



HAL
open science

De quelle(s) couleur(s) sont les Gilets jaunes ? Plonger des posts Facebook dans un espace idéologique latent

Jean-Philippe Cointet, Pedro Ramaciotti Morales, Dominique Cardon, Caterina Froio, Andrei Mogoutov, Benjamin Ooghe, Guillaume Plique

► To cite this version:

Jean-Philippe Cointet, Pedro Ramaciotti Morales, Dominique Cardon, Caterina Froio, Andrei Mogoutov, et al.. De quelle(s) couleur(s) sont les Gilets jaunes ? Plonger des posts Facebook dans un espace idéologique latent. *Statistique et Société*, 2021, Gilets jaunes et Grand Débat National : outils, données et analyses, 9 (1-2), pp.79-107. hal-03856216

HAL Id: hal-03856216

<https://sciencespo.hal.science/hal-03856216>

Submitted on 16 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

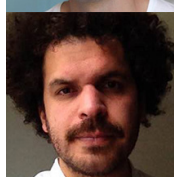


Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

De quelle(s) couleur(s) sont les Gilets jaunes ? Plonger des posts Facebook dans un espace idéologique latent



Jean-Philippe
COINTET¹



Pedro
RAMACIOTTI
MORALES²



Dominique
CARDON³



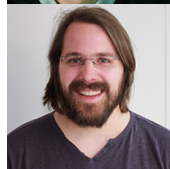
Caterina
FROIO⁴



Andreï
MOGOUTOV⁵



Benjamin
OOGHE TABANOU⁶



Guillaume
PLIQUE⁷

TITLE

What color are the Yellow Vests? – Ideological scaling of Facebook groups

RÉSUMÉ

Le refus de se positionner sur l'axe droite-gauche caractérise le mouvement des Gilets jaunes renvoyant sans cesse dos à dos les formations politiques plutôt que de prendre parti pour l'une d'entre elles. Pourtant les Gilets jaunes, lorsqu'ils font leur apparition en France, s'expriment dans un espace public déjà nourri de tensions et de structures idéologiques préexistantes. À ce titre, leur action est nécessairement située, elle s'inscrit dans cet espace et en hérite certaines propriétés. Il est dès lors légitime de s'intéresser à la place qu'occupe le mouvement, notamment dans sa déclinaison numérique sur Facebook. Comment les pratiques de citation en ligne trahissent-elles non pas la couleur politique du mouvement, mais l'espace politique dont ils se nourrissent et qu'ils alimentent ?

Cet article répond à cette question en introduisant un cadre méthodologique original qui permet d'étendre un plongement idéologique d'utilisateurs sur Twitter vers des posts publiés sur Facebook. Nous faisons d'abord appel à une analyse de correspondance pour réduire la matrice d'adjacence qui lie les parlementaires français à leurs followers sur Twitter. Cette première étape nous permet d'identifier deux axes latents qui sont déterminants pour expliquer la structure du réseau. La première dimension distribue les individus selon leur positionnement sur l'axe droite-gauche de l'espace politique. Nous interprétons la seconde dimension comme une mesure de la distance au pouvoir. Ces deux dimensions sous-tendent un espace dans lequel nous positionnons successivement des centaines de milliers d'utilisateurs de Twitter, les URLs et les médias cités sur cette plateforme et, par extension,

1. Sciences Po, médialab, Paris, jeanphilippe.cointet@sciencespo.fr
2. Sciences Po, médialab, Paris, pedro.ramaciottimorales@sciencespo.fr
3. Sciences Po, médialab, Paris, dominique.cardon@sciencespo.fr
4. Sciences Po, CEE, Paris, caterina.froio@sciencespo.fr
5. Sciences Po, médialab, Paris, andrei.mogoutov@sciencespo.fr
6. Sciences Po, médialab, Paris, benjamin.ooghe@sciencespo.fr
7. Sciences Po, médialab, Paris, guillaume.plique@sciencespo.fr

les publications de près de 1000 groupes Facebook parmi les plus actifs associés au mouvement des Gilets jaunes. Nous quantifions finalement l'évolution des publications de ces groupes dans l'espace idéologique latent pour donner à la fois un sens et une réponse à la question de l'inclinaison politique du mouvement. Les dynamiques observées renforcent l'interprétation d'un mouvement qui, d'abord positionné très à droite, a rapidement opéré un glissement vers la gauche tout en restant fidèle à une attitude contestataire. Cette description par l'usage que les Gilets jaunes font des médias sur Facebook illustre parfaitement l'idée d'un populisme polyvalent.

Mots-clés : *estimation idéologique, réseaux sociaux, mouvements sociaux, idéologies politiques, polarisation, populisme.*

ABSTRACT

«Neither right nor left» is the motto of the Yellow Vests, who prefer to dismiss traditional political formations back to back than to «take sides». Yet the Yellow Vests when they appear in France express themselves in a public space already nourished by tensions and pre-existing structures. As such, their action is necessarily situated, it is inscribed in this space, and inherits certain properties from it. It is therefore legitimate to take an interest in the place occupied by the movement, in particular in its digital version on Facebook. How do online citation practices betray not the political color of the movement, but the political space they feed off and feed?

This article answers this question by introducing an original methodological framework that makes it possible to extend the ideological embedding of Twitter users to posts published on Facebook. We first use a correspondence analysis to reduce the adjacency matrix that links French parliamentarians to their followers on Twitter. This first step allows us to identify two latent axes which are decisive in explaining the structure of the network. The first dimension seems to distribute individuals according to their position on the right-left axis of political space. We interpret the second dimension as a measure of distance to power. These two dimensions underlie a space in which we successively position hundreds of thousands of Twitter users, the URLs and media cited on this platform and, by extension, the publications of nearly 1,000 Facebook groups among the more active members of the Yellow Vests movement. We finally quantify the evolution of the publications of these groups in the latent ideological space to give both an interpretation and an answer to the question of the political orientation of the movement. The dynamics observed reinforce the interpretation of a movement which – initially fed by very right-wing resources – quickly operated a shift to the left while remaining faithful to a dissenting attitude. This description by the use that the Yellow Vests make of the media on Facebook perfectly illustrates the idea of a polyvalent populism.

Keywords: *ideology scaling, social media, social movements, political ideology, polarization, populism.*

1. Introduction

Le 17 novembre 2018, des centaines de milliers de manifestants arborant un gilet jaune se retrouvent dans la rue pour une grande journée de blocage des routes à travers toute la France. Ce premier acte de mobilisation, auquel succéderont de nombreux autres, naît d'une colère largement alimentée par les hausses successives du prix du carburant. La pétition publiée par Priscillia Ludosky sur la plateforme [change.org](https://www.change.org) et la vidéo de Jacline Mouraud, qui interpelle « Monsieur Macron » pour dénoncer la traque dont les conducteurs seraient victimes, sont souvent présentées comme les signes avant-coureurs d'un mouvement qui n'aura de cesse d'évoluer au gré des Actes qui s'incrémentent chaque samedi durant plus d'une année. Après avoir investi les péages autoroutiers, bloqué l'accès aux dépôts de carburant, occupé les ronds-points sur l'ensemble du territoire, les actions des Gilets jaunes convergent vers les lieux de pouvoir : les préfetures, l'Assemblée Nationale, jusqu'au très symbolique Arc de Triomphe.

Parallèlement, leurs revendications se sont élargies, appelant non seulement à la baisse des impôts mais aussi au renforcement des services sociaux de proximité, à une réforme constitutionnelle ambitieuse (le Référendum d'Initiative Citoyenne) et à une augmentation du pouvoir d'achat. À travers différents dispositifs de consultation, comme celui du Vrai débat (par opposition au Grand débat initié par le gouvernement) ou de l'Assemblée des assemblées, ils ont entrepris de définir un catalogue de revendications qui s'apparente à un programme politique (Suignard et al., 2021). Les Gilets jaunes apparaissent donc bien comme un mouvement social et politique de grande ampleur. Mais peut-on, pour autant, lui donner une couleur politique et l'inscrire dans la grille de positionnement idéologique qui nous sert traditionnellement à interpréter les mouvements sociaux d'une telle ampleur ?

Une telle question mérite d'être examinée avec précaution puisque cette mobilisation n'a cessé d'être scrupuleusement attentive à son indépendance et rétive à toute assignation idéologique. Elle s'est organisée indépendamment de tout parti politique et de toute structure syndicale. Elle a refusé de se donner une forme organisationnelle propre en interdisant tout mécanisme de délégation. Différents leaders de formations politiques ou de syndicats ont annoncé leur soutien au mouvement, se sont parfois prudemment affichés auprès des figures du mouvement afin de séduire leurs sympathisants. Mais dans les cortèges comme sur Facebook, les militants se sont gardés d'afficher des étiquettes indiquant leurs affiliations – et se sont fait rappeler à l'ordre par d'autres lorsqu'ils le faisaient. De façon relativement périphérique, deux listes « Gilets jaunes » ont émergé aux élections européennes de mai 2019, mais elles ont été massivement condamnées comme contraires à l'esprit du mouvement et n'ont rassemblé qu'un très faible nombre d'électeurs. Les travaux conduits sur les Gilets jaunes ont fait apparaître que le slogan « ni droite ni gauche » constituait une des caractéristiques les plus essentielles de son identité et un indispensable facteur d'unité de la mobilisation (Bendali et al., 2019). Les Gilets jaunes ont constamment cherché à s'émanciper des interprétations cherchant à les arraisonner dans la grille des idéologies politiques traditionnelles. Les enquêtes par questionnaire conduites auprès des participants aux mobilisations (André et al., 2019; Tristan Guerra, 2019) font ainsi apparaître leur refus de se situer sur l'axe gauche/droite : 52% des répondants se déclarent proche d'aucun parti politique et 35% refusent de se positionner sur un axe droite/gauche (André et al., 2019).

Si le mouvement des Gilets jaunes revendique une forme politique originale, il n'en a pas moins suscité un inlassable travail d'interprétation afin d'identifier sa nature politique et idéologique. C'est d'abord sur l'axe droite/gauche que beaucoup d'interprètes ont essayé de le fixer soulignant, notamment au début du mouvement, qu'il avait des affinités avec l'extrême-droite, avant qu'il n'apparaisse à d'autres observateurs comme un mouvement de gauche, puis que certains – en raison notamment de la présence de groupes violents dans les manifestations –, n'y voient une manifestation de l'ultra-gauche. Mais ces qualifications insatisfaisantes ont aussi donné jour à d'autres types de caractérisations, les éditorialistes prenant appui sur des épisodes de la mobilisation afin de lui donner une couleur d'ensemble : le partage de fausses informations a conduit à considérer que les Gilets

jaunes étaient « complotistes », les insultes reçus par Alain Finkielkraut qu'ils étaient « antisémites », la présence de Russia Today en Français aux côtés des manifestants qu'ils étaient « pro-russes », etc. Les intellectuels, eux aussi, ont proposé de nombreuses réflexions pour le rattacher à des mouvements historiques à la manière d'un nouveau mouvement des « sans culottes » (Sophie Wahnich dans Confavreux, 2019), d'une « révolte sociale populaire » (Noiriel, 2019) ou d'un « citoyennisme » (Samuel Hayat dans Confavreux, 2019).

Mais la plupart du débat – comme souvent – a tourné autour de la qualification controversée de « populisme » sans que celle-ci ne fasse consensus. L'effet combiné de revendications, de tactiques hétérogènes et de positions anti-gouvernementales fortes a conduit certains commentateurs à qualifier les Gilets jaunes de mouvement populiste. Tarragoni et Gaudillière (2019) font des Gilets jaunes la déclinaison française d'autres manifestations de populisme contemporaines de la Grande Récession qui a frappé l'Europe à la fin des années 2000 et qui a contribué aux succès électoraux de Podemos en Espagne ou de M5S en Italie (Della Porta et al., 2017). Mais cette qualification est récusee par d'autres, comme Etienne Balibar, qui y voit plutôt un « contre-populisme » (dans Confavreux, 2019). Sans doute l'une des spécificités des Gilets jaunes est-elle de ne pas se laisser caractériser facilement. Beaucoup de recherches sur la composition sociale du mouvement ont montré son hétérogénéité, la diversité des parcours biographiques, la variété des professions et des situations territoriales. La difficulté à établir une frontière entre engagés, soutiens et sympathisants a été soulignée, ainsi que la très grande fluidité des dynamiques de participation qui a vu la composition sociologique du mouvement se transformer tout au long de la mobilisation. Cette incertitude sur la qualification du mouvement ne doit cependant pas conduire à renoncer à dessiner l'espace idéologique au sein duquel il s'est déployé.

On voudrait montrer dans cet article que les méthodes numériques peuvent proposer une solution originale pour situer les Gilets jaunes dans un espace idéologique, sans imposer *a priori* une échelle de positionnement politique héritée de la traditionnelle division droite/gauche. L'objectif n'est pas tant ici de renouveler les catégories de l'analyse idéologique que de parvenir à les identifier à travers une démarche sans *a priori*, c'est-à-dire de type inductif.

On peut en effet considérer qu'il existe deux manières de conférer un positionnement idéologique aux acteurs sociaux. La première consiste à leur demander de se positionner de façon explicite sur un axe droite-gauche, ou bien d'inférer ce positionnement de leurs votes et de leurs engagements passés (Algan et al., 2019b). Cette méthode présente l'avantage de mesurer, de façon compréhensive, les attachements à l'axe gauche-droite que les acteurs expriment eux-mêmes – en supposant que ces coordonnées fassent l'objet d'une connaissance commune et partagée permettant à chacun de se positionner relativement aux autres sur une échelle. La seconde stratégie ne demande pas leurs avis aux acteurs, mais entreprend de les classer sur la base d'une analyse des enjeux saillants des programmes et des revendications (Volkens et al., 2009), du discours de leurs leaders ou des idées auxquelles ils se réfèrent. Il arrive dans les travaux comparatifs d'analyse de discours et d'idéologies que l'on sollicite des panels d'experts pour opérer ce travail de qualification au sein d'une grille de positionnement idéologique des phénomènes sociaux (Hooghe et al., 2010; Rooduijn et al., 2019).

