également décider de publier votre emplacement dans vos Tweets ou sur votre profil Twitter. Les listes que vous créez, les personnes que vous suivez et celles qui vous suivent, les Tweets que vous Aimez ou Retweetez sont également des informations publiques. Les diffusions Periscope que vous

### Et donc ils vous conseillent :

Vous êtes responsable de vos propres Tweets et des autres informations que vous fournissez par le biais de nos services, et vous devez réfléchir avec soin à ce que vous décidez de p ublier 1, surtout s'il s'agit d'informations sensibles. Si vous

### Car en plus de l'utilisation du site Twitter :

En plus de fournir directement vos informations publiques au monde sur Twitter, nous avons également recours à des technologies, telles que les interfaces de programmation d'application (API) et les éléments incorporés, pour mettre ces informations à la disposition de sites Web, d'applications et d'autres utilisateurs, par exemple en affichant des Tweets sur un site Web d'informations ou en analysant ce que disent les gens sur Twitter. En général, nous mettons ce contenu à disposition gratuitement s'il s'agit de quantités limitées et nous faisons payer des droits pour tout accès à grande

échelle. Nous avons élaboré des conditions standard régissant la façon dont ces données peuvent être utilisées et

ils utilisent d'autres interfaces (API) pour diffuser en masse les contenus, moyennant rémunération. C'est le cas avec Visibrain.

Twitter insiste (source):

Twitter est public et les Tweets peuvent immédiatement être vus et faire l'objet de recherches par qui que ce soit dans le monde entier. Nous mettons également à votre disposition

d'autant plus que c'est Twitter qui vend l'accès aux API qu'il a développées pour permettre de telles recherches... (source) :

# À propos des API Twitter

Twitter reflète ce qui se passe dans le monde et ce dont les gens parlent en temps réel. Vous pouvez accéder à Twitter via le Web ou sur votre appareil mobile. Pour partager les informations sur Twitter aussi largement que possible, nous fournissons aussi aux entreprises, aux développeurs et aux utilisateurs un accès programmatique aux données Twitter via nos API (interfaces de programmation d'applications). Cet article explique ce que sont les API Twitter et

### Tweets et réponses

Nous mettons les Tweets et réponses publics à la disposition des développeurs, et nous les autorisons à publier des Tweets via notre API. Les développeurs peuvent accéder aux Tweets en recherchant des mots-clés donnés, ou en demandant un échantillon de Tweets de comptes spécifiques.

Ces points de terminaison sont utilisés par des ONG et des organisations comme les Nations Unies pour identifier,

12/08/2018 10:46 17 of 38

comprendre et contrer les fausses informations relatives aux initiatives de santé publique. Par exemple, en Indonésie, des rumeurs persistantes faisaient courir le bruit que les vaccins contenaient du porc ou entraînaient la stérilité. Comprendre la façon dont ces rumeurs ont été lancées et propagées a permis aux Nations Unies d'envoyer une équipe sur le terrain pour dissiper ces mythes, à l'origine de préoccupations importantes dans cette nation à majorité musulmane. De la même façon, nous aidons les chercheurs à repérer les premières indications témoignant d'épidémies et à suivre leur propagation. Une équipe de l'université Northeastern a récemment développé une nouvelle technique de suivi de la grippe en utilisant des données Twitter qui prévoient les épidémies jusqu'à six semaines à l'avance, bien avant de nombreux autres modèles, sans toutefois sacrifier la précision.

Donc Twitter indique qu'il vend vos contenus, par exemple "aux ONG", aux "Nations-Unies", par exemple pour lutter contre le "complotisme anti-vaccins" (dans les pays "à majorité musulmane") ou lutter contre "les épidémies" telles "la grippe".

#### C'est beau...

Bon, ok, dans l'écrasante majorité des cas, ce sera des entreprises telles Boeing pour savoir ce que les twittos disent d'elles (voire des "chercheurs" voulant connaître vos opinions politiques) - mais on imagine qu'il ne faut pas se plaindre, puisqu'on vous dit que ca permet à des gens de ne pas mourir... Et Twitter est sympa (source) :

### Messages Privés

Nos points de terminaison de Messages Privés fournissent un accès aux conversations par MP des utilisateurs qui ont explicitement accordé une autorisation à une application spécifique. Nous ne vendons pas les Messages Privés.

