

#AffaireBenalla : déconstruction d'une polémique sur le rôle de la communauté Twitter « russophile » dans le débat politique français

Colin Gérard¹ et Guilhem Marotte²

Le 18 juillet 2018, le journal *Le Monde* révélait au public le premier volet de ce qui est désormais connu comme l'« affaire Benalla³ », du nom d'un chargé de mission de l'Élysée accusé – entre autres – d'avoir usurpé la fonction de policier en interpellant et violentant deux personnes lors d'une manifestation à Paris le 1^{er} mai 2018. Cette affaire a pris un tournant inattendu, lorsque la presse française a relayé les conclusions partielles d'une étude de l'ONG EU DisinfoLab, publiées par un de ses auteurs sur Twitter⁴. Selon cette étude, l'affaire Benalla aurait engendré une hyperactivité d'une petite communauté virtuelle qualifiée de « russophile », à l'origine d'un volume de publications – dont de nombreuses fausses informations – inédit sur le segment francophone de Twitter⁵. Rapidement récupérée par la classe politique

Hérodote, n° 177-178, La Découverte, 2^e et 3^e trimestres 2020.

1. Doctorant en géographie – Spécialité géopolitique, IFG Lab (université Paris 8), Inria (équipe Datasphère). Chercheur associé, centre de recherche Géopolitique de la datasphère (GEODE).

2. Docteur en géographie – Spécialité géopolitique, cartographe, IFG Lab (université Paris 8). Chercheur associé, centre de recherche GEODE.

3. Ariane Chemin, « *Le Monde* identifie, sur une vidéo, un collaborateur de Macron frappant un manifestant, le 1^{er} mai, à Paris », *Le Monde*, 18 juillet 2018.

4. Les conclusions partielles de l'étude ont été publiées par Nicolas Vanderbiest, alors doctorant à l'université de Louvain et membre du DisinfoLab. Voir <www.twitter.com/Nico_VanderB/status/1023871181387116545>.

5. EU DisinfoLab, « Affaire Benalla : Les ressorts d'un hyperactivisme sur Twitter », 8 août 2018, <www.spark.adobe.com/page/Sa85zpU5Chila/>, consulté le 18 décembre 2019.

avant même sa publication finale, l'étude a été utilisée par la majorité présidentielle, avec pour effet de détourner temporairement l'attention médiatique vers la Russie et ses médias internationaux RT (Russia Today) et Sputnik, accusés d'avoir cherché à profiter de l'affaire pour déstabiliser l'exécutif français.

Cette récupération politique a engendré une amplification de la polémique, induisant de fait une capacité de la Russie à structurer le débat politique français *via* son action sur les réseaux sociaux. Les chercheurs du DisinfoLab ne tirent pourtant à aucun moment de conclusions politiques de leur travail, purement quantitatif. Ainsi, leur étude soulève de nombreuses questions d'ordre méthodologique et politique. Elle interroge la manière dont peuvent être politiquement exploitée une polémique née sur les réseaux sociaux (l'annonce des conclusions partielles par un des auteurs de l'étude) et les données quantitatives issues des mêmes réseaux (volume de tweets, de retweets, etc.). En effet, les données utilisées par l'étude ont été perçues par le grand public et le monde politique comme des informations objectives, vraies puisque chiffrées. L'étude interroge donc également la place qu'occupent les réseaux sociaux dans nos sociétés. Considérés comme des outils œuvrant en faveur du progrès démocratique à leur création, les réseaux sociaux sont désormais pointés du doigt pour leur rôle potentiellement néfaste et leur dangerosité pour les démocraties.

Cette affaire s'apparente à ce que le sociologue Dominique Cardon qualifie de « vieille tendance au déterminisme technique, consistant à faire porter à la technologie la responsabilité d'un changement de climat politique, [qui] risque de détourner l'attention de transformations sociopolitiques beaucoup plus fondamentales » [Cardon, 2019, p. 264]. Le gigantesque volume de données produites sur les réseaux sociaux semble en effet provoquer une impression d'exactitude et d'exhaustivité lorsqu'on ne le regarde que d'un point de vue statistique. Il est toutefois possible, à partir des données disponibles issues de ces réseaux, d'adopter une approche critique des données [Cattaruzza, 2019] et de relativiser l'impact de leur volume extrêmement élevé. Nous allons voir que, en fonction des métriques utilisées, l'influence d'un acteur sur les réseaux sociaux peut être analysée de plusieurs manières.

Cet article propose donc de revenir sur le rôle croissant des réseaux sociaux dans les rapports géopolitiques, à travers l'exemple des médias russes internationaux et la polémique liée à l'affaire Benalla. RT et Sputnik, créés en réaction à un traitement de l'information jugé défavorable à la Russie, sont des symboles d'une transposition des luttes d'influence et des rivalités de pouvoir dans le cyberspace [Douzet, 2014]. Ils sont à l'origine de nouvelles tensions entre États et de représentations probablement exagérées de l'influence russe, qui reste difficile à estimer du fait des enjeux techniques qu'elle soulève et de la difficulté de la définir de manière consensuelle. Il est donc indispensable de développer des méthodologies d'analyse permettant d'évaluer avec précision l'impact de cette

stratégie. À partir de données issues des réseaux sociaux, nous allons voir qu'il est possible d'analyser et de cartographier une partie de l'activité qui s'y produit de manière qualitative et problématisée, un champ d'études jusqu'ici peu traité par les géographes, tant du fait de sa nouveauté que des compétences en informatique et en mathématiques requises pour l'extraction et la modélisation des données. Les données que nous utilisons dans cet article n'ont pas vocation à représenter les phénomènes que nous étudions dans leur exhaustivité, du fait de la méthodologie d'extraction que nous avons utilisée. Elles nous permettent néanmoins de saisir des grandes tendances en matière de propagation de l'information sur les réseaux sociaux. En analysant ces données sur des outils de visualisation algorithmique conçus par et pour des techniciens, nous sommes également parvenus à élaborer des représentations graphiques jusque-là inédites.

Les réseaux sociaux comme nouvel espace de projection de la puissance russe

Depuis le début de la décennie 2010, la Fédération de Russie développe une stratégie d'influence informationnelle à destination de l'étranger, articulée autour d'une forte présence sur Internet et les réseaux sociaux. Axe majeur d'une politique de diplomatie publique lancée au milieu des années 2000 [Kiriya, 2016], cette stratégie a connu plusieurs évolutions depuis sa mise en place. Elle passe rapidement de la promotion du modèle russe à une logique de dénigrement des sociétés libérales occidentales et de soutien à des gouvernements proches de la Russie, dans le monde arabe ou en Amérique latine. En France, l'influence informationnelle russe se diffuse *via* trois types d'acteurs. Ce sont, d'abord, des instruments médiatiques comme RT et Sputnik, deux supports de presse internationaux multilingues intégralement financés par l'État russe, dont les versions francophones ont été créées en 2015. Favorables aux intérêts géopolitiques russes, les contenus de ces médias et les représentations qu'ils véhiculent sont ensuite rediffusés par des communautés virtuelles composées d'utilisateurs d'Internet et des réseaux sociaux. Enfin, la diffusion des contenus de RT et Sputnik est également assurée par une galaxie de sites Internet et de blogs, qui répliquent les publications des médias russes sur leur site, par proximité ou opportunisme idéologique, voire par pur intérêt financier [Limonier, 2018].

Passés relativement inaperçus à leurs lancements respectifs en France, RT et Sputnik sont aujourd'hui sous le feu des projecteurs médiatiques et politiques. En effet, depuis la tentative russe d'influencer l'élection présidentielle américaine de 2016, la représentation d'une Russie capable de manipuler les processus démocratiques d'États tiers par le biais de ses activités cybernétiques et informationnelles est désormais bien ancrée en France [Limonier et Gérard, 2017; Mattelart,