La méthodologie que nous nous proposons de déployer dans cet article propose une troisième stratégie. On souhaite montrer qu'il est possible d'inférer le positionnement idéologique des acteurs en les plongeant dans un *espace idéologique latent* qui a été construit de manière inductive à partir de traces numériques à grande échelle. La méthode que nous allons décrire dans cet article n'est pas compréhensive, au sens où elle est indépendante des identifications produites réflexivement par les acteurs. Elle n'est pas non plus herméneutique, au sens où elle s'appuierait sur des constructions intellectuelles déjà établies dans l'histoire des idées. Ce qui la caractérise est de faire émerger un instrument que nous appellerons un *espace idéologique latent*, ni des déclarations des acteurs ni des interprétations des experts, mais plutôt d'un volume considérable de comportements numériques initialement induits par le fait, pour des milliers d'internautes, de suivre des personnalités politiques sur Twitter, puis de partager des liens hypertextes sur les réseaux sociaux. La structure de cet espace

multidimensionnel émerge du calcul automatique d'un très grand volume de traces numériques (dont le périmètre a été construit à partir des parlementaires et sénateurs). Il constitue une sorte d'architecture, une géométrie, au sein de laquelle il est ensuite possible de positionner, par extension, un ensemble très différent d'entités. Nous allons décrire, pas à pas, les différentes étapes qui permettent de construire cet espace puis de le déplacer progressivement vers les groupes Gilets jaunes de Facebook. Mais, pour l'introduire, il faut souligner que si une telle démarche n'est évidemment ni neutre, ni complètement immanente, elle invite cependant à identifier les forces et les limites pour les sciences sociales d'une approche inductive et comportementaliste basée sur des données massives. On montrera qu'elle permet de retrouver des résultats conformes à ceux obtenus par des méthodes plus traditionnelles en sciences politiques, mais qu'elle offre aussi de nouveaux instruments d'exploration et d'analyse.

2. De l'Assemblée nationale aux Gilets jaunes en passant par les médias

2.1. Un court panorama des méthodes d'inférence idéologique

La littérature sur l'inférence des positions idéologiques renvoie classiquement à une famille de méthodes très variées en sciences politiques (Imai et al., 2016) qui exploitent les données de vote de parlementaires. Ces méthodes sont entièrement inductives et ne nécessitent en général pas de connaître les étiquettes politiques des individus *a priori*, le positionnement idéologique latent des représentants résulte d'une opération de réduction de dimensionnalité qui assigne aux individus des positions qui dépendent exclusivement de leurs décisions de vote passées et plus précisément des différences entre leurs décisions de vote. Les travaux pionniers de Poole et Rosenthal (1991) ont ainsi permis de mesurer le degré de polarisation du Sénat américain (Bonica, 2014) mais aussi en Europe (Hix et al., 2006) ou de suivre les déplacements idéologiques de certains législateurs. Différentes méthodes d'inférence ont été développées sur des principes similaires (Poole et al., 2008; Clinton et al., 2004), la plus connue s'intitulant NOMINATE (pour Nominal Three-Step Estimation).

Ces méthodes ne permettent néanmoins que de positionner des législateurs. Si la forme de l'espace politique n'est pas spécifiée *a priori* et ne se réduit pas forcément à une unique dimension, seuls les membres d'une assemblée élue peuvent être soumis à un tel calcul. Plus récemment, des méthodes ont été développées qui permettent avec une précision croissante de prédire si un discours, ou une question parlementaire, a été posée par un individu provenant d'un parti ou d'un autre (souvent républicain ou démocrate) (Gentzkow et al., 2019; Jensen et al., 2012). Ces méthodes textuelles s'appuient sur une autre forme de signal que les votes parlementaires et peuvent prétendre à ce titre à étendre l'univers des entités susceptibles d'être plongées dans un espace idéologique latent associé. Pour autant, l'échantillon d'apprentissage sur lequel elles se fondent est si spécialisé que l'extension est limitée. Rheault et Cochrane (2020) déploient ainsi le formalisme des plongements lexicaux (*word embeddings*) pour inférer le positionnement idéologique de textes parlementaires. De façon connexe, mais à nouveau dans un domaine d'application relativement limité, Groseclose et Milyo (2005) mesurent le biais idéologique des médias en fonction de leur propension à citer certains *think tanks*.

L'arrivée des plateformes de réseaux sociaux a permis aux méthodes d'inférence de sortir des hémicycles pour rendre compte du positionnement relatif d'une grande variété d'acteurs dans l'espace public numérique (Ecornier-Nocca et Sauger, 2020). Bond et Messing (2015) ont été parmi les premiers à appliquer massivement ces méthodes à l'analyse des dynamiques de positionnement politique des internautes. Dans leur étude, plus de 6 millions d'utilisateurs de Facebook suivant au moins deux des 1200 pages politiques « officielles » hébergées par Facebook à l'époque sont projetés sur

un axe droite/gauche. Naturellement, ces données ne sont exploitables que par les chercheurs de Facebook, c'est pourquoi l'essentiel des efforts récents en matière d'inférence d'idéologie politique depuis les réseaux sociaux se sont portés sur Twitter, dont l'API permet de récolter un grand nombre de données aisément. Il est notamment possible de collecter les relations sociales en ligne qui lient les utilisateurs de la plateforme à leurs abonnés (*followers*). Pablo Barberà a popularisé l'utilisation des données de Twitter pour inférer un espace idéologique (Barberà, 2015; Barberà et al., 2015). Sa méthode s'inspire en grande partie de celle déployée par Bond et Messing (2015) sur Facebook, mais elle se fonde sur l'analyse de la liste des *followers* de l'ensemble des représentants élus d'un pays (députés et sénateurs). L'idée est simple, Barberà et ses collègues font l'hypothèse que deux individus dont les convictions politiques sont proches tendront à suivre les mêmes comptes politiques sur Twitter. Cette hypothèse d'homophilie classique leur permet d'estimer le positionnement politique de comptes à partir de la seule observation des comptes politiques qu'ils suivent. Les individus héritent ainsi d'un score sur un axe unidimensionnel censé refléter leur positionnement idéologique. Cette position politique est par la suite inférée à l'aide d'un modèle bayésien qui suppose que chaque individu est défini par un score d'idéologie latent dont la valeur infléchit dans un sens ou dans l'autre les probabilités qu'a l'individu de suivre tel sénateur, député ou média.

Dans l'article original, la méthode est appliquée à 6 pays différents (les États-Unis, l'Espagne, l'Italie, l'Allemagne et les Pays-Bas), et validée par de nombreuses mesures exogènes à Twitter (Barberà et Rivero, 2015). Cette méthode a déjà été testée en France par Briatte et Gallic (2015) qui ont mesuré les scores d'idéologie de 1000 comptes centraux dans le paysage politique français. Dans cet article nous adoptons le même principe, mais exploitons le positionnement idéologique de centaines de milliers de comptes dans l'espace induit, afin de caractériser *in fine* l'espace idéologique latent dans lequel évoluent des groupes Facebook de Gilets jaunes. Notre stratégie détaillée est exposée dans la section suivante.

2.2. Une méthode en trois étapes

Notre étude se fonde sur la façon dont les Gilets jaunes s'expriment sur Facebook. Sans y être réductible, la plateforme constitue un espace privilégié pour le mouvement (Boyer et al., 2020). Nos données n'ont pas vocation à représenter de manière exhaustive l'idéologie exprimée par les Gilets jaunes. Nous nous proposons donc de « sonder » et de cartographier l'espace idéologique des Gilets jaunes à travers les traces de leurs interactions numériques.

Notre méthode s'articule en trois étapes, représentées graphiquement dans la Figure 1. Un espace idéologique latent est d'abord construit à partir du réseau des utilisateurs de Twitter qui suivent les députés et sénateurs français. Ces scores sont par la suite extrapolés aux URLs que partagent les comptes Twitter positionnés dans l'espace. Enfin, on attribue aux posts individuels sur Facebook la « couleur politique » des liens qu'ils contiennent. Par suite l'espace idéologique qu'occupe un groupe Facebook est décrit par le nuage des positions idéologiques de chacune des URLs (et en toute rigueur, la distribution idéologique des comptes les ayant partagées). La même opération peut être effectuée à l'échelle d'un média en ligne. Un quotidien en ligne, ou une chaîne YouTube, forment une distribution dans l'espace idéologique qui correspond à l'ensemble des scores des comptes Twitter ayant partagé des URLs provenant de ce média.

In fine, cette analyse par les liens hypertextes nous renseigne sur la *variété* (au sens géométrique) que dessine le mouvement des Gilets jaunes (Section 5.2). Soumise à de fortes évolutions, une lecture de ses déplacements à différentes échelles permet de reconstruire une compréhension inédite du mouvement. L'espace idéologique induit depuis Twitter est, au sens propre du mot, virtuel, c'est-à-dire potentiel. Il se présente à la fois comme une *atmosphère* puisqu'il est extérieur aux acteurs qu'ils positionnent « virtuellement », et comme une *structure* à partir de laquelle il est possible d'établir des positionnements relatifs dans différentes dimensions qui peuvent être déplacées vers d'autres types

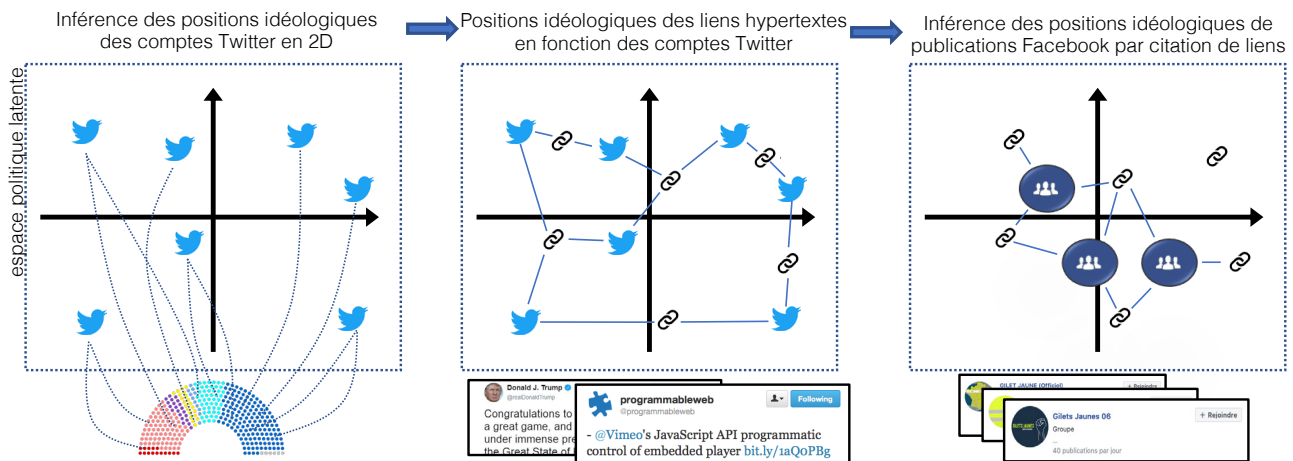


FIGURE 1 – La décomposition de notre méthodologie en trois étapes. De gauche à droite : inférence d'un espace idéologique des comptes Twitter (via leurs liens vers les députés et sénateurs), calcul des scores des URLs comme position moyenne des comptes les ayant partagées, propagation de ces scores aux posts puis aux groupes Facebook ayant partagé les mêmes liens.

de données. C'est précisément cette mobilité qui va nous permettre d'étendre, en propageant les étiquettes politiques *via* les liens hypertextes, l'espace idéologique latent des parlementaires vers les posts Facebook des Gilets jaunes.

Les trois étapes illustrées dans la Figure 1 serviront également de guide pour les trois parties suivantes. Dans la première, nous présentons la façon dont nous appliquons une méthode d'inférence des idéologies depuis des données collectées sur Twitter pour assigner à plusieurs centaines de milliers d'utilisateurs une position dans un espace politique que nous nous attacherons à décrire et valider (section 3). Ces positions individuelles nous servent par la suite à caractériser un ensemble d'URLs, provenant souvent de médias, qui sont partagées par nos utilisateurs étiquetés (section 4). Enfin, dans la dernière partie, on exploite ce positionnement des articles de presse, vidéos et autres ressources en ligne pour projeter les posts produits par les groupes Gilets jaunes sur Facebook dans cet espace virtuel (section 5.1).

3. Inférence de l'espace politique latent depuis Twitter

Avant de projeter les interactions des Gilets jaunes sur Facebook dans l'espace politique, nous construisons d'abord cet espace politique latent depuis Twitter. Nous introduisons la littérature sur l'inférence de scores idéologiques dans la première section 2.1, présentons les jeux de données sur lesquels nous nous appuyons (section 3.1), avant de décrire l'espace ainsi construit (section 3.2) et d'en proposer une évaluation empirique (section 3.3).

3.1. Collecte de données relationnelles sur Twitter

La première étape de cette méthodologie vise à inférer le positionnement politique d'un ensemble de comptes Twitter qui suivent des parlementaires. Dans son papier original, Barberá (2015) considère un ensemble d'acteurs relevant du monde de la politique : des représentants élus au niveau national, les comptes Twitter de partis politiques, et ceux des principaux médias d'information politique du pays. Les relations qu'établissent ou non des comptes Twitter anonymes avec ces comptes fortement discriminant politiquement permettent de mesurer leur position latente sous la forme d'un simple scalaire dont la valeur est inférée par une méthode de Monte-Carlo par chaînes de Markov. Nous

avons simplifié le protocole et uniquement considéré des élus parlementaires en cours de mandat à l'Assemblée nationale ou au Sénat en 2019.

Nous avons d'abord constitué à l'aide des données collectées et maintenues quotidiennement par l'association Regards Citoyens⁸, la liste des 883 comptes Twitter tenus par des députés (545) ou sénateurs français (338).

Nous avons ainsi collecté la liste exhaustive des « followers » des parlementaires utilisateurs de Twitter en mai 2019. Au total, 856 de ces comptes (dont la liste est annexée en section A) disposaient d'un compte actif dont nous avons extrait l'ensemble des followers le 22 mai 2019⁹. Nous avons ainsi identifié 4 487 430 comptes distincts qui suivaient au moins l'un de ces comptes. Ce réseau bipartite a ensuite été réduit à un réseau composé de 979 606 comptes suivant au moins 3 des 856 élus. Enfin, nous avons retiré des données initiales les comptes Twitter dont le profil était doublonné¹⁰ pour ne conserver finalement que 368 831 comptes.