Il ne vend pas vos messages privés! (les messages publics étant bien entendu vendus, eux). Car attention, Twitter a des "valeurs fondamentales" (source) :

12/08/2018 10:46 18 of 38

# Défendre et respecter les droits de nos utilisateurs

Défendre et respecter la voix des utilisateurs est l'une de nos valeurs fondamentales chez Twitter. Avec cette valeur, nous nous engageons en faveur de la liberté d'expression et de la confidentialité. La transparence constitue également une part importante de cet engagement. Dans cet esprit, Twitter souhaite rendre publique la manière dont nous traduisons cette valeur clé dans nos prises de décisions.

Et la "confidentialité" en fait partie. Ils font donc très attention dans leurs "prises de décision" - comme, par exemple, quand ils autorisent Visibrain à accéder à l'historique Twitter depuis 2006... Mais c'était prévu (source) :

# Partager vos données avec les partenaires commerciaux de **Twitter**

Lorsque nous recevons des données vous concernant, nous utilisons ces informations pour améliorer nos services et personnaliser votre expérience Twitter, ainsi qu'à d'autres fins décrites dans notre Politique de confidentialité. Dans certains cas, cela implique de partager des informations non publiques avec nos partenaires.

En conclusions, certains d'entre vous se diront peut-être : "Mais où le mal, ce sont des données publiques ?". Eh bien nous illustrons le problème dans la partie suivante...

### III. Un laboratoire du CNRS aurait aussi fiché les opinions politiques de près de 200 000 personnes!

Dans le cadre de nos recherches, nous sommes tombés sur cet article des *Décodeurs* du Monde du 4 décembre 2017 :

# Le blog du Décodex

UN BLOG DE LA RÉDACTION - Le Monde

# Comment des chercheurs ont utilisé les données du Décodex pour analyser la présidentielle

Sciences Po a réuni chercheurs, développeurs et journalistes pendant quatre jours pour exploiter différents jeux de données liés à l'élection présidentielle, un événement dont nous étions partenaires.

LE MONDE | 04.12.2017 à 15h32 • Mis à jour le 23.01.2018 à 17h16 |

Par Adrien Sénécat

Adrien Sénécat nous indique qu'ils ont pu croiser "une dizaine de bases de données sur la période électorale", lors de l'évènement Datapol :

Plus d'une dizaine de bases de données sur la période électorale ont été proposées aux participants, dont celle du Décodex, sous licence ODbL. Concrètement, nous avons apporté deux types de données :

- l'annuaire des sources du Décodex, qui fournit une typologie des sources d'informations en fonction de leur fiabilité et donne des informations sur chacune des références étudiées (les données sont accessibles ici au format JSON);
- 2 notre base de données des fausses informations qui ont circulé en ligne et des liens qui les ont fait circuler (les données sont accessibles ici au format JSON).

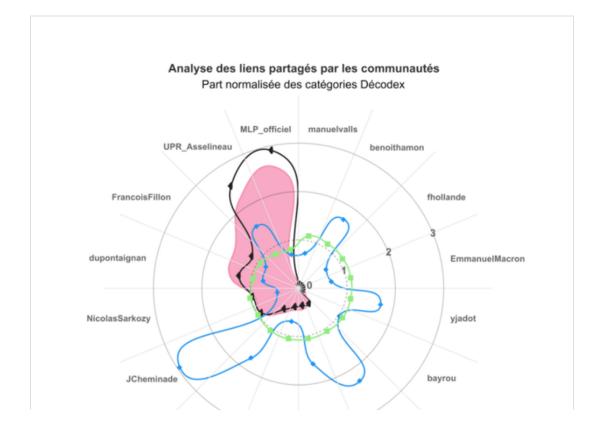
En croisant nos données avec d'autres bases et en faisant appel aux connaissances de participants venus d'horizons différents du nôtre, Datapol nous a permis de faire plusieurs découvertes intéressantes. En voici trois.

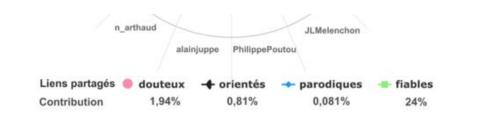
Ils ont réalisé une première étude sur "les fausses informations" (définies selon les critères méthodologiques du... Décodex). Puis une seconde qu'ils titrent "Les partisans de Marine Le Pen partagent plus de sources peu fiables que les autres" :

"Un autre enseignement intéressant est apparu en croisant les données du Décodex avec celles du Politoscope. [...] À partir de cette typologie et des données anonymisées du Politoscope, les participants à Datapol ont donc pu regarder dans quelle mesure les différentes communautés politiques partagent plus ou moins de liens vers les différents types de sources d'information identifiées dans le Décodex."