2018]. Cette représentation nourrit de nombreux débats et polémiques similaires à celles autour de l'affaire Benalla. Les lignes éditoriales de RT et Sputnik ont en commun de présenter une vision alternative et relativiste de l'information, consistant à remettre en question les fondements du modèle libéral-démocrate des sociétés occidentales. Leurs contenus révèlent une appétence certaine pour les questions liées aux dissensions internes de l'Union européenne ou de l'Otan, ainsi qu'à la moindre crise politique traversée par un État membre majeur de ces institutions. Ainsi, il est devenu courant lors de chaque manifestation parisienne de croiser des journalistes de RT et Sputnik, qui diffusent en direct sur les réseaux sociaux les échauffourées dans ou en marge des cortèges, qui leur permettent de présenter un pays en pleine crise. Toutefois, la ligne éditoriale des deux médias ne doit pas s'analyser par le seul prisme du « vrai/faux ». Régulièrement accusés de relayer propagande et fausses informations, les contenus de RT et Sputnik sont toutefois plus subtils, et s'apparentent plus à de l'information orientée qu'à de la désinformation, une notion difficilement employable dans le cas des médias russes tant elle est devenue subjective [Audinet, 2018]. Si des informations fausses ont pu être publiées par RT et Sputnik, elles prennent pour la plupart la forme d'un propos rapporté par un des deux médias, qui peuvent par la suite se dédouaner de toute volonté de manipulation de l'information⁶. Ces fausses informations restent toutefois très marginales lorsqu'on les met en relation avec le volume total de publications des deux médias. De plus, se sachant observés par leurs confrères français et par des organismes de régulation tels que le CSA, les journalistes de RT et Sputnik n'ont aucun intérêt à publier des informations délibérément fausses. Le 28 juin 2018, le CSA a d'ailleurs mis en demeure RT France pour « manquements à l'honnêteté, à la rigueur de l'information et à la diversité des points de vue [...] [en] contestant la réalité des attaques à l'arme chimique dans la région syrienne de la Ghouta orientale », observant que « la traduction orale des propos tenus par un témoin syrien ne correspondait en rien à ce qu'il exprimait à l'antenne ». Cette décision, confirmée par le Conseil d'État le 22 novembre 2019, est perçue par RT comme une atteinte à la liberté de la presse⁷. Un point de vue renforcé par les refus

6. Citons par exemple un article de RT France publié le 29 mars 2019, rapportant les propos du général de brigade russe Viktor Koupchinine, selon lequel la France et la Belgique préparaient une attaque chimique en Syrie, dans le but d'accuser la Russie par la suite : « Pour Moscou, les renseignements français prépareraient une provocation à l'arme chimique en Syrie », RT France, 29 mars 2019.

7. Voir la mise en demeure du CSA : Olivier Schrameck, « Manquements à l'honnêteté, à la rigueur de l'information et à la diversité des points de vue : la chaîne RT France mise en demeure », Conseil supérieur de l'audiovisuel, 28 juin 2018. Voir aussi la base de jurisprudence du Conseil d'État, décision n° 422790, 22 novembre 2019.

d'accréditation que subissent les journalistes de RT et Sputnik à l'Élysée et dans les ministères. Assumés par le gouvernement, ces refus ont pourtant pour effet de donner du crédit à l'aspect alternatif recherché par les deux médias, symbolisé par leurs slogans, « Question more » et « Tell the untold ».

La base de la stratégie d'ancrage de RT et Sputnik dans le paysage médiatique français repose sur les réseaux sociaux, qui leur ont permis de s'implanter de manière durable dans un contexte de remise en question de la confiance accordée aux médias en France. Cette stratégie s'est notamment avérée assez efficace depuis le début du mouvement des Gilets jaunes, que les deux médias ont couvert de manière plus positive que le reste des médias français. Pour s'en convaincre, il suffit de comparer l'accueil réservé par les Gilets jaunes aux journalistes de RT et Sputnik et à ceux des chaînes d'information en continu françaises. Tandis que les premiers sont parfois salués cordialement par les manifestants, les seconds ont été victimes de plusieurs agressions⁸. Outre le climat exceptionnel de défiance auquel font face les médias traditionnels, ce phénomène pourrait s'expliquer par le fait qu'en labourant depuis maintenant plusieurs années le terrain numérique sur lequel est né le mouvement des Gilets jaunes, les médias russes sont parvenus à encourager de manière opportuniste un certain nombre de thématiques chères aux manifestants. Ainsi, le thème des inégalités socio-économiques, quasiment inexistant dans la ligne éditoriale des médias russes il y a deux ans, est devenu un élément important de leur rhétorique en donnant beaucoup la parole à de nombreuses figures du mouvement⁹. L'angoisse du déclassement social, qui semble être l'un des moteurs de la contestation des Gilets jaunes, s'est également progressivement intégrée au dispositif narratif développé par les médias russes, venant enrichir une grille de lecture du monde où se mêlent condamnation de l'impérialisme moral américain, perte de souveraineté des États européens, affaissement des valeurs traditionnelles et angoisses migratoires. Cette stratégie opportuniste, qui consiste à intégrer à la grille de lecture proposée par les médias russes uniquement des thématiques qui suscitent l'émoi, semble donc trouver son paroxysme avec le mouvement des Gilets jaunes.

Par ailleurs, les formats des contenus publiés par RT et Sputnik sont adaptés au fonctionnement des réseaux sociaux : les vidéos en direct de RT ou les

8. Voir l'accueil de RT (RT France, « Des Gilets jaunes scandent "Merci RT !" à Paris », vidéo mise en ligne sur Youtube le 29 décembre 2018), et celui des journalistes des chaînes d'information en continu (BFMTV, « Plusieurs agressions de journalistes ont émaillé la mobilisation des gilets jaunes », vidéo mise en ligne sur Youtube le 24 novembre 2018).

9. Pour l'anniversaire des 100 jours du mouvement de protestation, l'émission « Interdit d'interdire » animée par Frédéric Taddei a par exemple organisé un débat entre François Boulo, Laëtitia Dewalle, Maxime Nicolle et Hayk Shahinyan, quatre figures des Gilets jaunes.

contenus *clickbait*¹⁰ de Sputnik assurent aux deux médias une base d'audience non négligeable. Au mois de décembre 2019, RT France est suivie par 1,1 million, 138 000 et 588 000 utilisateurs sur (respectivement) Facebook, Twitter et YouTube, les trois plateformes où les médias russes concentrent majoritairement leurs efforts¹¹. Sputnik France, dont les moyens financiers et humains sont bien inférieurs à ceux de RT¹², dispose d'une audience inférieure, avec (respectivement) 426 000, 74 000 et 80 000 abonnés sur les mêmes réseaux. Outre les utilisateurs abonnés, l'audience sur YouTube peut également être observée en termes de nombre total de vues des vidéos des deux médias : RT France en comptabilise plus de 238,5 millions, bien plus que les 12,7 millions de Sputnik. Il convient toutefois de relativiser ces chiffres et l'influence qu'ils sous-entendent en les comparant avec ceux d'autres médias français qui émettent à l'international. RFI recense par exemple plus de 4,1 millions d'abonnés sur Facebook, est suivie par 2,2 millions de personnes sur Twitter, et dispose de 179 000 abonnés sur YouTube, où ses vidéos ont été vues plus de 23 millions de fois.

Alors que l'audience des médias russes paraît relativement importante à la seule lecture de ces chiffres, leur impact concret sur l'opinion politique des citoyens français reste très difficile à estimer. En effet, la mesure de l'influence en ligne relève d'enjeux sociotechniques qu'on ne sait pas analyser avec précision. Le facteur comportemental, déterminant dans le choix d'un internaute de lire un article de Sputnik ou de visionner une vidéo de RT, est impossible à analyser en raison de son irrationalité. Ce facteur est imbriqué à des enjeux plus techniques, inhérents au fonctionnement même des réseaux sociaux. Ce sont par exemple les algorithmes de ces réseaux qui, en fonction du comportement en ligne d'un utilisateur (type de comptes suivis, interactions avec des publications, etc.), déterminent quels contenus afficher pour susciter le plus d'intérêt [Hong et Davison, 2010].

10. Le *clickbait*, « piège à clic » en français, est une technique de rédaction de contenus et de titres utilisée à des fins de marketing digital. Originellement conçue pour générer des revenus en ligne, cette technique consiste à attirer le plus de lecteurs possible par des titres racoleurs et en jouant sur les émotions de sorte à grossir l'audience du site voulu.

11. Statistiques d'audience consultées sur les différents comptes des médias russes le 19 décembre 2020.

12. RT France est dotée en 2018 d'un budget de 20 millions d'euros. Celui de Sputnik n'est pas connu mais reste bien inférieur, en attestent les budgets prévisionnels de 2020 pour les maisons mères des deux médias. Le groupe RT se voit allouer 22,9 milliards de roubles (environ 332 millions d'euros), dont 3 réservés au lancement de la chaîne RT Deutsch, contre 6,8 milliards de roubles (environ 98 millions d'euros) pour Rossiya Segodnya. Sergeï Sobolev, *Minfin predlozhit' uvelichit' subsidii «Pervomu kanalu» i VGTRK* («Le ministère des Finances a proposé d'augmenter les subventions de Pervyi Kanal et VGTRK»), RosBiznesConsulting (RBC), 26 septembre 2019.