3.2. Inférence des positions idéologiques des utilisateurs de Twitter

À partir de ces données, nous avons extrait une matrice d'adjacence A composée de 368 831 utilisateurs qui suivent au moins 3 des 856 personnalités politiques retenues. Soit i l'un des 368 831 utilisateurs et j l'un des 856 comptes des personnalités politiques retenues. La matrice d'adjacence finale est une matrice binaire définie telle que :

$$A(i, j) = \begin{cases} 1 & \text{si } i \text{ suit } j \\ 0 & \text{si } i \text{ ne suit pas } j. \end{cases}$$

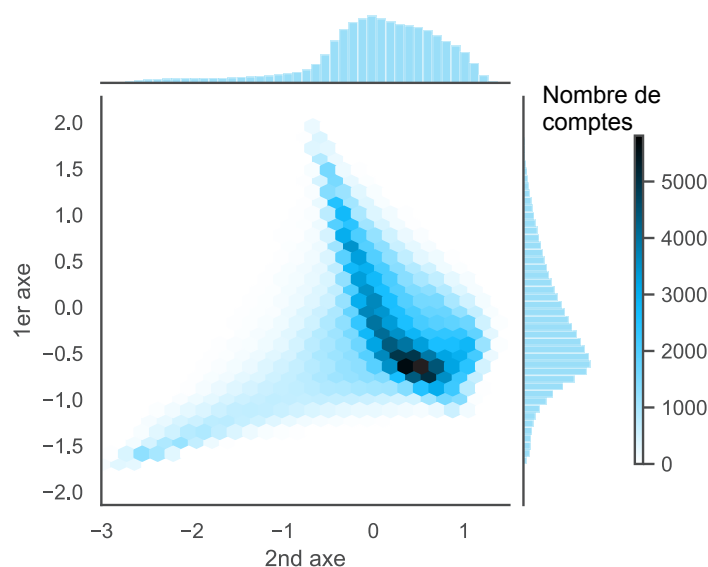


FIGURE 2 – Distribution des 368 831 comptes Twitter positionnés dans l'espace idéologique à deux dimensions

8. <https://github.com/regardscitoyens/twitter-parlementaires/blob/master/data/deputes.csv>,
<https://github.com/regardscitoyens/twitter-parlementaires/blob/master/data/senateurs.csv>

<https://github.com/regardscitoyens/twitter-parlementaires/blob/master/data/senateurs.csv>

9. Nos remerciements vont à Justin Clark pour avoir réalisé cette collecte, et de nous avoir largement guidés dans l'implémentation de la méthode d'inférence de Barberá.

10. Des centaines de milliers de comptes suivent exclusivement les 3 comptes 'MLP_officiel', 'jlmelenchon', 'manuelvalls'; des dizaines de milliers suivent exclusivement (parmi l'ensemble des députés et des sénateurs) les 5 comptes 'MLP_officiel', 'BrunoLeMaire', 'jlmelenchon', 'JVPlace', 'jclagarde', 'manuelvalls'. La très forte notoriété de ces comptes a tendance à déformer l'espace construit du fait de la très grande population de comptes qui occupent la même position.

L'analyse de correspondances¹¹ fait apparaître deux axes principaux qui expliquent respectivement 1.07% et 0.97% de l'inertie totale. On a représenté la distribution des positions des 856 comptes plongés dans cet espace latent à la Figure 2. Il apparaît clairement que notre population initiale de comptes est distribuée de façon très singulière. L'interprétation des axes ainsi construits nécessite de représenter certains comptes dont l'étiquette politique est connue. Il semble naturel de représenter les députés et sénateurs qui sont présents à la fois en ligne et en colonnes de la matrice d'adjacence de départ (ces derniers étant également susceptibles de suivre leurs collègues dans les différents hémicycles).

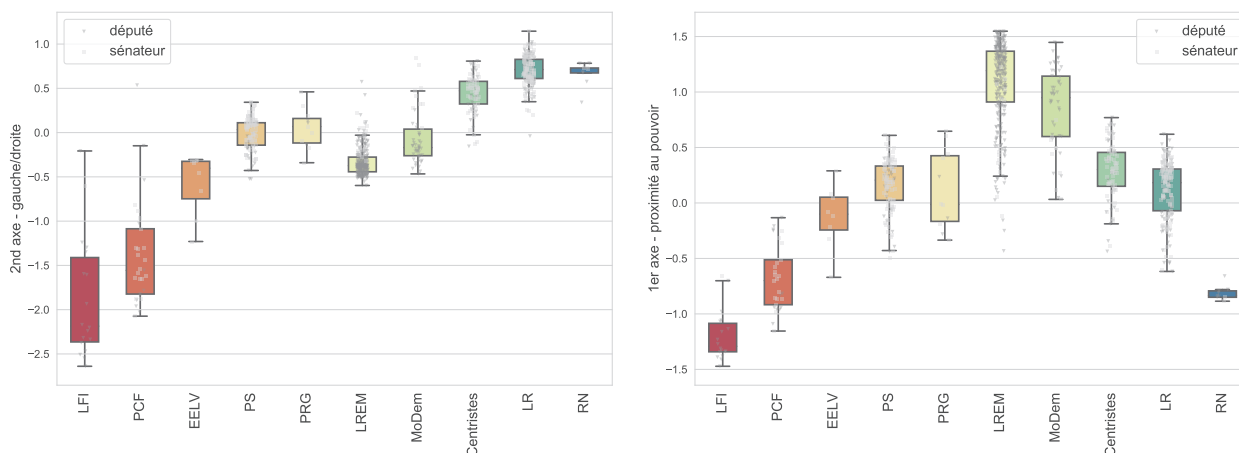


FIGURE 3 – Boîtes à moustaches présentant la distribution des principaux partis présents dans notre échantillon sur les deux axes principaux de l'espace politique latent. Le second axe ordonne clairement les partis selon leur orientation idéologique du parti le plus à gauche à celui le plus à droite ; le premier axe distribue les partis selon leur proximité aux institutions.

Nous avons d'abord repéré (Figure 3) la distribution sur les deux axes du positionnement idéologique de profils colonnes de l'ensemble des députés et sénateurs. L'ordre qui émerge permet de proposer une première interprétation de nos axes qui semble opposer pouvoir et opposition pour le premier, droite et gauche pour le second. Ces qualifications apparaissent évidentes même si la méthode d'inférence automatique produit aussi quelques variations à l'égard des résultats attendus par un politologue exercé. La distribution des députés et sénateurs est d'abord tendue sur le premier axe par le rapport au pouvoir et aux institutions. Les députés et sénateurs de la majorité gouvernementale (LREM et Modem) tirent la distribution en s'opposant aux députés et sénateurs de La France insoumise, du PCF et du Rassemblement national. La proximité ou la distance au gouvernement et aux institutions marque ainsi clairement la plus ou moins grande intégration aux instances de pouvoir en séparant les membres des partis ayant une expérience gouvernementale régulière des autres. Le deuxième axe distribue clairement les sénateurs sur la polarité droite/gauche. Il fait cependant apparaître une anomalie dans le positionnement des députés et sénateurs LREM qui sont positionnés légèrement plus à gauche que les parlementaires socialistes. Bien que les explications de ce phénomène soient certainement multiples, on peut penser que, nouveaux entrants dans la profession politique, les députés et sénateurs LREM n'ont pas été identifiés politiquement par les utilisateurs de Twitter au moment de la collecte (un an après l'élection) et que la distribution de leurs abonnés est en conséquence beaucoup plus flottante. Cette interprétation est confortée par le fait que, lorsque l'on observe unitairement la position des parlementaires, les quelques députés LREM disposant d'un capital parlementaire plus ancien (parce qu'ils étaient précédemment associés soit au PS soit aux Républicains) se positionnent alors sur l'axe droite/gauche de façon beaucoup plus conforme à ce qui est attendu.

11. Elle a été réalisée à l'aide de la librairie *Prince*.

Mais il est également possible de projeter dans ce même espace les profils lignes de n'importe lequel des comptes suivant au moins 3 sénateurs et députés de notre base et de nombreuses figures politiques – y compris les parlementaires eux mêmes – font naturellement partie de cet ensemble. On a ainsi projeté sur la Figure 4 l'ensemble des représentants élus français auxquels nous avons adjoint un certain nombre de figures politiques (anciens ministres, têtes de liste aux dernières élections européennes) que nous pouvions affilier à l'un des treize partis suivants : La France Insoumise (LFI - 39 membres), le Parti Communiste Français (PCF - 43), Europe-Écologie Les Verts (EELV - 31), Génération.s (16), le Parti Socialiste (PS - 236), le Parti Radical de Gauche (PRG - 16), La République en Marche (LREM - 365), le MoDem (49), les centristes (UDI, AGIR... - 107), Les Républicains (LR - 293), Debout la France (DLF - 16), les Patriotes (10) et le Rassemblement National (RN - 22).

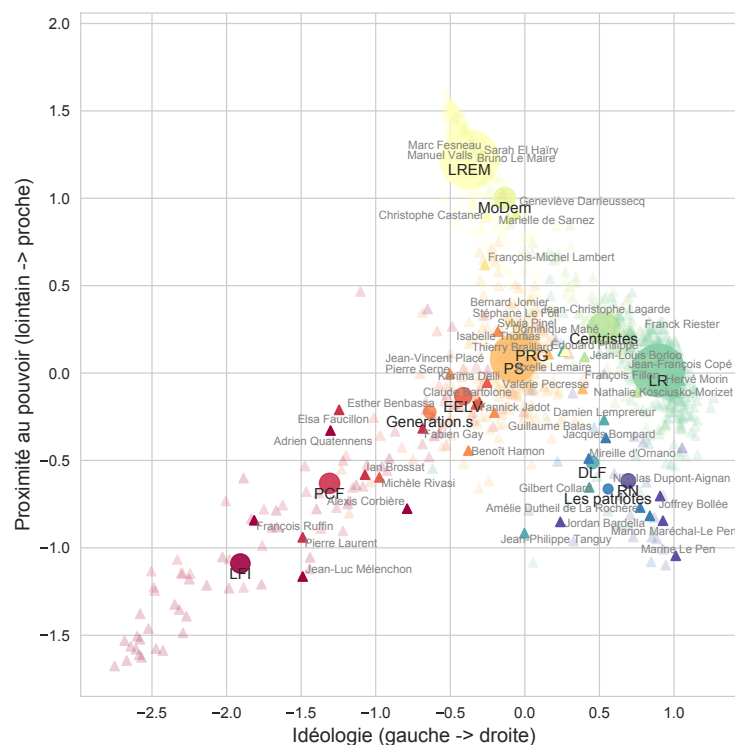


FIGURE 4 – Positionnement de près d'un millier de comptes politiques (députés, sénateurs, candidats aux dernières élections européennes, ministres ou anciens ministres) dans l'espace latent induit par l'analyse de correspondance. Quelques comptes individuels sont étiquetés. Les comptes de chaque personnalité sont colorés en fonction de leur parti, dont la position moyenne est figurée par un cercle et dont la taille est proportionnelle au nombre de leurs représentants. Les 4 comptes les plus suivis de chaque parti sont étiquetés par leur nom complet.

3.3. Validation par comparaison aux scores inférés à partir des scrutins

Une manière de valider notre méthode fondée sur l'analyse des réseaux sociaux est de comparer la distribution des députés et sénateurs aux scores que l'on obtient classiquement pour ordonner les représentants d'un hémicycle en fonction de leurs votes successifs. Pour faire cette vérification, nous avons corrélé les positions idéologiques des comptes des parlementaires obtenues sur Twitter avec celles que l'on peut inférer à partir de leurs votes au Sénat et à l'Assemblée nationale. Classiquement, l'analyse de données dites de « *roll call* » permet d'ordonner les députés ou les sénateurs en fonction de la similarité de leur choix de vote (Poole et Rosenthal, 1984; Clinton et al., 2004). En ce qui concerne l'Assemblée nationale, nous avons produit un score associé à chaque député à l'aide

des données de votes collectées sur la première année de la 15^e législature, soit 995 scrutins qui s’étendent entre le 4 juillet 2017 et le 26 juin 2018¹². Pour le Sénat, nous avons considéré l’ensemble des votes de la session parlementaire 2017-2018 auxquels nous avons accès à l’époque, soit 139 scrutins entre le 18 octobre 2017 et le 19 juin 2018¹³.

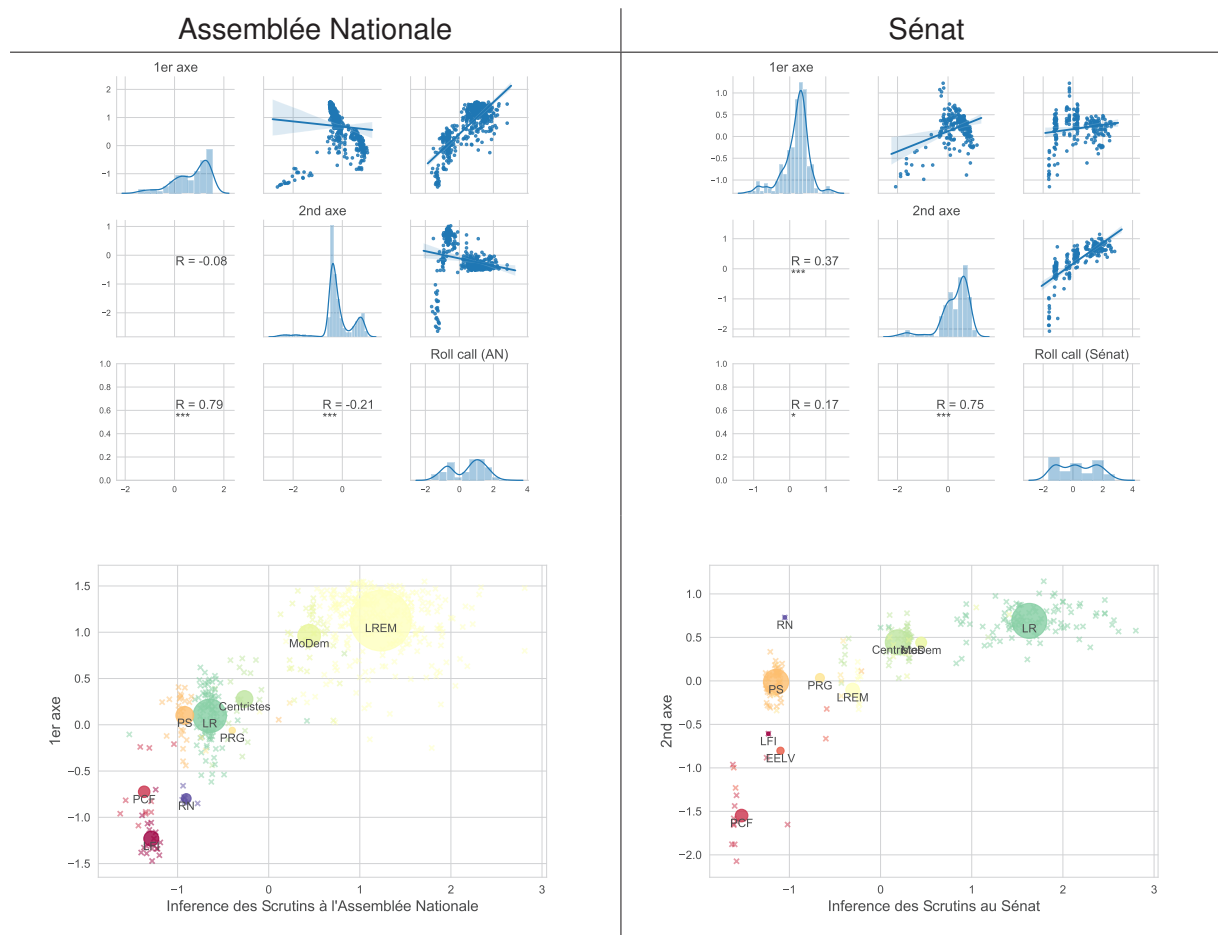


FIGURE 5 – Haut : corrélations entre les deux axes de notre analyse de correspondance (score en colonnes) et les scores attribués aux sénateurs et aux députés en fonction de leurs votes au Parlement. Les votes au Sénat sont très bien corrélés avec le second axe, les votes à l’Assemblée avec le premier axe. Bas : détail des corrélations par familles politiques à l’Assemblée nationale (gauche) et au Sénat (droite)