### L'infographie du Monde est ainsi construite :

- chaque candidat figure sur une tranche;
- plus un point est éloigné du 0, plus la communauté proche du candidat correspondant a partagé le type de sources en question :



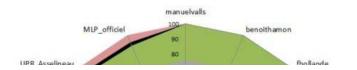


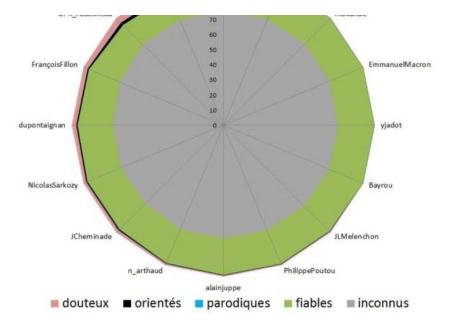
"Il apparaît ainsi que les partisans de Marine Le Pen et François Asselineau sont ceux qui, en moyenne, ont partagé le plus de liens vers des sources considérées comme peu fiables dans le Décodex. Une analyse qui mériterait d'être affinée par un travail approfondi, mais qui a le mérite d'apporter des données pour accréditer un comportement identifié par de nombreux observateurs pendant la campagne."

Avant de continuer sur notre axe, signalons quelques interrogations méthodologiques. D'abord, rappelons que la population des personnes inscrites sur Twitter n'est clairement **pas représentative** de la population générale (avec 49 % de CSP+ chez les 25-49 ans par exemple – source).

En fait le graphique a utilisé une présentation "douteuse". On a l'impression en regardant le graphique que les "communautés pro-Le Pen" et "pro-Asselineau" auraient inondé Twitter de liens vers des sites classés "très peu fiables". Or, les *Décodeurs* indiquent dans l'article que "*lorsqu'un point est proche du cercle de rayon 1, cela veut dire que le type de contenus correspondant a été partagé dans les mêmes proportions que la moyenne. Lorsqu'il est proche du trois, cela veut dire qu'il a été trois fois plus partagé." Ils représentent donc un simple écart à la moyenne, d'un phénomène marginal, d'où les forts écarts observés, qui sont en fait peu représentatifs. Ceci apparaît en fait clairement quand on observe la courbe verte peu visible de la diffusion de sites "fiables" : il y a assez peu d'écarts entre les candidats en réalité.* 

Illustrons. Nous avons reconstitué à la main les données du diagramme des *Décodeurs*. On peut dès lors reconstituer simplement ce que donnerait un graphique plus honnête représentant la diffusion de liens vers les sites classés par le *Décodex*:





Ce graphique donne donc une impression fort différente au lecteur de l'activité des personnes sur Twitter...

Mais en réalité, le fond de notre problématique est : mais comment savent-ils qui est partisan de qui?

Revenons donc au Politoscope de L'Institut des Systèmes Complexes Paris-Île-De-France (ISC-PIF, du CNRS). L'article insiste bien : "Dans le cadre de ce projet, l'équipe de l'Institut des systèmes complexes Paris Ile-de-France, un laboratoire du CNRS, a analysé sur la durée de la campagne les messages de milliers d'internautes sur Twitter. L'un des intérêts du Politoscope est qu'il identifie la proximité d'un utilisateur de la plate-forme avec tel candidat au fil du temps". Voilà ce qu'on trouve sur le site dédié du Politoscope qui explique ceci :

# Pourquoi un politoscope?

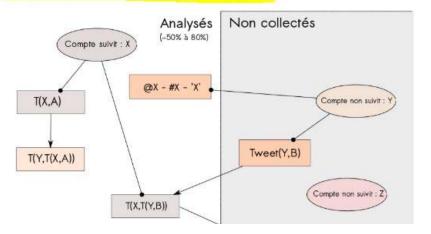
Twitter est devenu un nouveau média d'information ainsi qu'un passage obligé pour tout groupe qui cherche à influencer l'opinion

publique d'une manière ou d'une autre, y compris pour certains, en publiant de fausses informations (« fake news »).