Alors que le monde académique travaille depuis une dizaine d'années sur ces questions de mesure de l'influence d'un point de vue purement technique et statistique, ces travaux sont limités par les restrictions d'accès aux données imposées par certaines plateformes, et *in fine* par l'impossibilité de mettre en relation les données disponibles avec des facteurs humains [Bakshy *et al.*, 2011]. De ce fait, l'impact d'une stratégie d'influence ne se mesure pas tant selon des statistiques d'audience précises que par l'interprétation tirée de ces statistiques. En effet, nous allons voir que dans le cas de la polémique autour de l'affaire Benalla, c'est bien la question du volume particulièrement élevé de tweets relevé par le DisinfoLab qui a conduit certains médias et personnalités politiques à conclure à une influence en ligne réelle de RT et Sputnik.

L'impact de l'hyperactivité des médias russes sur Twitter pendant l'affaire Benalla

L'influence des réseaux sociaux et de l'espace numérique en question

Le 30 juillet 2018, Nicolas Vanderbiest, alors doctorant à l'université de Louvain et membre du EU DisinfoLab, publie sur Twitter les conclusions provisoires d'un rapport de l'ONG sur le traitement de l'affaire Benalla sur Twitter. Selon ses premiers résultats, l'affaire aurait engendré un volume de tweets exceptionnellement élevé, avec au total plus de quatre millions de tweets publiés depuis le 19 juillet 2018. Parmi ces tweets, récoltés sur la plateforme Visibrain, 44 % ont été publiés par seulement 1 % de comptes, dont une part importante (27 %) qualifiée de « russophile ». En d'autres termes, une poignée de comptes, dont environ 900 « russophiles », a produit presque la moitié du « bruit » autour de l'affaire Benalla sur Twitter. Vanderbiest parle alors de « gonflement numérique puissance 20 ». Remarqués par les journalistes, ses tweets vont rapidement donner lieu à la publication d'articles de médias français rapportant l'activité de ces comptes russophiles. Sur la base de ces articles, l'ancien député Les Républicains Frédéric Lefebvre, actuellement membre du parti En Marche!, publie le 1^{er} août un communiqué au nom de son parti Agir, appelant à l'ouverture d'une commission d'enquête sur « la manipulation attribuée aux comptes russophiles sur Twitter pour déstabiliser l'exécutif français¹³ ». Ce communiqué sera exploité par certains

13. Frédéric Lefebvre, « Affaire Benalla : la commission d'enquête doit se saisir de la manipulation attribuée aux comptes russophiles sur Twitter pour déstabiliser l'exécutif français », Agir, en ligne.

responsables de la majorité présidentielle, parmi lesquels l'ancien secrétaire d'État au numérique Mounir Mahjoubi ou l'ancien porte-parole du gouvernement Benjamin Griveaux, permettant ainsi de détourner momentanément l'attention médiatique de l'affaire en elle-même. Si la polémique autour d'un possible rôle russe s'est rapidement éteinte, elle soulève en revanche des points intéressants.

Dans l'étude, les auteurs qualifient de « russophiles » les comptes « partage[ant] publiquement et régulièrement des articles de RT et Sputnik » et qui ont « une sensibilité aux narratifs pro-russes ». Ici, la russophilie n'est « pas un jugement de valeur, mais un fait quantifiable selon une méthodologie ». C'est pourtant ce terme, accolé à RT et Sputnik, qui est mis en avant par les journalistes et la classe politique, créant ainsi une polémique avec pour toile de fond le rôle joué par la Russie sur les réseaux sociaux¹⁴. S'il n'est pas question ici de remettre en question le contenu du rapport, dont certains points méthodologiques sont très intéressants, les réactions qu'il a provoquées interrogent en revanche sur la capacité de la Russie à être présentée dans le débat public comme une menace pour la démocratie française. La représentation d'une puissance informationnelle russe capable d'influencer le débat politique et les élections en France *via* des réseaux sociaux facilement manipulables semble en réalité assez exagérée, et donne possiblement trop de pouvoir tant à la Russie qu'à l'Internet et aux réseaux sociaux. Dans leur ouvrage *Network Propaganda*, Yochai Benkler, Robert Faris et Hal Roberts estiment que « l'erreur fondamentale du narratif "Internet polarise" est qu'il adopte une vision trop naïve du fonctionnement de la technologie et sous-estime le degré auquel les institutions, la culture et la politique façonnent les modèles d'adoption et de diffusion de la technologie » [Benkler *et al.*, 2018, p. 8]. Dans le cas de l'affaire Benalla, si les millions de tweets rapportés par EU DisinfoLab semblent à première vue représenter un nombre très élevé, il convient de relativiser leur portée et leur impact sur la population.

Dominique Cardon rappelle à cet effet que :

les grands nombres du numérique sont désormais partout et doivent cesser de nous impressionner. Sur Facebook, les vingt infox les plus partagées lors de la campagne électorale qui a élu Donald Trump l'ont été 8 711 000 fois [...]. Le chiffre impressionne, mais il correspond à 0,006 % des informations partagées sur Facebook aux États-Unis pendant la même période ! On s'éviterait beaucoup de faux débats si on apprenait à ne pas regarder les chiffres du numérique, un monde dans lequel

14. Dans un tweet du compte anglophone de EU DisinfoLab, il est question de « comptes russes » (*Russian accounts*) <<https://twitter.com/DisinfoEU/status/1024936941559717889>>.

le volume des informations disponibles et consommées a radicalement changé d'échelle, avec les lunettes du kiosquier qui fait le bilan de ses ventes¹⁵.

Un impact limité de la communauté agrégée autour des médias russes

Pour relativiser ces chiffres, nous proposons ici une approche critique basée sur l'analyse de graphe, qui pose la question de l'impact du bruit généré par RT, Sputnik et la communauté d'utilisateurs agrégés autour de leurs comptes sur Twitter. Car s'il est clair, comme le montre l'étude du DisinfoLab, que nous avons assisté à une activité frénétique de ces utilisateurs dans le contexte de l'affaire Benalla, celle-ci semble avoir eu un effet très limité sur les utilisateurs de Twitter en général. À partir de données récoltées grâce à une des API de Twitter¹⁶, nous avons pu constituer une base de données de tweets publiés entre le 18 juillet et le 8 août 2018. Cette base est composée de 37 363 tweets, publiés par 14 808 comptes sur un réseau centré autour des comptes de RT et Sputnik¹⁷. Les données récoltées ont également été filtrées afin de ne conserver que les tweets qui contiennent le terme « Benalla ». Après les avoir analysées, nous avons anonymisé l'ensemble des données à l'exception de celles de RT et Sputnik. Il convient de préciser que ce jeu de données n'a pas vocation à représenter l'ensemble de l'activité générée en ligne au cours de l'affaire, mais bien d'en donner un aperçu à un instant *t* et de s'en servir de base pour comprendre et représenter la manière dont l'information publiée par les deux médias circule sur Twitter. En outre, en ne nous concentrant que sur Twitter, nous nous focalisons sur un réseau qui ne représente qu'une petite part de l'activité des Français sur les réseaux sociaux, mais qui reste central dans le débat politique français, la majeure partie de la classe politique s'y affrontant quotidiennement.

15. Dominique Cardon, « Pourquoi avons-nous si peur des fake news? (1/2) », *AOC*, 20 juin 2019, p. 2.

16. Une API (interface de programmation applicative) est une interface qui vise à simplifier l'extraction et l'exposition des données d'un site Internet. Twitter est le seul réseau social majeur qui met à libre disposition une API Streaming. Cette dernière permet de capturer en temps réel des tweets publiés par ses utilisateurs, donc de récolter un échantillon de données suffisamment représentatif de l'objet d'étude voulu, de façon à produire des analyses les plus pertinentes possible.

17. La méthodologie de collecte des tweets est définie selon une double approche par les acteurs (ici les médias russes) et par les contenus (des mots-clés, ici « benalla »). Ici, apparaissent donc l'ensemble des tweets publiés par les médias russes sur l'affaire Benalla. Ainsi, si nous sommes en mesure d'analyser une partie importante de l'activité générée par RT et Sputnik sur Twitter en fonction de thèmes précis, nos données n'ont toutefois pas vocation à représenter l'activité du réseau dans son exhaustivité, contrairement à celles de la plateforme Visibrain par exemple.