575 députés et 349 sénateurs figurent dans ce jeu de données dont nous avons inféré la position à l’aide d’un modèle de « *spatial vote* » à une dimension¹⁴. Nous avons ainsi pu associer un score à 484 députés et 224 sénateurs dont on a également inféré une position sur Twitter. Par suite, nous avons testé la corrélation (score de Spearman) entre les positions des représentants sur les deux axes obtenus à partir de l’analyse de Twitter avec le score résultant de leurs votes dans les deux chambres. Les tableaux de corrélations associés sont représentés à la Figure 5. Chaque dimension de notre analyse de correspondance est fortement corrélée à l’ordre produit par les votes des représentants dans les deux chambres. Le premier axe (que nous avons interprété comme une distance au pouvoir) est ainsi fortement lié aux logiques qui guident la façon dont les députés votent ($R = .79$). Le second axe (valence droite/gauche) est fortement lié à l’ordre produit en observant le vote des sénateurs

12. Ces données (ainsi que celles des sénateurs) ont été compilées par l’association Regards Citoyens et sont hébergées sous une forme agrégée à cette adresse : <https://git.regardscitoyens.org/regardscitoyens/sprint-scrutins/-/tree/master>

13. Voir le site du Sénat pour un détail des scrutins : <https://www.senat.fr/scrutin-public/scr2017.html>

14. La librairie “pscl” (Jackman et al., 2017) a été utilisée pour analyser (via MCMC) ces données. Nous remercions Jason Kaiser pour ses conseils dans son utilisation.

($R = .75$). La valeurs des coefficients de corrélation n'est pas aussi importante que celle mesurée sur une seule dimension par Barberá (2015), mais elle est suffisamment importante sur les deux axes pour indiquer que l'espace latent construit capture bien un ordre provenant de l'activité parlementaire.

4. Espace des médias

Une fois cet espace idéologique latent constitué, la méthode mise en œuvre dans cette recherche consiste à le propager vers de nouvelles entités afin de les positionner dans l'espace bi-dimensionnel que les parlementaires nous ont permis de constituer. L'architecture de l'espace reste la même, mais nous proposons d'en modifier la population. Afin de propager ces étiquettes vers les Gilets jaunes qui s'expriment sur Facebook, la deuxième étape de notre méthodologie consiste à positionner dans le même espace idéologique latent les URLs partagées sur Twitter. En agrégeant les URLs provenant d'un même domaine¹⁵, il est possible de décrire la position occupée par des médias, des associations ou des chaînes YouTube en la définissant simplement comme la somme des orientations politiques des internautes qui la partagent.

4.1. Données de partage d'URLs sur Twitter

Nous avons sélectionné, dans un premier temps, un échantillon aléatoire de 40 000 comptes Twitter parmi les comptes engagés dont nous pouvons inférer la position idéologique (*i.e.* les comptes qui suivent au moins 3 sénateurs et/ou députés). Nous avons ensuite capturé leurs tweets les plus récents (jusqu'à 3 200 pour les plus actifs) grâce à l'API de Twitter. Nous avons enfin extrait de ce corpus de tweets, l'ensemble des liens hypertextes partagés (39 392 tweets ont partagé au moins un lien). Au total nous dénombrons 23 534 803 liens citant 13 279 718 URLs distinctes. Après normalisation des noms de domaines, il est possible d'identifier pas moins de 419 256 domaines différents dans les données. Différentes opérations peuvent être réalisées à l'échelle de ces domaines qui correspondent souvent à des médias.

4.2. Un média, des publics variés

La première opération à laquelle nous nous sommes livrés a simplement consisté à représenter la projection dans l'espace idéologique d'un ensemble d'articles publiés sur un même site de média.

La Figure 6 (partie gauche) illustre ainsi la façon dont des articles publiés sur un même média peuvent générer un écho sur Twitter auprès de publics idéologiquement variés. Les articles de *Russia Today en France*, média qui joue constamment de l'ambivalence de son positionnement, en sont une parfaite illustration : certains sont partagés par des comptes Twitter positionnés à gauche quand d'autres sont en revanche partagés par des comptes positionnés à l'extrême-droite.

Pour autant, lorsqu'on agrège l'ensemble des partages associés à un média ou un domaine particulier, la distribution bidimensionnelle associée est presque toujours unimodale. La répartition des articles partagés par *Le Monde*, *Reporterre* et *Valeurs actuelles* sur la Figure 6 fait apparaître très clairement le positionnement contrasté de ces publications dans l'espace idéologique latent. Nous avons systématiquement testé les domaines les plus partagés totalisant plus de 5000 partages (soit environ 400 sites), et pour chacun d'entre eux, la statistique de Phi-Means (Siffer et al., 2018) utilisée a détecté une distribution unimodale dans 97% des cas¹⁶. L'essentiel des domaines ne présentant pas une distribution unimodale sont des sites dont on pouvait croire qu'ils se projettent de façon

15. Le nom de domaine est la première entité d'une adresse du web, comme <http://www.insee.fr>.

16. La statistique a été calculée avec un taux de confiance de 95%.

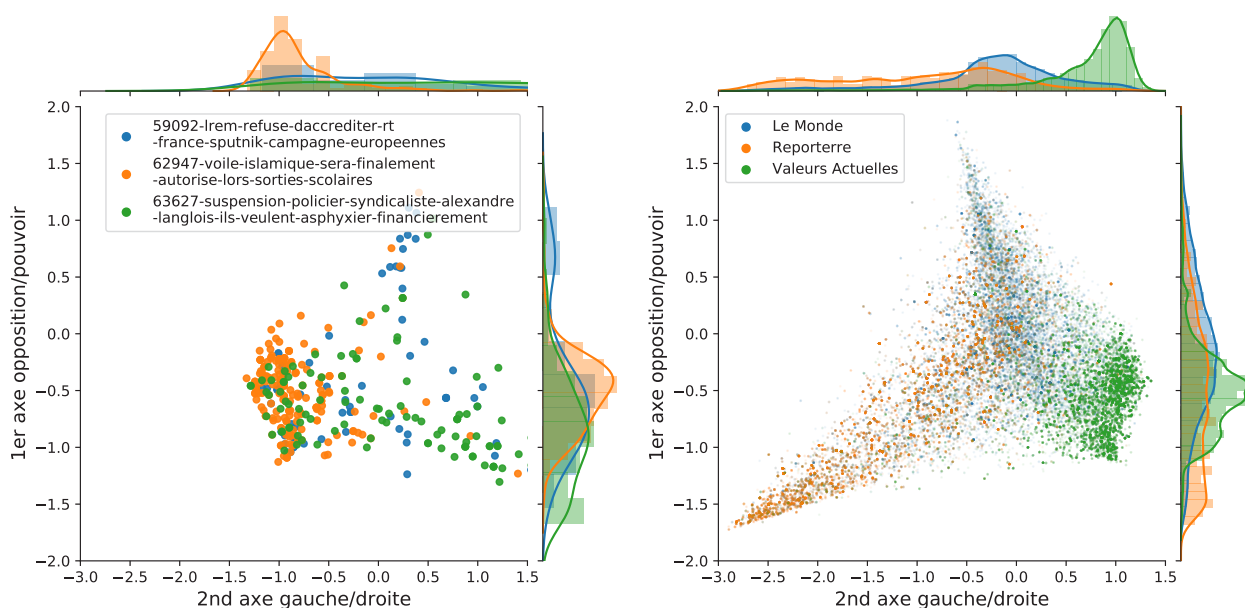


FIGURE 6 – [Gauche] Profil idéologique de trois URLs partagées sur Russia Today. L'article sur l'autorisation du voile islamique lors des sorties scolaires a été largement repris par une audience plutôt à gauche. A contrario les articles sur le refus d'accréditer RT et Sputnik pour couvrir les élections européennes et sur la suspension d'Alexandre Langlois, policier syndicaliste à la CGT, ont majoritairement été partagés par un public plutôt droitiste (et certains plus proches du pouvoir pour le premier). [Droite] Distribution idéologique d'un échantillon aléatoire de 20,000 partages d'URLs provenant du Monde, de Reporterre et de Valeurs Actuelles. L'orientation politique des publics des différents médias est bien visible dans leur distribution jointe sur les deux axes.

isotrope dans l'espace idéologique soit parce qu'ils correspondent à des services web, ou à des sites marchands (*unfollow.fr*, *shrd.com*, *Clubic.com*, *etsy.com*, *cdiscount.com*, *hugedomains.com*, *fno.fr*¹⁷). Par ailleurs quelques rares sites ont des distributions réellement bimodales. C'est notamment le cas de La France Insoumise (LFI) ou du site *melenchon.fr* dont nous reproduisons les distributions à la Figure 7.

4.3. La distribution politique des médias

L'exemple du blog de Jean-Luc Mélenchon est néanmoins l'exception qui confirme la règle. Dès lors, il semble raisonnable de résumer en première approche la position idéologique d'un domaine du web à la moyenne des positions des comptes ayant partagé ses URLs. C'est ce principe qui motive la réalisation de la cartographie de la Figure 8 représentant la distribution spatiale des 20 000 médias les plus partagés sur Twitter par notre échantillon. La projection des médias dans l'espace idéologique latent montre que, à partir de principes d'organisation semblable à l'espace bi-dimensionnel des députés et sénateurs, les tensions sur les axes de la distance au pouvoir et droite/gauche fournissent un cadre d'interprétation pertinent. Les nouvelles entités que nous avons projeté dans cet espace invitent cependant à spécifier autrement l'axe de la distance au pouvoir et aux institutions. Celui-ci oppose des médias périphériques, fortement politisés et à faible audience aux médias centraux, professionnels et à très forte audience. Le haut et le centre de la carte fait apparaître les grands titres de la Presse quotidienne et régionale, *l'Équipe* ou les principales radios et télévisions comme *France info.tv*. Traduisant l'effet du premier axe marquant la proximité au pouvoir, on observe aussi la place

17. la fondation nationale des orthophonistes

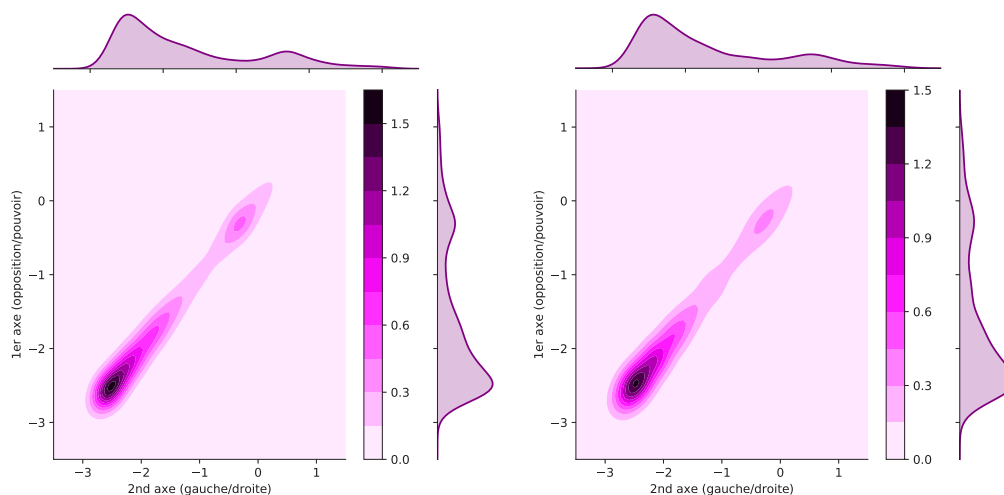


FIGURE 7 – *Distribution idéologique correspondant aux 13 112 actions de partages du site de la France insoumise et les 6 571 partages du blog de Jean-Luc Mélenchon. Les profils sont naturellement semblables mais étonnément bi-modaux (avec un ensemble de points se concentrant d'abord sur la gauche radicale puis sur un espace correspondant à une gauche plus « institutionnelle »).*

de la presse d'information économique comme *Les échos*, *La Tribune* ou *Challenges*. En descendant un peu dans la carte, s'exerce alors l'effet de la deuxième dimension (axe droite-gauche) qui sépare légèrement les titres professionnels et centraux des grands quotidiens ou magazines plutôt orientés au centre-gauche (*Le Monde*, *France Culture*, *Libération*, *Télérama*, *The Guardian*) et ceux qui sont plutôt orientés au centre-droit (*Le Figaro*, *L'opinion*, *Paris-Match*, *Atlantico*). Au bas de la carte, c'est la polarisation politique sur l'axe droite-gauche qui clive en deux camps distincts deux types de médias qui sont tous deux dans une position oppositionnelle beaucoup plus marquée à l'égard du pouvoir et des institutions. Ils ont aussi une audience beaucoup moins importante que les médias centraux. La galaxie constituée dans la partie gauche de la carte, fait apparaître après les médias les plus intégrés au pôle institutionnel comme *Alternatives économiques*, *Médiapart* ou *Reporterre*, des médias plus politisés comme *L'Humanité*, *Révolution permanente* ou *Lundi.am*. Sur l'autre frange oppositionnelle, s'observe le déploiement des médias de droite (*Causeur*, *Valeurs actuelles*) dans la zone intermédiaire jusqu'aux sites d'extrême-droite identitaire (*François de souche*) et national-révolutionnaire. Plongée dans notre espace idéologique latent, le positionnement des médias apparaît ici étonnément conforme à des résultats obtenus dans une autre recherche (Cardon et al., 2019) s'appuyant pourtant sur des principes de regroupements différents puisque fondés sur les liens hypertextes que s'échangent les médias entre eux.