Le but du *Politoscope* est d'inverser les rôles en permettant au "grand public" de plonger dans les masses de données générées sur les réseaux sociaux à l'approche des présidentielles grâce à des outils et méthodes d'analyses produites par la recherche.

Depuis août 2016, l'ISC-PIF analyse les interactions de plus de 3000 comptes Twitter liés à la campagne présidentielle française 2017 (élus, partis, candidats): (re)tweets, mentions, citations, etc. Une plateforme collecte et traite ainsi chaque mois plusieurs millions d'événements auxquels ont participé au total plus d'un million d'utilisateurs uniques. Une série d'analyses portant sur la structure des interactions et leur contenu est présentée dans ce dispositif, qui fait partie du *Politoscope*. L'objectif est de resituer dans le temps des événements perçus dans le flux continu de l'actualité. C'est une plongée inédite dans les différentes temporalités de la campagne présidentielle qui est proposée ici.

Le Politoscope suit depuis août 2016 plus de 3000 comptes Twitter appartenant à des médias et des personnalités du monde des médias, ainsi que certains flux ciblés thématiquement à partir de mots-clés et hashtags dans la limite du 1% Twitter (ex. #présidentielle2017). Plusieurs centaines de milliers de tweets politiques sont ainsi récoltés et analysés chaque jour.





**Résumé.** À l'occasion de l'élection présidentielle française de 2017, nous avons développé le Politoscope, une plateforme ouverte d'analyse des tweets politiques qui a collecté et analysé près de 60 millions d'échanges Twitter entre plus de 2,4 millions d'utilisateurs interagissant avec les personnalités politiques françaises. Dans cet article, nous nous intéressons

En fait, dans cette étude méthodologique, ils expliquent qu'ils sont partis de 3 700 comptes Twitter de figures politiques françaises, dont ils recueillent ainsi les tweets, mais surtout toutes les informations lors de retweets de ceux-ci ou de réponses (cf schéma précédent):

### 7.1 Collection focused on main political accounts

We established a list of more than 3 700 accounts of French political figures (candidates, deputies, senators, persons with a significant party mandate). We gather informations from these accounts with the follow API of Twitter (see the inset).

The documentation describes exactly the data collected for a followed user account:

From Twitter follow API documentation<sup>a</sup>. For each user specified, the stream will contain:

- · Tweets created by the user.
- · Tweets which are retweeted by the user.
- · Replies to any Tweet created by the user.
- · Retweets of any Tweet created by the user.
- Manual replies, created without pressing a reply button (e.g. "@twitterapi I agree").

The stream will not contain:

- Tweets mentioning the user (e.g. "Hello @twitterapi!").
- Manual Retweets created without pressing a Retweet button (e.g. "RT @twitterapi The API is great").
- Tweets by protected users.

https://dev.twitter.com/streaming/overview/request-parameters

Given an account X, a content A and T(X,A) a tweet published by X containing A, we collected, within the Twitter API volume limit, for each followed account X and any account Y the tweets T(X, A), T(Y, T(X, A)) (e.g. retweet), and if A is a response to X(@X ..."), all T(Y, A) (see the Fig. 10).

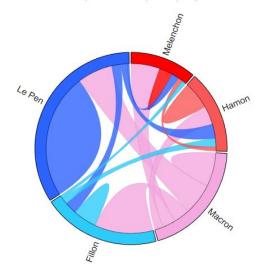
Et partant de là – vous le voyez venir -, quand l'échantillon de tweets est suffisamment important, ils peuvent (eux-aussi) inclure un compte dans une "communauté politique" :

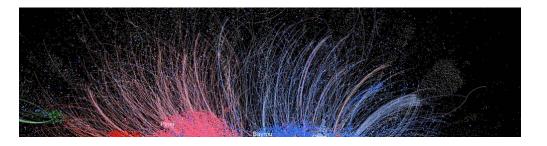
### Thèmes et communautés

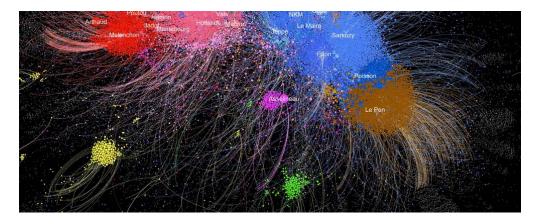
Les communautés politiques autour des candidats ont été définies algorithmiquement sur la base de l'analyse du partage de tweets à l'identique entre comptes Twitter (retweets "secs"). Les thèmes ont été identifiés grâce à des outils de fouille de texte développés à l'ISC-PIF. Voir la méthodologie pour plus de détails.