À partir des tweets récoltés nous avons donc construit un graphe, composé d'arêtes et de nœuds. Les nœuds représentent des comptes uniques, tandis qu'une arête reliant deux nœuds A et B représente un retweet entre ces deux comptes¹⁸ (voir figure 1). Pour définir des communautés sociopolitiques sur Twitter, nous utilisons un algorithme de partitionnement de graphes sur le logiciel Gephi¹⁹, qui fait ressortir 227 communautés, un nombre trop élevé pour en analyser l'ensemble. Ce nombre s'explique par la quantité importante de nœuds et de liens dans le graphe. En éliminant les communautés représentant moins de 1 % du graphe dans notre analyse, nous pouvons toutefois réduire considérablement ce nombre. En dessous de ce pourcentage, les communautés n'ont pas d'influence réelle sur la structure du graphe et se situent pour la plupart à sa périphérie, autrement dit sans connexion notable avec le reste des nœuds. Ce faisant, nous arrivons à réduire le partitionnement à quatre communautés, qui comprennent 35 158 tweets postés par 14 551 comptes : parmi eux, 18,7 % sont des tweets organiques, c'est-à-dire des tweets rédigés, alors que les 81,3 % restants sont des retweets, c'est-à-dire des répliques de tweets organiques.

La position des nœuds dans le graphe est ici définie par Force Atlas 2, un algorithme de spatialisation du logiciel Gephi qui permet d'obtenir la représentation graphique d'un réseau [Jacomy *et al.*, 2014]. L'idée est de simuler un système mécanique pour spatialiser le graphe. Les nœuds se repoussent comme des aimants et les liens attirent les nœuds comme des ressorts. Ces forces créent des mouvements jusqu'à ce que le système trouve une position d'équilibre. Cette représentation est intéressante pour comprendre les interactions entre les différents acteurs d'un réseau.

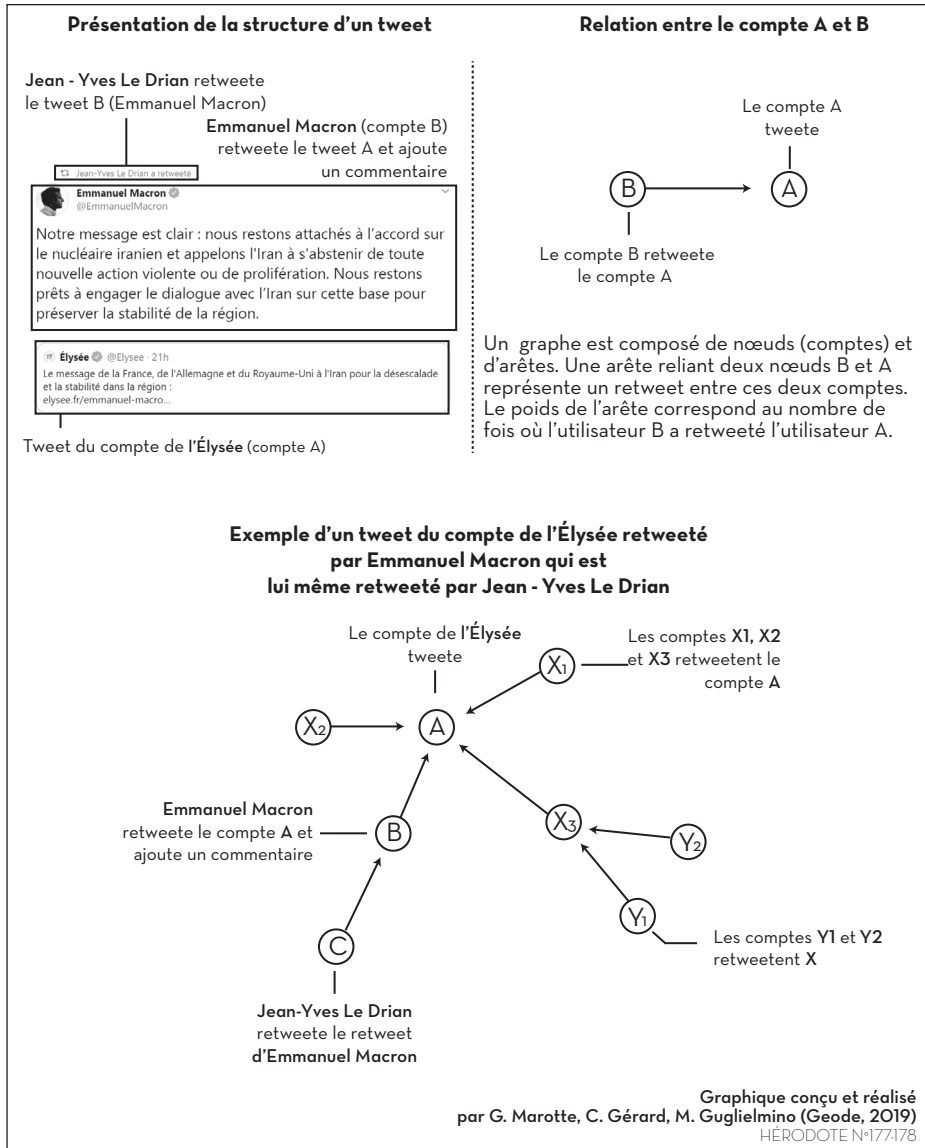
Ici, la notion d'importance d'un compte au sein d'une communauté est déterminée en fonction d'un score, obtenu par un calcul appliqué à chaque compte apparaissant dans la base de données²⁰. Ainsi, plus un compte est important dans

18. Ici, le poids n de l'arête indique que l'utilisateur A a retweeté n fois l'utilisateur B.

19. Gephi est un logiciel libre d'analyse et de visualisation de réseaux. Les algorithmes de clustering permettent de partitionner les données en communautés de manière non supervisée, c'est-à-dire indépendamment d'une quelconque source d'information autre que l'objet initial. Ici, nos communautés sont déterminées par une mesure de modularité, basée sur l'algorithme de Louvain. Ces communautés regroupent entre elles des observations similaires. Les algorithmes de clustering dépendent donc fortement de la façon dont est définie cette notion de similarité, qui est souvent spécifique à son domaine d'application. L'idée est ici de toujours maximiser les relations (poids des arêtes) entre les nœuds d'une même communauté et minimiser les liens avec toutes les autres communautés. C'est un problème combinatoire complexe pour lequel nous utilisons des approximations mathématiques proposées par Gephi.

20. Le calcul consiste à multiplier le degré de chaque nœud (c'est-à-dire la somme du nombre de fois où un compte A a retweeté d'autres comptes et du nombre de fois où il a été retweeté) par le poids de chaque lien dirigé vers ce nœud (le nombre d'interactions entre un compte et tous les comptes avec lesquels il interagit).

FIGURE 1. – REPRÉSENTATION SUR UN GRAPHE DIRIGÉ DES INTERACTIONS SUR TWITTER



Hérodote, n° 177-178, La Découverte, 2^e et 3^e trimestres 2020.

une communauté, plus son score sera élevé. Une fois obtenus, les scores de chaque compte sont additionnés de sorte à obtenir un score total par communauté. Afin d'analyser des échantillons représentatifs de chaque communauté, nous avons décidé de vérifier au cas par cas les comptes les plus importants d'une communauté dont la somme des scores correspond à au moins 30 % du score total de cette communauté. Si les communautés sont ici déterminées par des critères mathématiques, il est toutefois indispensable d'analyser de manière qualitative les comptes qui les composent. L'idée est de déterminer si les comptes regroupés au sein d'une même communauté ont une appartenance ou sensibilité politique commune. Nous nous intéressons donc aux caractéristiques principales de chaque compte en fonction des informations et des publications directement accessibles sur son fil Twitter. Certains comptes sont très facilement identifiables lorsqu'ils appartiennent à une personnalité publique (comme un homme ou une femme politique) et sont certifiés par Twitter. Pour les autres, c'est l'analyse des publications, des retweets, des personnes ou les hashtags utilisés qui permet d'attribuer une orientation politique au compte. Parfois, les bannières, les photos et les symboles utilisés par un compte fournissent des informations sur l'appartenance politique²¹. Une fois l'analyse qualitative des comptes effectuée, les données sont anonymisées, à l'exception de celles de RT, Sputnik et des personnalités politiques et médias de premier plan.

Sur la figure 2 apparaissent différentes communautés déterminées par la mesure de modularité de Gephi. Ainsi, on retrouve à droite du graphe une communauté qui concentre plutôt des utilisateurs de gauche et d'extrême gauche. À gauche du graphe, une communauté d'utilisateurs de droite et d'extrême droite. Au centre, une communauté de comptes agrégés autour de ceux de Sputnik et RT et, en haut, une communauté située à la marge du graphe, majoritairement composée de comptes de soutiens ou de membres de La République en Marche (LREM), bien que certains utilisateurs se déclarent « anti-Macron ».