Par ailleurs, la Figure 8 fait aussi apparaître la dispersion des partages des médias sur l'axe droite/gauche (représenté en couleur). Cette mesure de diversité idéologique doit être interprétée avec prudence tant l'espace idéologique est in-homogène. Pourtant, les résultats qu'une telle mesure fait apparaître sont intéressants. Les médias dont les articles partagés sont les moins distribués sont ceux du pôle central dominant plus proche du pouvoir et les médias de droite et de l'extrême-droite. Les articles de ces médias sont partagés sur Twitter par des comptes relativement proches idéologiquement, alors que les grands médias du centre-gauche et, de façon plus singulière, les médias plus distants du pouvoir et plus à gauche voient leurs articles partagés par des comptes idéologiquement plus dispersés. Ces résultats donnent une représentation assez juste de la forme idéologique de l'espace médiatique français et, notamment, des partages sur Twitter dont le barycentre se trouve au centre-gauche (Charon, 2006; Boyadjian, 2014).

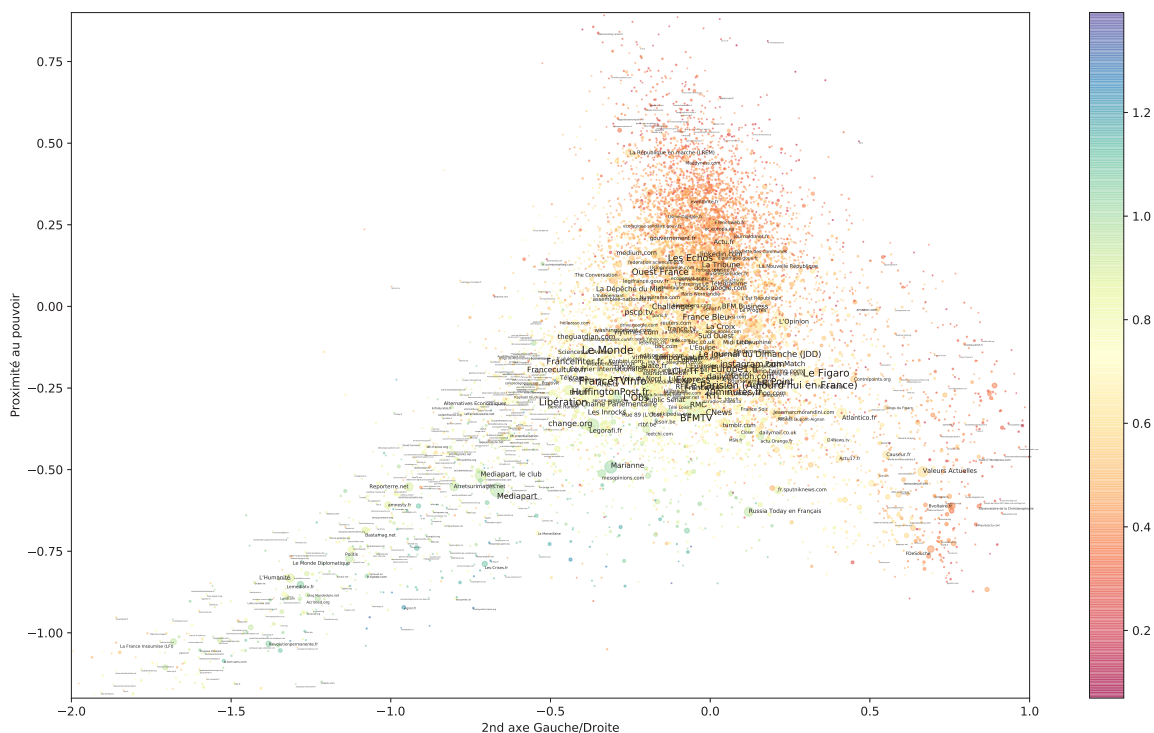


FIGURE 8 – Distribution des 20 000 domaines cités par le plus grand nombre de comptes sur Twitter. Les 100 les plus cités sont étiquetés ainsi qu’une sélection aléatoire de sites à la périphérie du centre de l’espace politique. La taille des domaines est proportionnelle au nombre de citations reçues depuis Twitter. La couleur des domaines correspond à la déviation standard de la position idéologique des sites sur l’axe droite/gauche.

5. Les groupes Gilets jaunes sur Facebook

Au terme de ce long détour destiné à déplacer l’espace idéologique latent vers les Gilets jaunes, la dernière phase de notre méthode consiste à propager les positions idéologiques des URLs extraites dans la partie précédente des comptes Twitter vers les posts qui citent ces URLs sur Facebook, et plus précisément au sein des posts publiés par les groupes Gilets jaunes.

5.1. Collecte de données de groupes sur Facebook

Pour collecter les publications des groupes Gilets jaunes sur Facebook, nous utilisons l’API Crowdtangle de Facebook. Dans une première collection exploratoire, nous avons collecté tous les groupes Facebook qui ont utilisé l’expression « Gilets jaunes », dans l’une de ses variantes (singulier, pluriel, attaché, détaché, précédé ou non d’un hashtag). Cette exploration a permis d’identifier 21 047 groupes. En retenant les 1 000 groupes les plus actifs, nous avons ensuite procédé à leur examen manuel afin d’exclure ceux qui n’étaient pas des groupes militants Gilets jaunes (certains groupes, commentant l’actualité politique mais n’étant pas des groupes sympathisants ont ainsi été exclus). Notre corpus contient ainsi les 892 groupes Facebook Gilets jaunes les plus actifs, réunissant en

cumulé 2 667 810 membres¹⁸ au 7 octobre 2019, date limite de la collecte.

Pour l'ensemble des 892 groupes les plus actifs, nous avons procédé à l'extraction de tous leurs messages, du 1^{er} novembre 2017 au 7 octobre 2019. 9 755 214 publications constituent ainsi notre corpus de départ. La présence de groupes Gilets jaunes avant 2018 s'explique par le fait que certains groupes ont changé de nom au cours de cette année en devenant des groupes Gilets jaunes lorsque le mouvement a émergé ; nous avons décidé de les conserver pour intégrer dans notre analyse l'espace idéologique dans lequel les revendications des Gilets jaunes ont commencé à se diffuser (Figure 9). Le nombre de posts collectés avant le mois de novembre 2018 est néanmoins dérisoire comparé au volume de données générées quotidiennement à compter du 1^{er} acte.

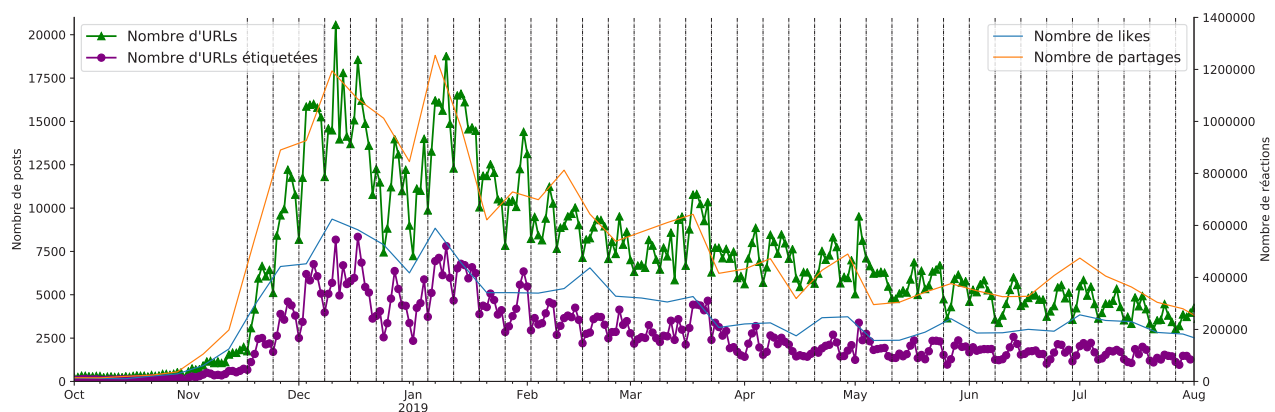


FIGURE 9 – Nombre de posts publiés parmi nos groupes, et nombre de likes et partages suscités

5.2. Les Gilets jaunes dans l'espace politique

Par la suite, nous nous concentrons sur près de 2 millions de posts qui citent ou illustrent leurs propos à l'aide d'une ressource extérieure à Facebook (nous avons précisément dénombré 1 875 462 posts se référant à 2 377 875 URLs dont 456 435 sont uniques¹⁹). Il est intéressant de noter qu'il s'agit dans près d'un quart des cas de vidéos hébergées sur la plateforme YouTube (576 277 posts citant 57 699 URLs différentes).

Compte tenu de notre stratégie de collecte de données (fondée sur un échantillon de 40 000 utilisateurs sur Twitter) et du fait que tous les contenus partagés sur Facebook ne circulent pas nécessairement sur Twitter, nous ne sommes capables d'associer une position idéologique qu'à une portion réduite de ces derniers. Parmi les 1 371 794 d'URLs pour lesquelles nous avons pu définir une étiquette politique calculée comme la moyenne des positions des utilisateurs qui les ont partagées (avec la contrainte qu'ils soient au moins 5), seulement 85 742 d'entre elles sont présentes dans les posts de nos groupes Gilets jaunes. Néanmoins, elles sont citées à 788 104 reprises entre le 1^{er} octobre 2018 et le 1^{er} août 2019, dans le corpus, ce qui représente près d'un tiers de l'ensemble des posts citant une URL (précisément 36.4% avant le 1^{er} septembre 2019). La Figure 9 montre que la proportion de posts contenant des URLs étiquetées par notre méthode est relativement stable tout au long du mouvement. Il est délicat de caractériser le biais éventuel qui résulte de ce déficit de couverture des posts. En tout état de cause, les résultats indiquant que les URLs retrouvées sur Facebook se distinguent par une forte distance aux institutions, nous supposons que le biais systématique que pourraient générer des URLs qui ne circulent pas sur Twitter serait plutôt de nature à amplifier le phénomène observé.

18. Ce chiffre ne correspond pas au nombre de personnes suivant les groupes Gilets jaunes de notre corpus, beaucoup d'entre elles étant membres de plusieurs groupes.

19. La librairie `urld` nous a permis de standardiser et d'homogénéiser autant que possible les URLs de façon à réduire les variantes d'une adresse qui pointent en réalité vers une même page.

5.3. Analyse statique

Pour identifier la « couleur » idéologique des Gilets jaunes, nous proposons de positionner leurs publications dans notre espace idéologique latent. La Figure 10 présente deux distributions côte à côte qui représentent toutes les deux des distributions spatiales d'URLs partagées. Conformément au calcul réalisé pour les médias dans la partie précédente, nous définissons la position d'une URL comme la moyenne des positions idéologiques inférées des comptes Twitter l'ayant partagée. Seules les URLs partagées au moins 5 fois sont conservées. Dès lors, il est possible de mettre en perspective – figure de gauche – l'enveloppe idéologique que compose l'ensemble des URLs citées par notre échantillon aléatoire de 40 000 individus « engagés » sur Twitter depuis l'été 2018, et – figure de droite – la distribution des URLs partagées dans les groupes Facebook des Gilets jaunes de notre corpus.

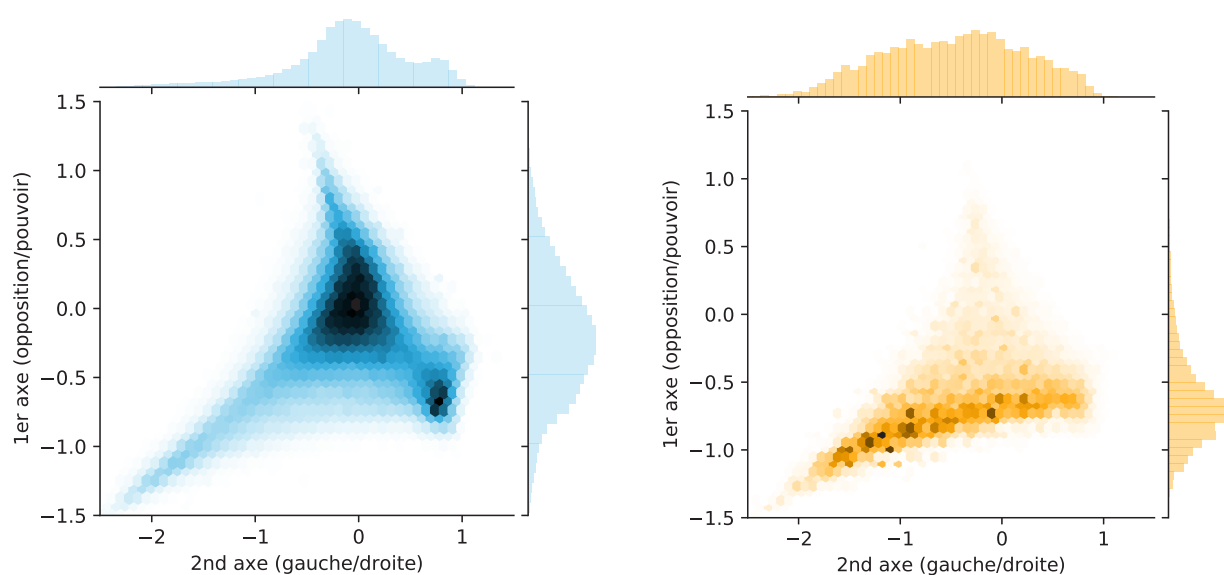


FIGURE 10 – *Distribution de l'espace idéologique occupé [à gauche] par les URLs partagées sur Twitter (40 000 comptes aléatoires), [à droite] les URLs partagées sur les groupes Gilets jaunes sur Facebook*

Le contraste est saisissant. À l'inverse de la distribution standard des liens sur Twitter, l'essentiel de la distribution des Gilets jaunes sur Facebook se situe sur la frontière la plus « oppositionnelle » (distante au pouvoir) de l'espace idéologique latent. Très distant du pôle du pouvoir institutionnel, le positionnement des Gilets jaunes est en revanche très peu marqué sur l'axe droite/gauche, ce qui révèle une distribution politique très hétérogène du mouvement. Au regard des deux dimensions déployées dans cet espace, c'est bien la première dimension, la distance aux institutions, qui constitue la caractéristique dominante et unificatrice des publications des Gilets jaunes et pas la seconde.