### Attention des communautés pour la journée du 18.06.17

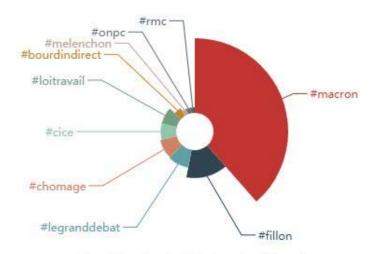
Détail de l'attention de chaque communauté pour les principaux candidats concurrents (survolez pour obtenir les infos)







Et ils font alors des statistiques :



Répartition des hashtag les plus fréquents pour ce thème dans la communauté de @JLMelenchon





Et on apprend alors (discrètement) combien de comptes ils ont ainsi classé politiquement durant la présidentielle : 187 619 !

| Name                  | # people present | Average time in days |
|-----------------------|------------------|----------------------|
| Marine Le Pen         | 35 812           | 33.8                 |
| François Fillon       | 30 516           | 30.7                 |
| Jean-Luc Mélenchon    | 42 206           | 21.8                 |
| Emmanuel Macron       | 57 651           | 16.2                 |
| Benoît Hamon          | 37 748           | 17.7                 |
| François Asselineau   | 14 774           | 11.8                 |
| Philippe Poutou       | 51 908           | 11.6                 |
| Nicolas Dupont-Aignan | 30 366           | 5.6                  |
| Nathalie Arthaud      | 45 450           | 4.5                  |
| Jacque Cheminade      | 29 153           | 4.4                  |
| Global                | 187 619          | 55.8                 |

Table 2: Table showing, for each candidate, the number of persons present at least once in his/her community, as well as the average time spent inside the community. The communities of Le Pen and Fillon are clearly different to the others, in the sense that they are the most stable with a very high average time. The fact that the average time spent by an account within a community (last line) is very high indicates that a significant proportion of accounts change from one community to another, without disappearing from Twitter.

Si on cumule chaque ligne, on arrive à un total 375 000 personnes, car leur méthode est assez frustre (ils n'analysent pas le contenu des tweets par exemple). Ainsi une personne peut-elle est comptée comme pro-Poutou une semaine, puis pro-Arthaud la suivante, puis pro-Mélenchon la suivante, selon ses retweets (car cette sensibilité politique a plusieurs candidats possibles proches). Mais on voit que les auteurs indiquent

que les communautés définies comme pro-Le Pen et pro-Fillon sont très stables... (sic.)

Et le plus fort est que leur opération ne s'est pas terminée avec la présidentielle – la base a doublé depuis :



Le Politoscope en est à plus de 126 millions de "tweets politiques", émanant de plus de 6 millions de comptes Twitter! (dont une seule fraction est analysée politiquement comme on l'a vu)

L'avantage de ce laboratoire du CNRS, cependant, est qu'ils ne diffusent pas la base publiquement – mais évidemment, la simple existence d'un tel fichier pose de très lourds problèmes – sachant que n'importe quel gros utilisateur de Visibrain (ou de l'API de Twitter directement) peut faire de même, et largement perfectionner les attributions politiques s'il le souhaite...

Après, certains diront que ce sont juste des chercheurs animés de bonnes intentions. Ce qui est vrai.

Mais on répondra : "mais quelle est la réelle valeur ajoutée de ces travaux ?". Bien sûr on apprend des choses, mais est-ce vraiment si important ? Cela vaut-il le risque que de tels fichiers soient constitués – et donc peut-être utilisés un jour (ou simplement piratés) ?

Êtes-vous à l'aise avec le fait qu'un tel fichier existe ?

Big Brother?

On imagine cependant que le laboratoire aura fait montre de prudence que DisinfoLab.

Mais cela pose néanmoins de nombreuses questions : les données sont-elles anonymisées, comment sont elles-stockées, est-il vraiment impossible de lever l'anonymat si la base était piratée ? (cela semble difficile, car il suffit de retrouver certains tweets dans Twitter, etc.)