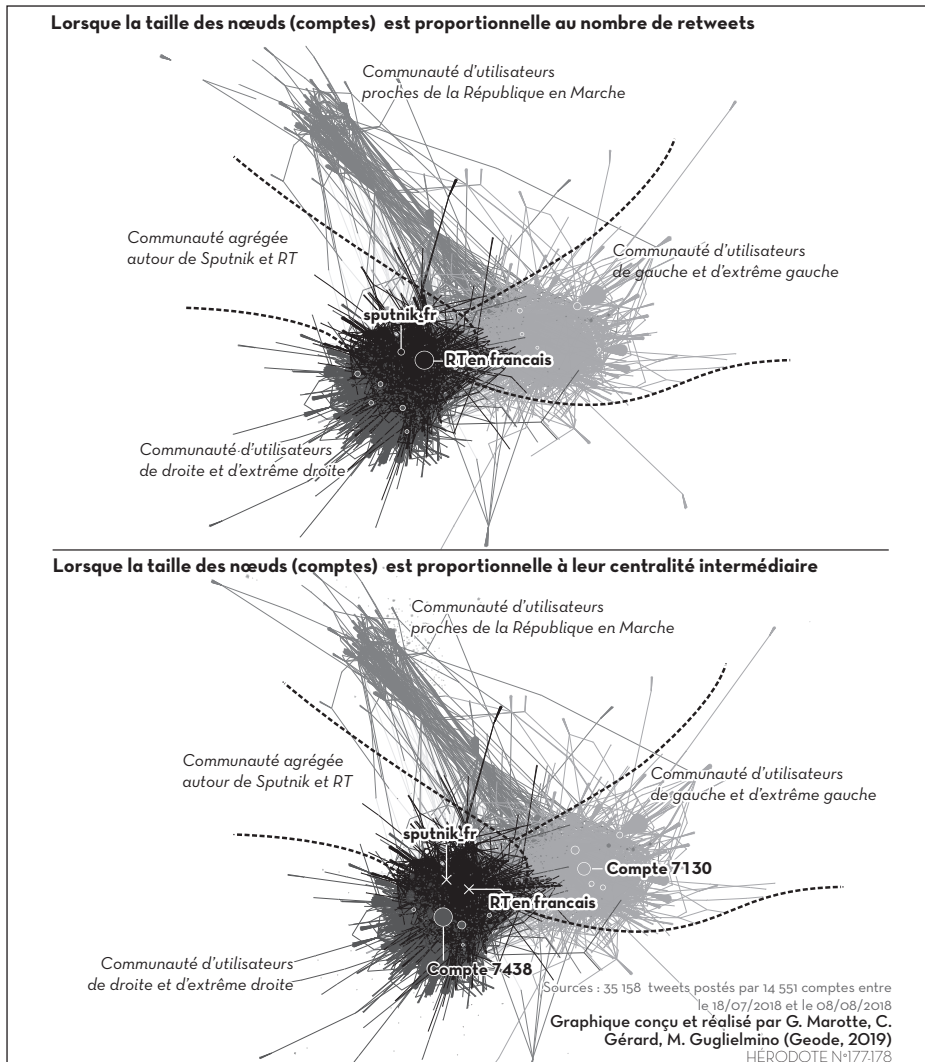
Il existe plusieurs métriques permettant d'évaluer l'influence d'un utilisateur à une période donnée (donc l'influence d'un nœud dans le graphe) :

1. le degré, qui correspond au nombre d'arêtes entrantes et sortantes d'un nœud, c'est-à-dire le nombre de fois où un utilisateur a retweeté (degré sortant) et a été retweeté (degré entrant). Dans cet article, nous qualifions les comptes qui ont été le plus retweetés de centrales médiatiques ;

21. Par exemple, le signe Ⲁ est souvent utilisé sur les réseaux sociaux pour afficher son soutien aux chrétiens d'Orient et a été repris par une partie de la droite identitaire. Si tous les comptes arborant ce signe ne font pas partie de l'extrême droite, la présence du Ⲁ est un indice pour le chercheur souhaitant qualifier un compte. À gauche de l'échiquier politique, c'est la présence de la lettre grecque « phi » (φ) apposée à côté du nom d'un utilisateur qui indique l'appartenance ou le soutien de ce dernier au mouvement de la France Insoumise.

FIGURE 2. – INTERACTIONS SUR TWITTER AUTOUR DE L'AFFAIRE BENALLA.

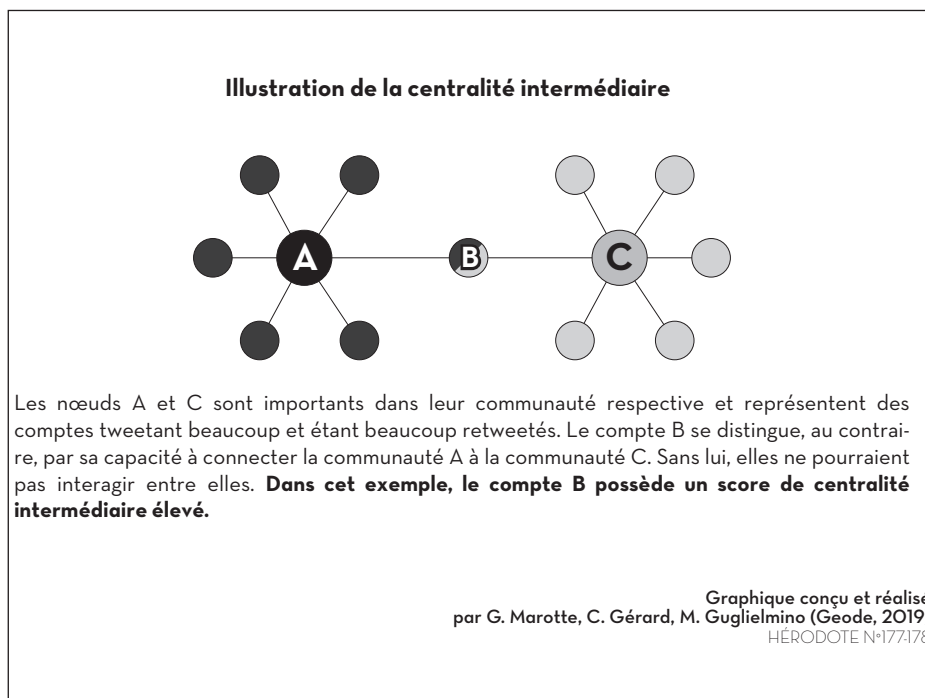
En haut, la taille des nœuds est proportionnelle à leur degré entrant.
En bas, elle est proportionnelle à leur score de centralité intermédiaire.



Hérodote, n° 177-178, La Découverte, 2^e et 3^e trimestres 2020.

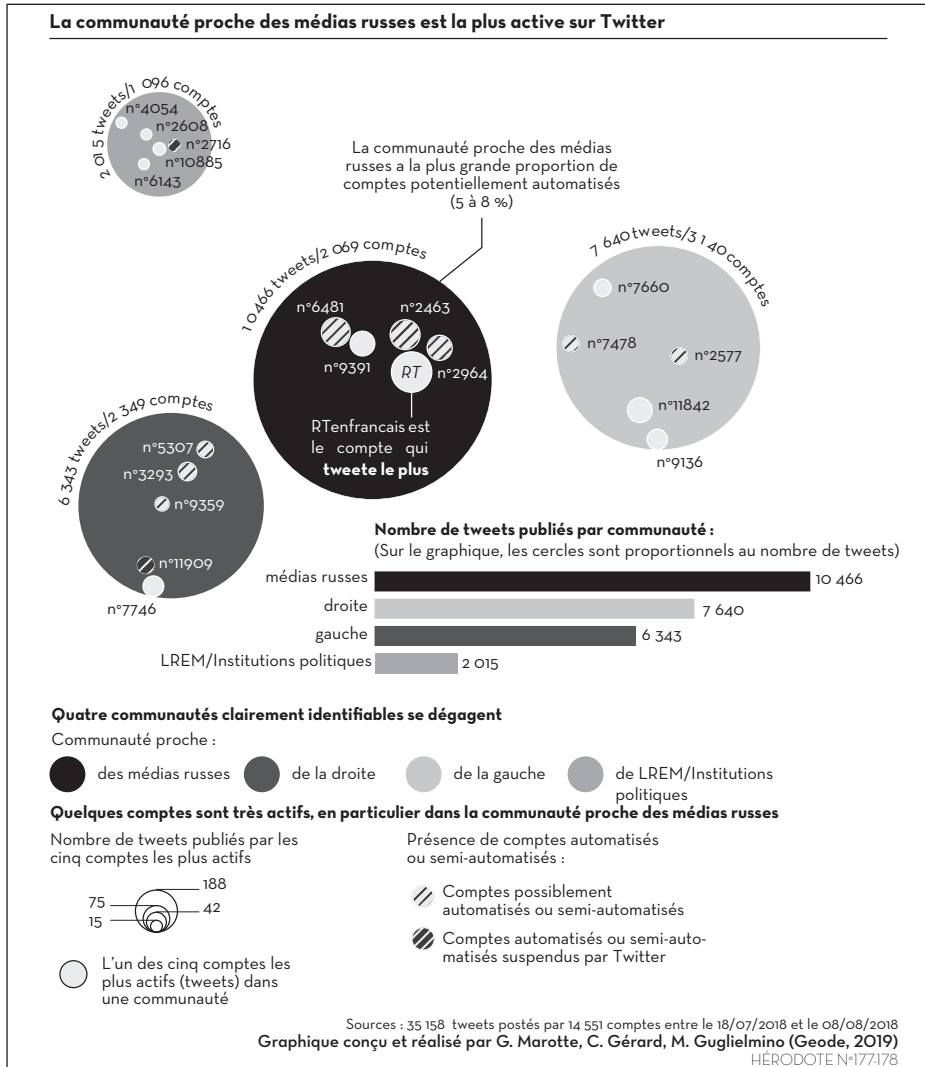
- la centralité intermédiaire (*betweenness centrality*), notion importante en théorie des graphes, qui définit mathématiquement pour chaque nœud le nombre de chemins les plus courts entre deux points passant par ce nœud. En d'autres termes, plus le score de centralité d'un utilisateur est élevé, plus l'utilisateur disposera d'une capacité à faire circuler une information au sein du réseau. Nous qualifions les comptes au score de centralité supérieur à 1 000 de hubs d'informations.