Transposée désormais vers le mouvement des Gilets jaunes, le premier axe que nous avons défini dans l'espace parlementaire comme proximité ou distance au pouvoir gouvernemental et aux institutions peut alors ici être considéré comme une bonne approximation de ce qui est entendu dans les approches de sciences politiques sous le nom de « populisme ». Alors que les spécialistes ont longtemps débattu de la définition de cette notion (Rooduijn, 2019), un certain consensus se dégage désormais autour de l'usage non normatif de cette étiquette pour qualifier des formations qui partagent une vision du monde opposant « le peuple pur » et les « élites corrompues » (Canovan, 1999; Laclau, 2005; Taggart, 2004; Mudde et Kaltwasser, 2012; Hawkins et Rovira Kaltwasser, 2017; Van Kessel, 2014; Rooduijn et al., 2019; Müller, 2017; Moffitt, 2020).

Les recherches sur le populisme ont en effet d'abord mis l'accent sur les variantes de droite de ce phénomène au cours des trois dernières décennies (Mudde, 2007; Albertazzi et McDonnell, 2007). Ce n'est que récemment que les formations de la gauche ont pu attirer directement l'attention sur la relation entre le populisme et la politique de gauche en Europe (March, 2007; Stavrakakis et Katsambekis, 2014). À cet égard, les Gilets jaunes qui réunissent des positionnements qui se marquent aussi bien très à droite ou très à gauche s'inscrivent en revanche dans une interprétation du « populisme » qui ne cherche pas à le qualifier sur l'axe droite/gauche – sur lequel les acteurs refusent de se positionner –, mais comme une opposition constante aux institutions et au pouvoir des élites et des représentants. Cette distance oppositionnelle apparaît bien alors comme l'effet d'une construction morale opposant le peuple aux élites (Moffitt, 2020; Mudde et Kaltwasser, 2012) sans pourtant que cette construction manichéenne ne fasse disparaître la variété des positionnements sur l'axe droite/gauche. La tension centrale (l'opposition au pouvoir) n'écrase pas l'hétérogénéité des références politiques des Gilets jaunes qui empruntent leurs points d'appui aux deux extrémités de l'axe gauche-droite. À l'instar de la qualification proposée par Pirro (2018) pour Le Mouvement Cinq Étoiles italien, les Gilets jaunes apparaissent comme un « populisme polyvalent » (*polyvalent populisme*) qui articule, sans nécessairement les mélanger, des traits politiques situés aux deux bords de l'échiquier politique. D'autres travaux conduits à partir d'enquêtes d'opinion ont souligné que la pluralité des adhésions au mouvement des Gilets jaunes qui s'expriment plus volontiers sur les franges gauche et droite des échelles d'opinion, n'effaçaient pas la diversité des idéologies et ne permettaient pas de les réunir dans un ensemble homogène (Algan et al., 2019a). C'est aussi ce qui apparaît dans la distribution des groupes Facebook de Gilets jaunes dans l'espace idéologique latent construit de façon inductive par notre méthode.

5.4. Analyse Dynamique

La structuration politique que laisse apparaître la Figure 10 est très claire, mais pourrait résulter d'un artefact de construction *ex-post* de notre corpus qui tend à aplatir une année de débats et de mobilisation en une seule représentation, fut-elle bi-dimensionnelle ? Il semble clair que les Gilets jaunes se situent sur Facebook dans un espace politique à distance des institutions de la République. Par contre, il est théoriquement possible que le mouvement ait muté suffisamment entre le 17 novembre 2018 et le printemps 2019 pour que la forte dispersion qui nous apparaît soit en réalité la résultante de l'agrégation temporelle de moments successifs de la mobilisation dont le centre de gravité politique aurait migré. L'avantage de notre méthode est précisément qu'elle permet de suivre la distribution de l'espace idéologique qu'occupent les Gilets jaunes en opérant des regroupements temporels variés.

Pour répondre à cette question, nous avons construit (Figure 11) quatre périodes qui nous semblaient correspondre à un découpage possible du mouvement des Gilets jaunes qui distingue une première phase de préparation du mouvement avant le premier Acte, une seconde phase d'intensification de la mobilisation à Paris et en France qui culmine avec les détériorations de l'Arc de Triomphe, une troisième phase qui installe le mouvement dans l'hiver jusqu'au fort regain de mobilisation de l'Acte VIII, et enfin une quatrième et dernière phase qui correspond à essoufflement progressif de la mobilisation. Cette décomposition temporelle permet de répondre à la question initiale, mais la réponse n'est pas univoque. Oui, le mouvement des Gilets jaunes a, dans ses premières semaines, et notamment avant que les premiers Actes ne s'égrainent, fait appel sur Facebook à de nombreuses citations vers des ressources situées très à droite. Durant cette première phase, la distribution est extrêmement concentrée sur cette partie de notre espace latent. Les deux phases suivantes (du 1^{er} Acte jusqu'à l'Acte VIII marqué par la médiatisation de l'ancien boxeur Christophe Dettinger défiant les gendarmes mobiles à poings nus sur la passerelle Léopold-Sédar-Senghor) font clairement apparaître un glissement progressif du mouvement dont le centre de gravité se déplace clairement vers la gauche du spectre politique. Pour autant, et la quatrième période semble confirmer cette hypothèse, l'espace

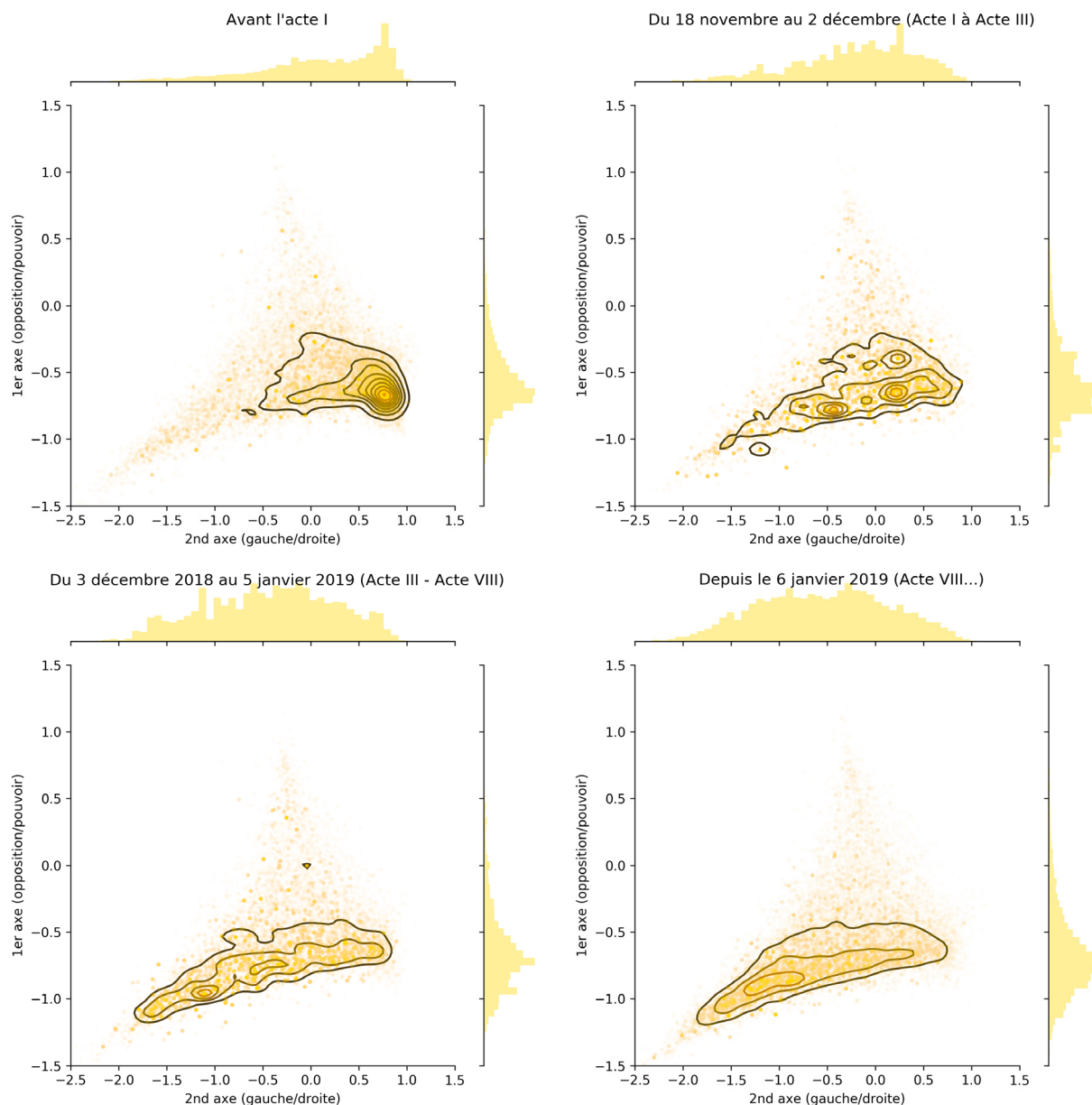


FIGURE 11 – *Distribution idéologique des posts des Gilets jaunes en fonction des URLs citées au cours de quatre périodes successives. Pour chaque période 28 000 URLs aléatoires sont échantillonnées de façon à construire des distributions correspond à un nombre d'entités constant dans chaque période. Les courbes de niveau concentriques successives correspondent également à un niveau de densité constant d'une période à l'autre qui permet d'observer l'affaissement progressif de la distribution.*

occupé par le mouvement a également tendance à lentement s'aplatir sur l'axe droite/gauche. La très forte concentration initiale (sur la droite du spectre idéologique) est remplacée au fil des mois par une distribution beaucoup plus étalée qui illustre assez bien cette notion de populisme polyvalent.

5.5. Géométrie des discours

Plongés dans l'espace idéologique latent, le contenu des discours des Gilets jaunes peut lui aussi être positionné afin de faire apparaître différents registres expressifs et thématiques. Pour se faire,

nous avons procédé en deux étapes. Dans un premier temps nous avons découpé l'espace idéologique occupé par l'ensemble de nos posts en 16 rectangles regroupant un nombre identique de posts²⁰. Puis nous avons isolé les 4 rectangles correspondant aux posts les plus extrêmes selon nos deux dimensions d'analyse. Chaque ensemble regroupe ainsi environ 20 000 publications positionnées dans ces rectangles les plus éloignés du centre de gravité du mouvement. On distingue ainsi les posts les plus proches du pouvoir et les plus à gauche (Gauche-Pouvoir), les plus proches du pouvoir et les plus à droite (Droite-Pouvoir), les plus éloignées du pouvoir et les plus à droite (Droite-Opposition), les plus éloignées du pouvoir et les plus à gauche (Gauche-Opposition). Une extraction terminologique²¹ a permis par la suite d'identifier les termes les plus fréquents²².

Pour la représentation finale, nous n'avons retenu que les termes dont la distribution dans nos 4 ensembles pré-définis est la plus déséquilibrée de façon à faire apparaître dans un anneau les termes les plus spécifiquement associés aux différentes directions de l'espace idéologique. La Figure 12 permet ainsi d'identifier dans chaque cadran les personnalités (nous avons choisi de distinguer les acteurs en les coloriant en rouge), les revendications ou les organisations.

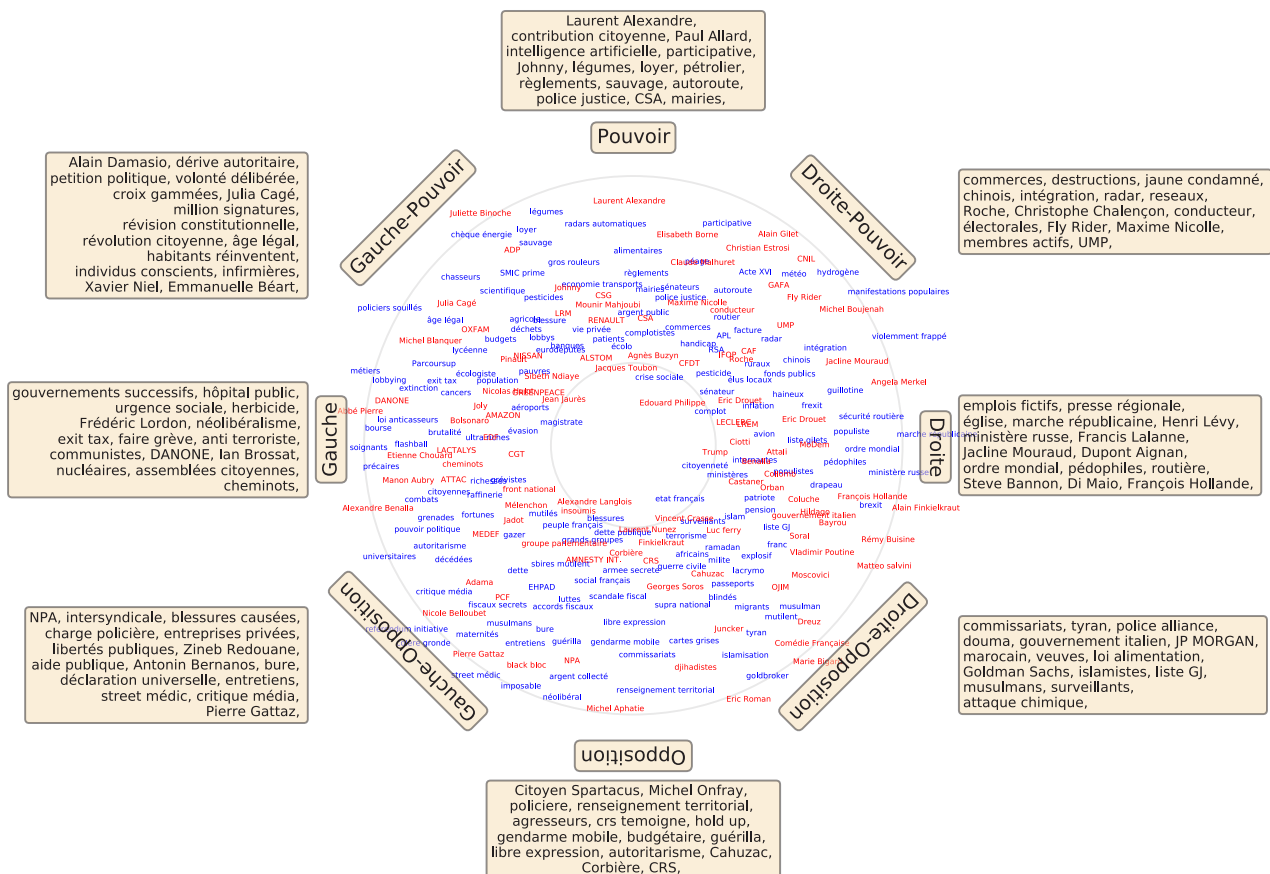


FIGURE 12 – Les mots des Gilets jaunes. Les termes (figurés en rouge lorsqu'ils correspondent à des acteurs) sont positionnés dans un anneau tendu par les dimensions Gauche/Droite, Opposition/Pouvoir en fonction de leur fréquence d'apparition relative dans l'espace idéologique. Nous avons listé dans les boîtes périphériques les termes les plus exclusifs à chaque catégorie.