Espérons en tout cas que ce laboratoire répondra rapidement à nos interrogations, qui restent à prendre au conditionnel...

### IV. Réglementation et Discussion

Rappelons que le Code pénal précise ceci :

Article 226-16

Le fait, y compris par négligence, de procéder ou de faire procéder à des traitements de données à caractère personnel sans qu'aient été respectées les formalités préalables à leur mise en oeuvre prévues par la loi est puni de cinq ans d'emprisonnement et de 300 000 euros d'amende.

Article 226-18

Le fait de collecter des données à caractère personnel par un moyen frauduleux, déloyal ou illicite est puni de cinq ans d'emprisonnement et de 300 000 euros d'amende.

Art. 226-19

Le fait, hors les cas prévus par la loi, de mettre ou de conserver en mémoire informatisée, sans le consentement exprès de l'intéressé, des données à caractère

12/08/2018 10:46 30 of 38

personnel qui, directement ou indirectement, font apparaître les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses, ou les appartenances syndicales des personnes, ou qui sont relatives à la santé ou à l'orientation ou à l'identité sexuelle de celles-ci, est puni de cinq ans d'emprisonnement et de 300 000 € d'amende.

### Art. 226-22

Le fait, par toute personne qui a recueilli, à l'occasion de leur enregistrement, de leur classement, de leur transmission ou d'une autre forme de traitement, des données à caractère personnel dont la divulgation aurait pour effet de porter atteinte à la considération de l'intéressé ou à l'intimité de sa vie privée, de porter, sans autorisation de l'intéressé, ces données à la connaissance d'un tiers qui n'a pas qualité pour les recevoir est puni de cinq ans d'emprisonnement et de 300 000 € d'amende.

La divulgation prévue à l'alinéa précédent est punie de trois ans d'emprisonnement et de 100 000 € d'amende lorsqu'elle a été commise par imprudence ou négligence.

L'article 8 de la loi n°78-17 du 6 janvier 1978 (CNIL) apporte toutefois un bémol :

I. – Il est interdit de traiter des données à caractère personnel qui révèlent la prétendue origine raciale ou l'origine ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale d'une personne physique ou de traiter des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique.

II. – Dans la mesure où la finalité du traitement l'exige pour certaines catégories de données, ne sont pas soumis à l'interdiction prévue au I:

1° Les traitements pour lesquels la personne concernée a donné son consentement exprès, sauf dans le cas où la loi prévoit que l'interdiction visée au l ne peut être levée par le consentement de la personne concernée ;

4° Les traitements portant sur des données à caractère personnel rendues publiques par la personne concernée;

Il y a bien sûr deux types de données. Celles publiées par l'utilisateur, et celles déduites par des algorithmes à partir des précédentes.

C'est ainsi que, par exemple, Samuel Laurent a pu dire ceci, interpellé à propos des fichiers DisinfoLab:



C'est à dire qu'on est sur un réseau public, et qu'on parle de chercheurs. Donc bon

12/08/2018 10:46 32 of 38

13:24 - 6 août 2018

Ce n'est évidemment pas faux – pour la partie publique (et peut-on véritablement parler de "chercheurs" pour des personnes non universitaires ?).

Mais, première objection : quand ce jeune étudiant indique son nom et sa couleur politique, pour agir sur Twitter avec ses... 23 abonnés ;



À t-il bien véritablement donné son accord pour que de nombreux analystes créent des fichiers avec son nom et ses opinions politiques?

Il l'a certes dit publiquement sur Twitter, mais si c'est public, cela reste discret. Et un tweet s'oublie vite dans le fil, et le compte peut s'effacer. Mais ce n'est pas possible si des chercheurs créent des fichiers en permanence, non purgés.

Avec ce raisonnement, on pourrait d'ailleurs ficher plein de personnes : on vous reconnait

12/08/2018 10:46 33 of 38

en train de participer à une manifestation France Insoumise, vous sortez du local Les Républicains, vous avez défendu Macron dans l'affaire Benballa lors d'une discussion au bistro (surtout, ne prenez pas la route!). C'était public – peut-on vous ficher avec vos opinions politiques ?

Et par ailleurs, on peut vraiment se demander si "la finalité du traitement exige" une telle utilisation des données.