FIGURE 3. – SCHÉMA EXPLICATIF DE LA CENTRALITÉ INTERMÉDIAIRE



Ici, nous distinguons donc bien deux types d'influence : la production du discours (degré entrant) et sa propagation sur le réseau social (centralité intermédiaire).

FIGURE 4. – REPRÉSENTATION SYNTHÉTIQUE DE L'ACTIVITÉ DES COMMUNAUTÉS SOCIOPOLITIQUES SUR TWITTER PENDANT L'AFFAIRE BENALLA



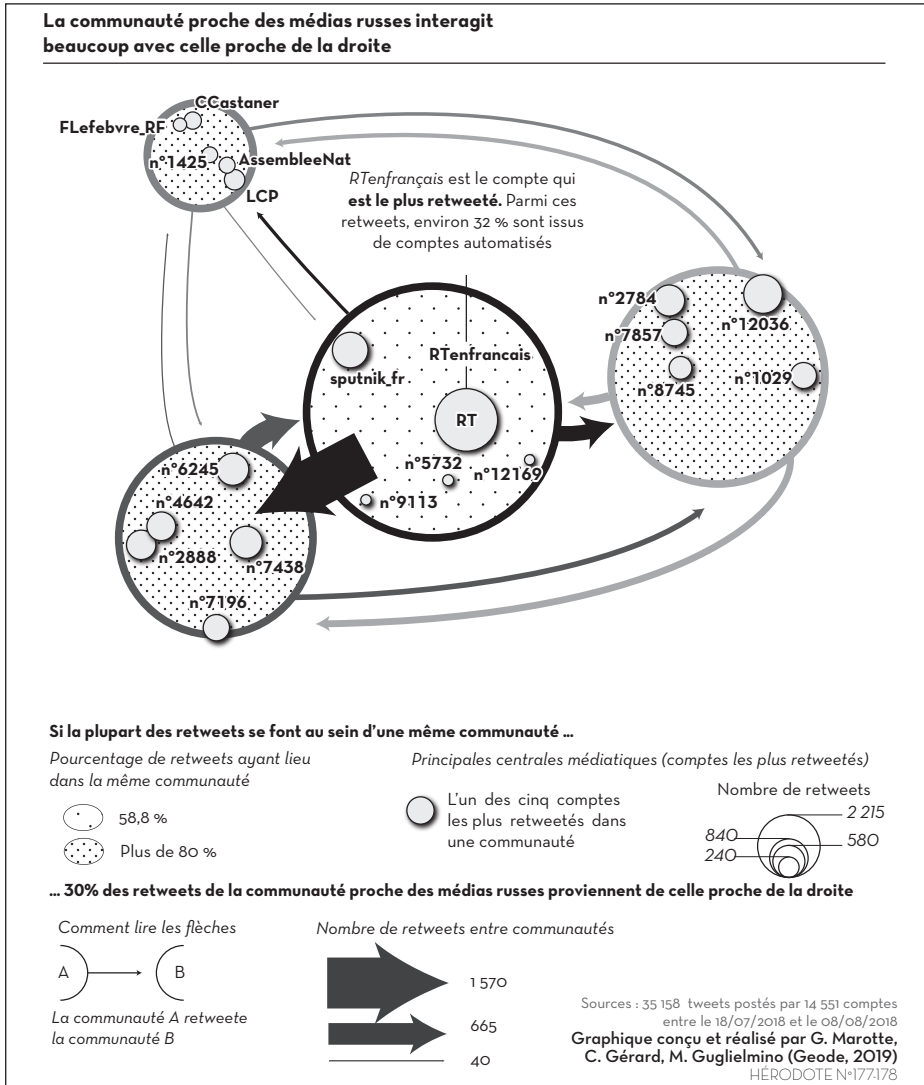
Hérodote, n° 177-178, La Découverte, 2^e et 3^e trimestres 2020.

Ici, sont représentés les comptes ayant publié le plus de tweets.

À partir des données issues du graphe, il est possible de représenter graphiquement les informations les plus importantes. La figure 4 présente par exemple le volume de tweets publiés par les communautés sociopolitiques détectées, ainsi que les comptes les plus actifs. On remarque d'abord que si la communauté agrégée autour des médias russes n'est pas la plus grande en termes de volume de comptes (2069 comptes contre 3 140 pour la communauté proche de la gauche), elle a toutefois été la plus active en termes de tweets publiés. Nous rejoignons donc ici une des conclusions de l'étude d'EU DisinfoLab, selon laquelle une petite poignée de comptes a généré un nombre important de tweets sur l'affaire Benalla. Afin de compléter l'analyse, nous nous sommes intéressés aux comptes potentiellement automatisés, susceptibles de gonfler les statistiques d'activité. Pour détecter ces bots, nous utilisons l'outil Botometer, développé par des chercheurs de l'université d'Indiana [Davis *et al.*, 2016]. L'outil utilise plus de mille caractéristiques pour prédire la probabilité qu'un compte soit automatisé. Ces catégories sont basées sur les métadonnées des utilisateurs (le nombre de tweets, de followers ou encore la longueur du pseudonyme), sur de l'analyse de sentiments (contenu du tweet), l'intégration de l'utilisateur dans le graphe (degré entrant, coefficient de clustering), le contenu des tweets ou encore leur fréquence. Botometer utilise un algorithme d'intelligence artificielle (*machine-learning*) pour évaluer la probabilité que l'utilisateur soit un bot, en faisant de l'apprentissage sur des bots déjà connus. Cette méthode a des limites qui dépassent le cadre de cette étude, mais permet de dégager de grandes tendances [Rauchfleisch et Kaiser, 2020]. Ainsi, nous pouvons constater que parmi les quatre communautés analysées, celle agrégée autour des médias russes dispose de la plus grande proportion de comptes potentiellement automatisés (entre 5 % et 8 %).

En nous basant sur la métrique du degré entrant, nous pouvons ensuite étudier les retweets entre les différentes communautés, et donc déterminer un certain niveau de proximité entre elles. Ici, la première information importante est que toutes les communautés, à l'exception de celle agrégée autour des médias russes, ont une part d'interactions intracommunautaires bien supérieure à la part d'interactions intercommunautaires. En d'autres termes, les comptes des communautés de gauche, de droite et proche de LREM ont tendance à retweeter majoritairement des comptes de leur propre communauté, alors que les comptes de la communauté agrégée autour des médias russes interagissent à un niveau plus élevé avec ceux de la communauté de droite (30 % des retweets), soulignant ainsi une certaine proximité entre les utilisateurs des deux communautés. L'extrême gauche, pourtant régulièrement taxée de faire le jeu de la stratégie d'influence russe, semble ici cantonnée à la promotion de ses seuls narratifs : les comptes de cette communauté se sont retweetés très majoritairement entre eux, à hauteur de 89 %. Cela n'exclut toutefois pas la possibilité que les narratifs employés à gauche soient communs à ceux diffusés par les acteurs affiliés à la Russie. On remarque également que les comptes de RT et Sputnik font

FIGURE 5. – REPRÉSENTATION SYNTHÉTIQUE DE L'ACTIVITÉ DES COMMUNAUTÉS SOCIOPOLITIQUES SUR TWITTER PENDANT L'AFFAIRE BENALLA



Hérodote, n° 177-178, La Découverte, 2^e et 3^e trimestres 2020.