Plus on se trouve proche du pouvoir et des institutions (i.e. dans la partie haute du graphe), plus les thèmes discutés, tant sur le flanc droit que sur le flanc gauche, sont liés au fonctionnement

20. Nous avons fait appel à une décomposition de l'espace en KD Tree pour réaliser le découpage.

21. Nous nous sommes inspirés de la librairie *Scattertext* de Jason Kessler pour produire cette visualisation.

22. Après élimination des mots vides et nettoyage des scories présentes dans les posts (caractères isolés, problème d'encodages), nous avons identifié près de 5200 termes (incluant les bi-grammes) dont la fréquence d'apparition dans le sous-corpus composé de nos quatre rectangles est supérieure à un seuil arbitraire de 80. Un dernier nettoyage manuel a permis d'éliminer certains doublons (singulier/pluriel, acronymes formes développées).

des institutions et de la démocratie. Les revendications participatives et constitutionnelles, comme le Référendum d'initiative citoyenne (RIC) réclamé par une partie du mouvement des Gilets jaunes apparaissent dans le pôle le plus intégré à l'espace central et institutionnel. C'est dans cette partie haute que les membres du gouvernement, les termes liés à des politiques publiques (argent public, service public, taxe, transports, RSA ou impôts) se trouvent positionnés. En revanche, plus on descend dans le bas du graphe sur la dimension de distance aux institutions, plus les revendications thématiques se différencient entre les pôles situés très à gauche et très à droite. Du côté de la gauche oppositionnelle les revendications se fixent sur les questions de contestation du pouvoir économique (*aide économique, hôpital public, urgence sociale*) dans la partie intermédiaire, mais la thématique des violences policières sature presque totalement la zone la plus oppositionnelle de la gauche de la figure (*blessures, charges policières, libertés publiques, mutiler, Zineb Redouane, etc.*) en laissant une petite place aux acteurs qui interviennent pour soutenir le cadrage d'extrême-gauche des revendications des Gilets jaunes (*NPA, Frédéric Lordon, Mélenchon, Ruffin, etc.*). Sur la partie droite oppositionnelle (en bas à droite de la carte), la question de la police apparaît aussi mais cette fois pour la défendre. Apparaissent aussi les termes de *migrants* et d'*islamisation* qui témoignent de l'empreinte des thématiques de l'extrême-droite. Positionné au sein de l'espace idéologique latent construit dans cette recherche, le lexique des Gilets jaunes sur Facebook fait lui aussi apparaître l'extrême hétérogénéité politique de ce populisme polyvalent.

6. Conclusion : propager un espace idéologique latent

Partant d'une analyse de la structure du réseau social qui lie les parlementaires français à leurs followers sur Twitter, nous sommes parvenus à positionner l'idéologie qui sous-tend les pratiques de partage d'information des Gilets jaunes sur Facebook. En dépit de leur refus d'être réduits à un parti ou une orientation politique, la méthode d'inférence que nous avons déployée permet de positionner les Gilets jaunes sur Facebook selon deux dimensions qui correspondent à un axe droite-gauche traditionnel ainsi qu'un second axe opposant Pouvoir et Contre-Pouvoir. Près de 2 millions de posts publiés sur quasiment un millier des groupes publics Facebook associés au mouvement des Gilets jaunes sont ainsi positionnés dans cet espace idéologique latent. Les posts ne se distribuent pas de façon homogène dans l'espace, ils se regroupent de façon très majoritaire dans un espace opposé au pouvoir (partie basse). Le suivi de leur positionnement au fil des actes ayant scandé le mouvement montre que le mouvement a d'abord été inspiré par des ressources très droitières avant de très vite opérer un élargissement vers la gauche tout en restant fidèle à une attitude contestataire, typique d'un populisme polyvalent.

Mais comment un *espace idéologique latent* construit à partir des *followers* des parlementaires sur Twitter peut-il être un outil pertinent pour positionner des millions de posts de Gilets jaunes sur Facebook ? La qualité des résultats mis en évidence par la technique de propagation utilisée dans cette recherche invite à interroger les raisons de son efficacité. Plongée dans cet espace idéologique latent, la mobilisation des Gilets jaunes se déploie au sein de coordonnées qui permettent de la rendre plus intelligible : des traits déterminants de sa dynamique apparaissent, l'hétérogénéité du mouvement se laisse cartographier et les thèmes des revendications peuvent être associés à des positions politiques. Bien qu'il soit encore difficile d'évaluer correctement les apports et les écueils de ces nouvelles méthodologies, nous voudrions proposer quelques réflexions pour engager la discussion autour des apports des approches *data driven* aux recherches en sciences sociales.

La première caractéristique de cette méthode réside dans son caractère inductif. Elle permet de ne pas assigner les acteurs à une échelle politique qu'ils contestent (auto-positionnement sur l'axe droite-gauche, inférence depuis les votes passés, indifférence aux sans-réponses et aux répondants refusant de se positionner). L'originalité de cette approche est de faire apparaître, virtuellement, ces positionnements comme s'ils émergeaient à *côté des acteurs* en traçant virtuellement la géométrie

d'un espace dans lequel leurs comportements prennent sens relationnellement les uns par rapport aux autres. Le choix des données d'entrée (des parlementaires), du moyen de l'étendre (la plateforme Twitter), du véhicule permettant de propager l'espace (le partage des URLs) n'a rien de neutre et contribue grandement à orienter la construction de cet espace dans un horizon qui est celui de la définition, la plus légitime, de la politique. Mais il n'en demeure pas moins que les dimensions produites par cet espace n'ont pas été définies par les acteurs politiques centraux ou les chercheurs, mais par les comportements de milliers de comptes Twitter. La particularité d'une telle approche est qu'elle manque à l'évidence de précision lorsque les positionnements sont observés localement, à l'échelle d'une URL individuelle peu partagée, mais qu'en revanche elle présente au niveau global une consistance remarquablement stable.

La deuxième caractéristique de cette technique de propagation tient à la souplesse de l'architecture de cet *espace idéologique latent*, instrument de recherche d'un nouveau type dont on ne connaît pas encore bien les potentialités et les défauts. Les dimensions qu'il fait apparaître prennent des colorations différentes selon les entités qu'on y plonge tout en exprimant cependant le même genre de propriétés. L'opposition entre partis de gouvernement et partis d'opposition qui apparaît comme la qualification la plus ajustée du premier axe de cet espace au moment de sa gestation dans les enceintes parlementaires, prend une acception différente lorsque l'on plonge les médias dans ce même espace. L'opposition entre les médias centraux à forte audience et des médias plus radicaux et périphériques constitue alors une autre manière de qualifier la même dimension de cet espace idéologique latent. Appliquée enfin aux comportements des Gilets jaunes, la distribution sur l'axe de la distance au pouvoir apparaît cette fois comme un index de « populisme » suffisamment pertinent pour projeter un éclairage original sur le mouvement. Structure à la fois stable et souple, l'espace idéologique latent constitue ainsi un instrument d'exploration qui peut être utilisé pour lire et interpréter nombre de phénomènes sociaux qui laissent des inscriptions numériques. Cette méthode inductive est ainsi particulièrement adaptée pour caractériser des acteurs collectifs spontanés comme les Gilets jaunes sans attaches formelles à des partis ou syndicats.

La dernière caractéristique, enfin, a trait à l'explication de la « stabilité souple » qu'enferme cette méthode. Deux réponses viennent à l'esprit. La première est que les URLs partagées sur les réseaux sociaux que, par simplification, on a appelés ici l'espace médiatique, constituent une structure suffisamment stable, et fréquemment mobilisée, pour propager à longue distance, vers des entités *a priori* très éloignées, des positions idéologiques qui sont nées dans le suivi par certains internautes des comptes Twitter des députés ou sénateurs. Parce que le positionnement des sites web des médias les uns par rapport aux autres compose une topographie à la structure relativement stable, elle constitue un intermédiaire pertinent pour faire migrer les positions idéologiques vers toutes sortes d'acteurs numériques qui partagent des URLs vers des sites d'informations. La seconde explication, étroitement liée à la première, est le caractère réglé et organisé des comportements des internautes. La propagation de la structure de cet espace tient à une certaine forme de rationalité pratique des attitudes des internautes : dans la composition de leurs abonnements sur Twitter et dans les pratiques de partage de liens hypertextes vers des médias auxquelles ils se livrent, ils rassemblent, rapprochent, opposent les informations de façon beaucoup plus diverse et variée que s'ils avaient à régler leurs pratiques selon une partition de valeurs idéologiques stables et contraignantes. Cependant, l'effet composé d'un très grand nombre de partage d'information par les internautes, laisse apparaître un comportement suffisamment réglé et ordonné. L'originalité d'une perspective guidée par les données numériques tient à la possibilité d'extraire d'une masse considérable de micro-comportements numériques des informations qui font émerger un ordre suffisamment cohérent pour offrir un cadre d'interprétation déplaçable d'un espace social vers un autre. Cette démarche parvient ainsi à fournir un cadre d'analyse abstrait et distant qui est produit, sans intention explicite, par le caractère ordonné des comportements individuels lorsque ceux-ci sont mesurés à grande échelle.

Remerciements

Nous remercions Justin Clark pour son aide dans la constitution du réseau social des figures politiques et pour ses conseils précieux (ainsi que ceux de Jonas Kaiser) pour l'implémentation de la méthode d'inférence des positions politiques des acteurs. Nous remercions également Max Halford qui a développé le module d'analyse de correspondance Prince que nous utilisons.

Ce travail a bénéficié du soutien financier de trois projets : le *SAB YellowPol* de Sciences Po et le projet "I read it on Facebook" du programme Social Science One du SSRC (*Social Science Research Council*), et enfin le projet JCJC GOPI (ANR19-CE38-0006-01).

En conformité avec le Règlement général à la protection des données 2016/679, le projet de recherche *Ideology scaling Twitter France* dont on exploite les données a été déclaré le 19 mars 2020 au registre des traitements de données de Sciences Po.

Références

- Albertazzi, D. et D. McDonnell (2007), *Twenty-first century populism : The spectre of Western European democracy*, Springer.
- Algan, Y., E. Beasley, D. Cohen, et M. Foucault (2019a), *Les Origines du populisme*, Le Seuil.
- Algan, Y., E. Beasley, D. Cohen, M. Foucault, et M. Péron (2019b), «Qui sont les gilets jaunes et leurs soutiens», Rapport technique, Technical report, CEPREMAP et CEVIPOF.
- André, A., C. Bedock, Z. Bendali, A. B. de Raymond, A.-G. Beurier, P. Blavier, L. Bonin, A. Courty, M. Della Sudda, G. Delozière, et al. (2019), «Enquêter in situ par questionnaire sur une mobilisation en cours : une étude sur les gilets jaunes», *Revue Française de Science Politique*, vol. 69, n° 5-6, pp. 869–892.
- Barberá, P. (2015), «Birds of the same feather tweet together : Bayesian ideal point estimation using twitter data», *Political analysis*, vol. 23, n° 1, pp. 76–91.
- Barberá, P., J. T. Jost, J. Nagler, J. A. Tucker, et R. Bonneau (2015), «Tweeting from left to right : Is online political communication more than an echo chamber?», *Psychological science*, vol. 26, n° 10, pp. 1531–1542.
- Barberá, P. et G. Rivero (2015), «Understanding the political representativeness of twitter users», *Social Science Computer Review*, vol. 33, n° 6, pp. 712–729.
- Bendali, Z., R. Challier, M. Della Sudda, O. Fillieule, É. Agrikoliansky, et P. Aldrin (2019), «Le mouvement des gilets jaunes : un apprentissage en pratique (s) de la politique?», *Politix*, vol. 128, n° 4, pp. 143–177.
- Bond, R. et S. Messing (2015), «Quantifying social media's political space : Estimating ideology from publicly revealed preferences on facebook», *American Political Science Review*, vol. 109, n° 1, pp. 62–78.
- Bonica, A. (2014), «The punctuated origins of senate polarization», *Legislative Studies Quarterly*, vol. 39, n° 1, pp. 5–26.
- Boyadjian, J. (2014), «Twitter, un nouveau «baromètre de l'opinion publique»?», *Participations*, vol. 8, n° 1, pp. 55–74.