Au-delà, il faut bien comprendre que **cette notion légale de "données rendues publiques", acceptable en 1976, ne l'est plus aujourd'hui.** En effet, par exemple avec Twitter, il y a différentes façons de connaitre vos opinions politiques :

- vous l'avez dit en dur dans votre biographie ou votre libellé de compte, comme l'étudiant ci-dessus (public);
- vous l'avez dit dans un tweet le soir du 1er tour de la présidentielle (public) ;
- vous relayez surtout un dirigeant politique (inférence très probable) ;
- vous parlez en général de vos opinions de nature politique, sans être partisan "il faut libérer les énergies, on paie trop d'impôts pour trop de fonctionnaires, etc." (inférence probable)
- vous parlez beaucoup de vous sur Twitter, vous suivez et retwittez beaucoup de personnes. Et c'est là que le Big Data peut inférer vos opinions politiques avec une précision que vous n'imaginez probablement pas. (inférence possible)

Ainsi, désormais, les pouvoirs publics ne doivent pas seulement s'occuper des données non publiques, mais également de la collecte et surtout la diffusion de masses de données publiques, en particulier sur les réseaux sociaux.

Twitter est ainsi gravement fautif – finalement bien plus que DisinfoLab ou le ISC-PIF du CNRS. Après tout, si vous donnez des millions de données à des chercheurs, certains finiront immanquablement par les utiliser ainsi...

Bref c'est bien les conditions d'utilisation de l'API de Twitter qu'il faut viser dans le

combat pour la protection des données de la population (source) :



L'affaire Benalla enflée par le "gonflage numérique", qui arrive aussi en Belgique

Contrairement à Facebook ou d'autres, le réseau Twitter offre assez de transparence pour observer les interactions entre utilisateurs. Il est par ailleurs une "bulle d'influence" en pleine consolidation, souligne le directeur d'EU Disinfo Lab, ce qui en fait un excellent baromètre.

Frédéric Rohart, Journaliste Source: L'Echo

Quand on pense que la CNIL a été créée en 1978 après une vive émotion dans l'opinion publique suite à un projet gouvernemental visant à identifier chaque citoyen par un numéro et interconnecter, via ce numéro, tous les fichiers de l'administration – on se rend compte de l'énorme baisse de notre vigilance, et des risques pour les Libertés publiques...

12/08/2018 10:46 35 of 38

### IV. Plainte

Nicolas Vanderbiest a déclaré :



et:



Pour vérifier cela, j'ai porté plainte à la CNIL. - comme beaucoup d'entre vous l'ont fait. Nous soutenons en effet que l'identifiant Twitter (pseudo ou pas) est bien évidemment une donnée personnelle.

J'ai également mandaté mon avocat afin de prévenir le Procureur de la République de ces faits particulièrement graves.

Nous allons évidemment saisir également la CNIL belge et la CNIL irlandaise (siège de Twitter), ainsi que les services Protection des données de la Commission européenne.

Si des gens veulent nous aider pour agir à l'international, ou pour prévenir d'autres utilisateurs concernés et afin de grouper les plaintes, ou si vous êtes spécialiste de ce sujet de protection des données et pouvez nous aider, vous pouvez nous écrire ici.

(Billet édité)

# #POLITOSCOPE Ce projet du @CNRS

Résumé. À l'occasion de l'élection présidentielle française de 2017, nous avons développé le Politoscope, une plateforme ouverte d'analyse des tweets politiques qui a collecté et analysé près de 60 millions d'échanges Twitter entre plus de 2,4 millions d'utilisateurs interagissant avec les personnalités politiques françaises.

| Name               | # people present |
|--------------------|------------------|
| Marine Le Pen      | 35 812           |
| François Fillon    | 30 516           |
| Jean-Luc Mélenchon | 42 206           |
| Emmanuel Macron    | 57 651           |
| Benoît Hamon       | 37 748           |



Dossier: #FichagePolitique - EU Disinfo Lab

12/08/2018 10:46 37 of 38

#2 #3 #4 #5

Un laboratoire du CNRS aurait aussi fiché les opinions politiques de près de 200 000 personnes! #FichagePolitique?

Tous fichés? Comment Twitter permet le fichage politique de la population

L'étude de l'influence russophile sur le Benallagate: d'énormes failles méthodologiques

Le EU DisinfoLab à l'origine d'une intox : l'affaire Benalla, produit de la « russosphère »?