Ici, sont représentées les interactions entre les communautés.

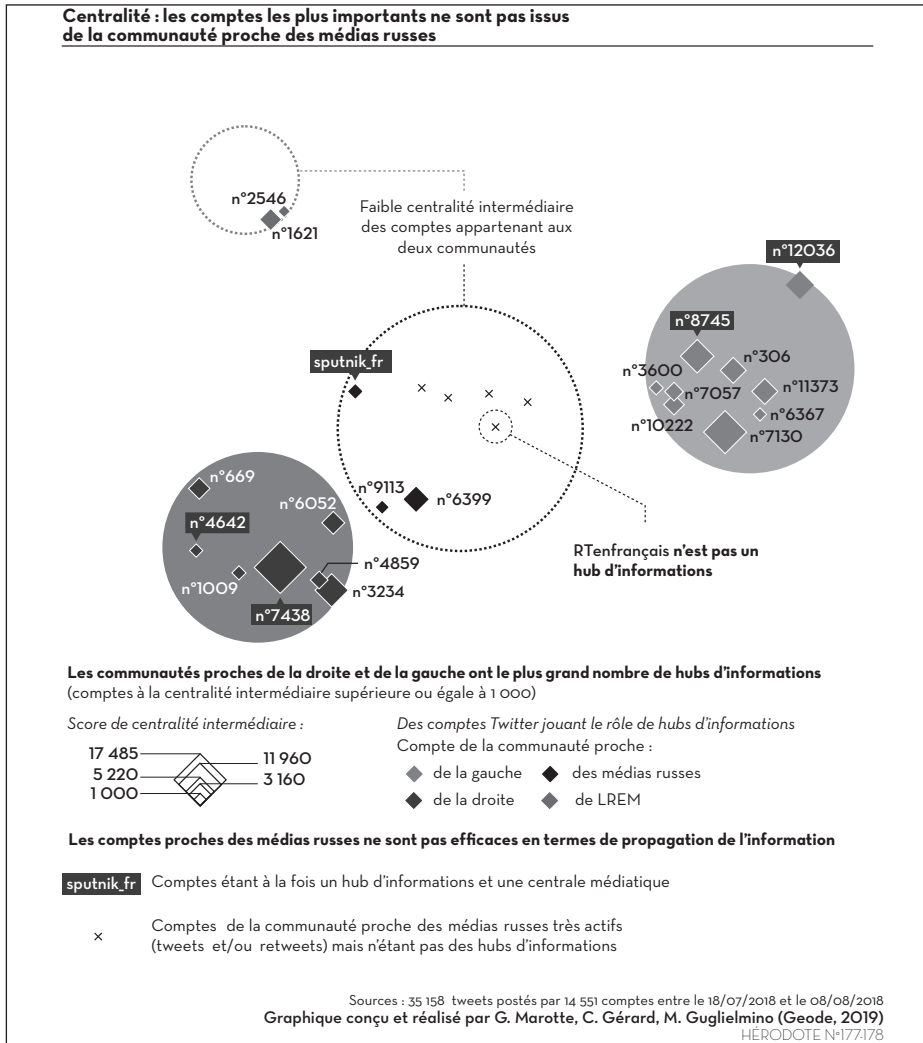
partie de ceux qui ont été les plus retweetés du graphe, une statistique qui s'explique, outre la popularité des contenus des deux médias pendant l'affaire Benalla, par le fait que notre méthodologie d'extraction est centrée sur ces deux comptes, dont elle récolte la totalité des tweets. Jusqu'ici, la thèse d'une hyperactivité des comptes proches des médias russes semble se confirmer.

Toutefois, il est possible de relativiser l'impact concret de cette hyperactivité en analysant les comptes du graphe en fonction de leur score de centralité intermédiaire, qui nous permet de distinguer deux types d'influence en ligne. D'une part, la production de contenus qui s'analyse avec le nombre de tweets et de retweets, et d'autre part la capacité à diffuser ces contenus, déterminée par la centralité. Cette métrique met en lumière les différents échanges entre les utilisateurs lorsque ces derniers se retweetent. Dans un réseau de télécommunication par exemple, les nœuds dotés d'une centralité importante sont les éléments nécessaires à une bonne transmission ; leur absence handicaperait le flot de propagation des données. Le score de centralité intermédiaire mesure donc l'importance d'un compte dans la transmission de l'information au sein d'un réseau. Ainsi, la suppression d'un compte possédant un score élevé perturberait la transmission de l'information au sein du graphe. Sur Twitter, les principaux producteurs de contenus sont généralement les comptes de médias ou les comptes officiels de partis politiques, qui émettent de l'information pour être le plus repris, sans toutefois avoir vocation à faire transiter l'information au sein du réseau global.

C'est notamment le cas de RT dont le score de centralité est nul, ce qui sous-entend que ses contenus ne se sont pas particulièrement propagés en dehors de sa communauté. En d'autres termes, RT n'a pas interagi avec d'autres utilisateurs, ce qui semble s'apparenter au comportement en ligne d'un média classique. Sputnik dispose de son côté d'un score de centralité de 1 711. Sa capacité à diffuser ses contenus a donc été sur la période étudiée supérieure à celle de RT, mais reste toutefois dix fois inférieure à celle du compte n° 7438, le plus central du graphe, issu de la communauté proche de la droite. Nous estimons donc que l'activité des deux médias russes et l'hyperactivité des comptes agrégés autour d'eux n'ont pas eu d'influence en dehors de leur propre communauté. En d'autres termes, le flot d'information est à sens unique pour les comptes agrégés autour de RT et Sputnik : comme beaucoup de médias, ils disposent d'une large audience, mais n'interagissent pas avec. En effet, seuls trois comptes de cette communauté ont un score de centralité supérieur à 1 000 et jouent par conséquent le rôle de hubs d'informations. La forte activité évoquée plus haut et décrite dans l'étude du Disinfo Lab ne semble donc pas avoir eu autant d'impact que ce que certaines personnalités politiques et médiatiques ont pu un temps affirmer.

Pour appuyer ce constat, il aurait été intéressant de comparer le volume de tweets et de retweets des deux médias russes avec celui d'autres grands médias

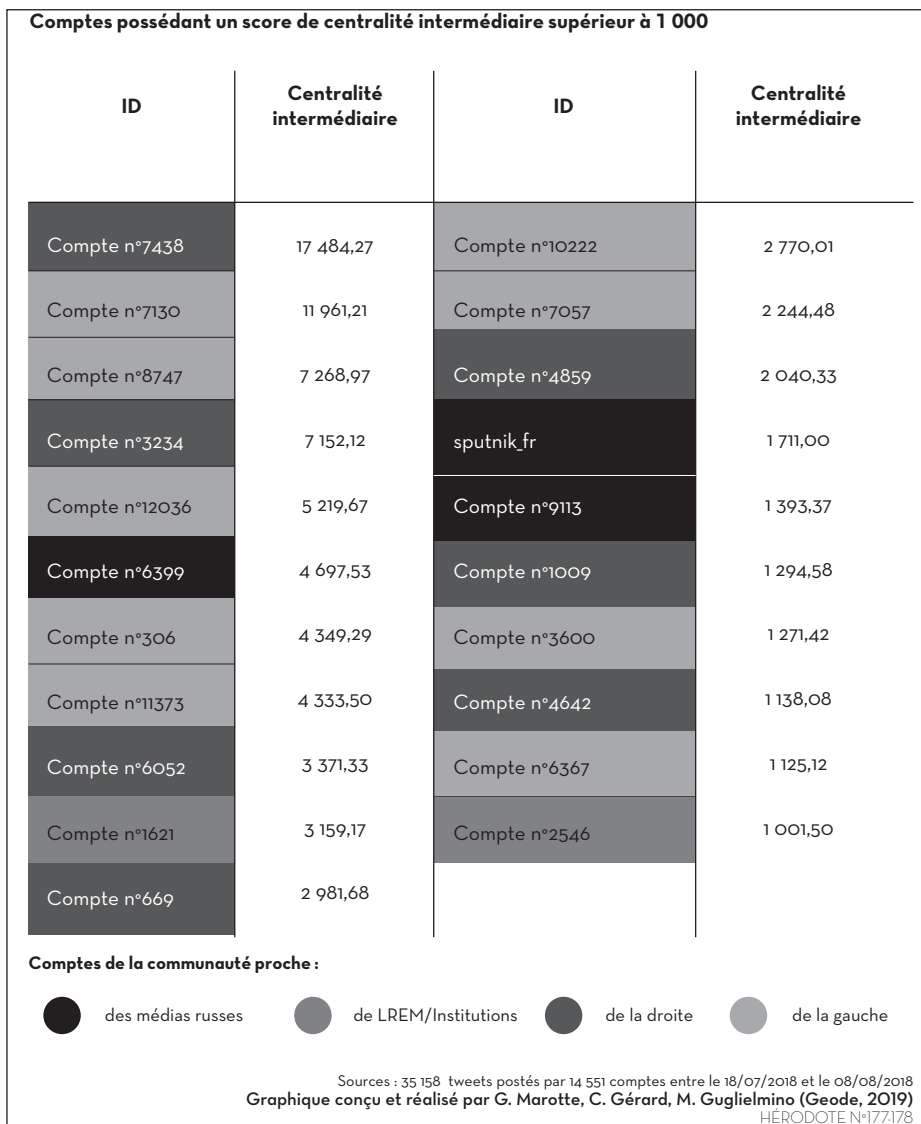
FIGURE 6. – REPRÉSENTATION SYNTHÉTIQUE DE L'ACTIVITÉ DES COMMUNAUTÉS SOCIOPOLITIQUES SUR TWITTER PENDANT L'AFFAIRE BENALLA



Hérodote, n° 177-178, La Découverte, 2^e et 3^e trimestres 2020.

Nous représentons ici la capacité de certains comptes à propager des contenus au sein du réseau global.

FIGURE 7. – COMPTES POSSÉDANT UN SCORE DE CENTRALITÉ INTERMÉDIAIRE SUPÉRIEUR À 1 000



Hérodote, n° 177-178, La Découverte, 2^e et 3^e trimestres 2020.

français sur la même période, ce qui nous aurait probablement permis de constater que l'activité de RT et Sputnik n'a rien eu d'extraordinaire au début de l'affaire Benalla. Toutefois, notre méthodologie d'extraction initiale ne nous permet pas de faire cette comparaison ici, l'ensemble des tweets publiés par des médias comme *Le Monde* ou *Libération* n'ayant pas été extrait.

En revanche, en dehors des comptes de grands médias ou de partis politiques, la centralité souligne l'influence dans des utilisations plus conventionnelles de Twitter. Un utilisateur poste un tweet, en retweete un autre : il y a ainsi de l'échange, et c'est fondamentalement cette notion que la centralité met en lumière. C'est ici le cas pour cinq comptes, qui sont à la fois des hubs d'informations (centralité) et des centrales médiatiques (degré entrant) comme le montre la figure n° 6.

Par conséquent, l'amplification d'une controverse sur un réseau structuré comme Twitter semble nécessiter, d'une part, une production de contenus, et d'autre part, des interactions entre le compte producteur et d'autres comptes. Ces interactions sont en effet déterminantes pour diffuser de l'information au-delà de sa propre communauté, et donc toucher un public le plus large possible. L'interaction entre les médias russes et d'autres comptes du graphe aurait par conséquent semblé nécessaire à l'amplification de la controverse autour de l'affaire Benalla. En effet, si cette controverse a bien eu lieu, son utilisation à des fins politiques provient plus de comptes proches de la gauche ou de la droite que de comptes « prorusses ». De même, la révélation de cette affaire au grand public n'est pas du fait des médias russes, mais bien du journal *Le Monde*. Ainsi, l'hypothèse selon laquelle la controverse autour de l'affaire Benalla aurait été amplifiée par une hyperactivité des médias russes et des comptes agrégés autour d'eux sur Twitter ne se vérifie qu'à moitié. Nous estimons donc que le rôle joué par RT et Sputnik a été similaire à celui joué par d'autres grands médias qui ont tous couvert de très près l'affaire Benalla.

Conclusion

Depuis la tentative d'ingérence russe dans l'élection présidentielle américaine de 2016, la stratégie d'influence informationnelle de la Russie est aujourd'hui un thème récurrent dans le débat politique de nombreux États. Les termes de « trolls russes » et de « bots russes », parfois employés de manière irrationnelle, sont à l'origine d'une représentation exagérée de la capacité de la Russie à influencer et manipuler les populations d'États tiers. Alors qu'aux États-Unis une partie importante des démocrates américains tient toujours la Russie comme responsable de l'élection de Donald Trump, la polémique liée à l'affaire Benalla illustre le fait que la représentation d'une menace informationnelle russe a traversé l'Atlantique

pour atteindre la France. En effet, grâce à l'activité de quelques centaines de comptes Twitter dont il est impossible de prouver qu'ils sont gérés par des Russes, la représentation d'une puissance informationnelle de la Russie basée sur une manipulation des réseaux sociaux en sort considérablement renforcée. Ainsi, l'influence concrète de la stratégie russe pourrait relever, outre ses caractéristiques techniques, de la manière dont elle est perçue et présentée par les acteurs politiques et médiatiques des pays concernés.

S'il n'est bien sûr pas question de nier la volonté de la Russie d'influencer les citoyens d'autres États par le biais de ses médias internationaux, l'exemple de l'affaire Benalla montre que l'impact de l'activité russe en France peut être relativisé si on l'analyse avec précision. En effet, notre article permet de distinguer deux types d'influence sur Twitter, que sont, d'une part, la production de contenus (les médias), et d'autre part, leur diffusion (les utilisateurs, qui agissent comme des autoroutes pour l'information produite). Dans le cas étudié, les médias russes ont simplement joué le rôle de producteurs de contenus, sans toutefois jouer un rôle central dans leur diffusion. L'hypothèse selon laquelle l'impact d'une stratégie d'influence ne se mesure pas uniquement selon des statistiques d'audience précises, mais également par l'interprétation qui est tirée de ces statistiques semble donc confirmée. S'il n'est pas exempt de biais méthodologiques, cet article s'inscrit dans la lignée des travaux entrepris depuis plusieurs années par le monde académique visant à comprendre ces nouveaux phénomènes d'influence en ligne. Le défi méthodologique reste toutefois conséquent, et devra faire l'objet de nouveaux travaux issus de différentes disciplines et étendus à d'autres plateformes que Twitter, de sorte à atteindre un modèle analytique optimal.

Bibliographie

- AUDINET M. (2018), «Diplomaties publiques concurrentielles dans la crise ukrainienne. Le cas de RT et Ukraine Today», *Revue d'études comparatives Est-Ouest*, vol. 2, n° 2.
- BAKSHY E., HOFMAN J., MASON W. et WATTS D. (2011), «Everyone's an influencer: quantifying influence on Twitter», *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*, WSDM, p. 65-74.
- BENKLER Y., FARIS R. et ROBERTS H. (2018), *Network Propaganda, Manipulation, Disinformation and Radicalization in American Politics*, Oxford, Oxford University Press.
- CARDON D. (2019), *Culture numérique*, Paris, Presses de Sciences Po.
- CATTARUZZA A. (2019), *Géopolitique des données numériques. Pouvoir et conflits à l'heure du big data*, Paris, Le Cavalier Bleu.
- DAVIS C. A., VAROL O., FERRA E., FLAMMINI A. et MENCZER F. (2016), «BotOrNot: a system to evaluate social bots», *Proceedings of the 25th International Conference Companion on World Wide Web*, p. 273-274.

- DOUZET F. (2014), « La géopolitique pour comprendre le cyberspace », *Hérodote*, vol. 152-153, n° 1, p. 3-21.
- HONG L. et DAVISON B. D. (2010), « Empirical study of topic modeling in Twitter », *Proceedings of the First Workshop on Social Media Analytics*, p. 80-88.
- JACOMY M., VENTURINI T., HEYMANN S. et BASTIAN M. (2014), « ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software, *PloS one*, 9(6) », e98679.
- KIRIYA I. (2016), « Russia Today comme dispositif diplomatique de la “Nouvelle Russie” », in MATTELART T. et KOCH O., *Géopolitique des télévisions transnationales d’information*, Paris, Mare et Martin Éditions, « Media critic ».
- LIMONIER K. (2018), « Diffusion de l’information russe en Afrique : essai de cartographie générale », note n° 66 de l’IRSEM.
- LIMONIER K. et GÉRARD C. (2017), « Guerre hybride russe dans le cyberspace », *Hérodote*, vol. 166-167, n° 3, p. 145-163.
- MATTELART T. (2018), « Russia Today et le renouveau de la diplomatie médiatique russe : de la projection mondiale du Kremlin à la “ré-information” en France », *Annuaire français de relations internationales*, vol. XIX.
- RAUCHFLEISCH A. et KAISER J. (2020), « The False Positive Problem of Automatic Bot Detection in Social Science Research », Berkman Klein Center Research Publication, N° 2020-3.