- Boyer, P. C., T. Delemotte, G. Gauthier, V. Rollet, et B. Schmutz (2020), «Les déterminants de la mobilisation des gilets jaunes», *Revue économique*, vol. 71, n° 1, pp. 109–138.
- Briatte, F. et E. Gallic (2015), «Recovering the french party space from twitter data», Preprint, Presented at the "Science Po Quanti" conference, Sciences Po, Paris, 2015.
- Canovan, M. (1999), «Trust the people ! populism and the two faces of democracy», *Political studies*, vol. 47, n° 1, pp. 2–16.
- Cardon, D., J.-P. Cointet, B. Ooghe, et G. Plique (2019), «Unfolding the multi-layered structure of the french mediascape», Preprint.
- Charon, J.-M. (2006), «Les journalistes politiques : qui sont-ils ?», *Le Temps des médias*, vol. 7, n° 2, pp. 176–190.
- Clinton, J., S. Jackman, et D. Rivers (2004), «The statistical analysis of roll call data», *American Political Science Review*, vol. 98, n° 2, pp. 355–370.
- Confavreux, J. (2019), *Le fond de l'air est jaune. Comprendre une révolte inédite*, Seuil.
- Della Porta, D., J. Fernández, H. Kouki, et L. Mosca (2017), *Movement parties against austerity*, John Wiley & Sons.
- Ecormier-Nocca, F. et N. Sauger (2020), *Measuring Elite-Mass Linkages through the Internet*, Research Handbook on Political Partisanship edited by Henrik Oscarsson, Sören Holmberg.
- Gentzkow, M., J. M. Shapiro, et M. Taddy (2019), «Measuring group differences in high-dimensional choices : Method and application to congressional speech», *Econometrica*, vol. 87, n° 4, pp. 1307–1340.
- Groseclose, T. et J. Milyo (2005), «A measure of media bias», *The Quarterly Journal of Economics*, vol. 120, n° 4, pp. 1191–1237.
- Hawkins, K. A. et C. Rovira Kaltwasser (2017), «What the (ideational) study of populism can teach us, and what it can't», *Swiss Political Science Review*, vol. 23, n° 4, pp. 526–542.
- Hix, S., A. Noury, et G. Roland (2006), «Dimensions of politics in the european parliament», *American Journal of Political Science*, vol. 50, n° 2, pp. 494–520.
- Hooghe, L., R. Bakker, A. Brigevid, C. De Vries, E. Edwards, G. Marks, J. Rovny, M. Steenbergen, et M. Vachudova (2010), «Reliability and validity of measuring party positions : The chapel hill expert surveys of 2002 and 2006», *European Journal of Political Research*, vol. 49, n° 5, pp. 687–703.
- Imai, K., J. Lo, et J. Olmsted (2016), «Fast estimation of ideal points with massive data», *American Political Science Review*, vol. 110, n° 4, pp. 631–656.
- Jackman, S., A. Tahk, A. Zeileis, C. Maimone, J. Fearon, Z. Meers, M. S. Jackman, et M. Imports (2017), «Package 'pscl'», See <http://github.com/atahk/pscl>.
- Jensen, J., S. Naidu, E. Kaplan, L. Wilse-Samson, D. Gergen, M. Zuckerman, et A. Spiraling (2012), «Political polarization and the dynamics of political language : Evidence from 130 years of partisan speech», *Brookings Papers on Economic Activity*, pp. 1–81.
- Laclau, E. (2005), *Populism : What's in a Name*, Verso London.
- March, L. (2007), «From vanguard of the proletariat to vox populi : Left-populism as a shadow of contemporary socialism», *SAIS Review of International Affairs*, vol. 27, n° 1, pp. 63–77.

- Moffitt, B. (2020), *Populism*, John Wiley & Sons.
- Mudde, C. (2007), *Populist radical right parties in Europe*, Cambridge University Press.
- Mudde, C. et C. R. Kaltwasser (2012), *Populism in Europe and the Americas : Threat or corrective for democracy ?*, Cambridge University Press.
- Müller, J.-W. (2017), *What is populism ?*, Penguin UK.
- Noiriel, G. (2019), *Les gilets jaunes à la lumière de l'histoire*, Editions de l'Aube.
- Pirro, A. L. (2018), «The polyvalent populism of the 5 star movement», *Journal of Contemporary European Studies*, vol. 26, n° 4, pp. 443–458.
- Poole, K. T., J. B. Lewis, J. Lo, et R. Carroll (2008), «Scaling roll call votes with w-nominate in r», *Available at SSRN 1276082*.
- Poole, K. T. et H. Rosenthal (1984), «The polarization of american politics», *The journal of politics*, vol. 46, n° 4, pp. 1061–1079.
- Poole, K. T. et H. Rosenthal (1991), «Patterns of congressional voting», *American journal of political science*, vol. 35, n° 1, pp. 228–278.
- Rheault, L. et C. Cochrane (2020), «Word embeddings for the analysis of ideological placement in parliamentary corpora», *Political Analysis*, vol. 28, n° 1, pp. 112–133.
- Rooduijn, M. (2019), «State of the field : How to study populism and adjacent topics ? a plea for both more and less focus», *European Journal of Political Research*, vol. 58, n° 1, pp. 362–372.
- Rooduijn, M., S. Van Kessel, C. Froio, A. Pirro, S. De Lange, D. Halikiopoulou, P. Lewis, C. Mudde, et P. Taggart (2019), «The populist : An overview of populist, far right, far left and eurosceptic parties in europe», Preprint.
- Siffer, A., P.-A. Fouque, A. Termier, et C. Largouët (2018), «Are your data gathered ?», in «Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining», pp. 2210–2218.
- Stavrakakis, Y. et G. Katsambekis (2014), «Left-wing populism in the european periphery : the case of syriza», *Journal of political ideologies*, vol. 19, n° 2, pp. 119–142.
- Suignard, P., C. Escoffier, L. Charaudeau, et M. Brugidou (2021), «Que peuvent les algorithmes de plongement de mot pour l'analyse sociologique des textes ? Analyser les discours et caractériser les locuteurs des plateformes « Grand Débat National » et « Vrai Débat »», *Statistique et Société*, vol. 9, n° 1.
- Taggart, P. (2004), «Populism and representative politics in contemporary europe», *Journal of political ideologies*, vol. 9, n° 3, pp. 269–288.
- Tarragoni, F. et J.-P. Gaudillière (2019), «Chercher une définition sociologique de l'élite et du peuple dans les nouveaux mouvements sociaux revient à trahir le populisme qui leur est commun», *Mouvements*, vol. 100, n° 4, pp. 43–54.
- Tristan Guerra, C. A., Frédéric Gonthier (2019), «Populist attitudes among the french yellow vests», *Populism*, vol. 1, n° 1, pp. 1–12.
- Van Kessel, S. (2014), «The populist cat-dog : Applying the concept of populism to contemporary european party systems», *Journal of Political Ideologies*, vol. 19, n° 1, pp. 99–118.

Volkens, A., J. Bara, et I. Budge (2009), «Data quality in content analysis. the case of the comparative manifestos project», *Historical Social Research/Historische Sozialforschung*, vol. 34, n° 1, pp. 234–251.

Annexes

A. Liste des comptes de politiques

_DidierParis, _richardramos, 3807Limon, A_Ardisson, AANZIANI, AbadieCaroline, ac_loisier, ACanayer, ACazabonne, AChristine_Lang, adrienmorenas, AFBrunet_4403, AGenetet, agnesfirmin, ainakuric, AJoyandet, AKaram973, al_petel, Alain_Bruneel, AlainDavid_AN, AlainFouche, alainhoupert, alainneri, alainramadier, AlainTourret, AlainVasselle, AlbaneGaillot, ALCattelot, aldemontgolfier, alexIholroyd, alexiscorbiere, AliceThourot, AlineArchimbaud, AllizardPascal, AMBERTRAND, amelia_lkf, andre_reichardt, AndreChassaing, AndreGattolin, Andretrillard, AngelePreville, Anissa_Khedher, AnnaigLeMeur_AN, anne_chain, AnneBlanc_12, AnneBrugnera, anneEmeryDumas, AnnickBillon, anniedavid38, AnnieGenevard, anthonycellier, antoineherth, AntoineLAON, Apourceau, AQuatennens, arnaud_bazinVO, arnaudviala, AubeyF, AudeAmadou, AudeBono, AudeLuquet2017, audreydufeu, AurelienPradie, Aurelientache, auroreberge, AVFreschi, AVilliers2017, avyelimas, b_abba, b_belhaddad, B_BessotBallot, b_perrut, B_Peyrol, baillydom, Bannier_G, barbarapompili, basphilippe, BBrocard, BCouillard33, BDeflesselles, beatricepiron, BeaudouinSo, beauvais2017, BechtGroff2017, Benedict-Taurine, BenjaminDirx, benoitpotterie, BenoitSimian, bernard_buis, bernardcazeau, bernarddelcros, BernardJomier, bernardreynes, BertrandBouyx, BertrandPancher, BertrandSorre, bessonjean, BGriveaux, BHam mouche5708, BJoncour, bmicouveau, BNAzerot, BockelJeanMarie, bonnefoy_n, BonnardE, BorisVallaud, BouchetGilbert, bourquinmartial, BoyerJ_M, BQuestel, bretonxavier, BrigBourguignon, Brigitte_Kuster, brigitte_liso, bru_vincent, bruno_fuchs, bruno_sido, BrunoBilde, BrunoBonnellOff, BrunoDuverge, brunogilles13, BrunoMillienne, brunoretailleau, brunostuder67, BuonTAN, C_deCourson, C_Frassa, C_MorinDesailly, c_naegelen, C_Troendle, CalvetFrancois, Capo_Canellas, CaroleBBonnard, carolinecayeux, CarolineFiat54, CarolineJanvier, Castellani_, catherine_dumas, CatherineTasca, CBE16, CBonfantiDossat, CConconne, cderoche_senat, Cdganay, CdLavergne, Cecile_Rilhac, CecileCukierman, ceciluntermaier, CedricRoussel06, celine_brulin, celinecalvez, cfabreAN, CGrandjean54, Ch_Hennion, Chalumeau_P, chantal_jouanno, chapelierannie, ChassaingPh, ChauvinMcm, ChCambon, chevenement, ChPrunaud, Chris_Bouillon, ChrisCloarec, christianfavier, ChristianHutin, christoblanchet, christophe_p, christophearend, ChristopheBechu, ChristopheEuzet, Cisaacsibille, cjerretie, CKamowski3805, clairepitollat, clapotmireille, claude_kern, ClaudeGoasguen, ClaudeMalhuret, claudinelepage, Clem_Autain, CMUSCHOTTI, CoDubost, colettemelot, ConstanceLe-Grip, CORDIERPierre08, CorinneBouchoux, CorinneFeret, corinnevignon, CornanoJacques, COURTOIS_JP, cperrin90, cpiresbeaune, CProcaccia, CyrilPellevat, damienabad, damienadam76, Damien-Pichereau, damienregnard, DanielChasseing, DanielDubois80, Daniele_herin, DanielPercheron, danydegert, DanyWattebled, dassouline, DAUBRESSE_MP, david_lorion, david_rachline, DavidDhabib, DBagarry, dbaichere, dbrulebois, DDasilva_LREM, delaverpilliere, delphbataille, delphinebatho, DelponMichelEM, demoulinem, denis_Masseglia, DeniseSaintPe, DenisSommer, DeputeCabareP, DeputeCausse, Deputee_Obono, DeputeRatenon, desarnez, DFasquelle, DiardEric, Didier_Marie76, didier_rambaud, didierlegac, DidierMandelli, didierquentin, DimitriHoub17, DinoCinieri, Dipompeo-Christ, Djebbari_JB, dmartindijon, Domdelegge, dominiq_estrosi, DominiqueDavid_, DominiqueGillot, doorjean, DoStephanie_77, draoulsenateur, DuboisMarianne, DubyMuller, dumasfrederique, dupontaignan, E_Doineau, e_renaud_G, eanthoine26, ebareigts, ebothorel, ECiotti, EdouardCourtial, egold_senat, eguerelLREM, ejeansannetas, ElianeAssassi, ElianeGiraud, EliseFajgeles, elodiejala, elsafaucillon,

EmilieCariou, EmilieCChalas, emmanuel_maquet, EmmanuelCapus, EPoulliat, erica-lauzet, ericbocquet, ericcoquerel, EricKerrouche, EricPAUGET1, ericstraumann, ericwoerth, erwanbalanant, estherbenbassa, etoututpicard, eyonnetSenat, F_BACHELIER, F_Bonhomme, f_cartron, F_Charvier, F_Lardet, F_Menonville, f_patriat, f_personne, fabien_gay, Fabien_Rssl, fabiennecolboc, FabriceBrun, famatras, faureolivier, fchouat, FCommeinhes, FCornutGentille, fdescrozaille, fdumasdeputee, FEspagnac, FGouttefarde, fgranjus, fionalazaar, FJolivet36, florent_boudie, flprovendier, fm_lambert, FMARC_Senat, fmeunier19, FMsenateur59, fnb_officiel, FontainDomeizel, FournierSenat, fpetitan, fpupponi, FranceLorho, franckmarlin, FranckMontauge, Francois_Ruffin, FrancoisAndre35, francoisbaroin, FrancoiseGatel, Fred_Barbier, FVerkamer, fzocchetto, g_gouffiercha, G_Peltier, G_ROUILLARD, g_vuilletet, GabrielServille, GaelLeBohec, GarciaLaxou, Gauvain_Raphael, GBessonMoreau, gchevrollier, Genestjacques1, genevievejean84, genevievelevy83, GeorgesPatient, gerard_larcher, GerardCherpion, gerardcollomb, gerardcornu, gerardlonguet55, GerbaudFrdriqu1, ggorce, GilbertCollard, gilbertroger_93, GillesCarrez, GillesLeGendre, gilleslurton, GirardinEric, GiseleBiemouret, GiseleJourda, GLabazee_senat, GLarrive, gonthier_maurin, GrazieMelchior, GrosdidierSenat, GuillaumeArnell, GuillaumeChiche, guillaumegarot, guillaumekasba, GuillaumGontard, GuyDoKENNEL, guyteissier, HConwayMouret, HdeRaincourt, HeleneZannier, HenrietPierre, Herve-Berville, hervemarseille, HervePellois, HerviauxO, hleroymandelieu, HMaurey, hsaulignac, HSaury, hub_laferriere, hubertfalco, HuguesPortelli, huguesrenson, HuguetteBello, HuguetteLREM46, hvchristophe, iacovellixavier, IanBoucard, IDebre, iflorennes92, imbertycorinne17, imullerquoy, isabelle_rauch, isavalentin43, J_C_Boulard, J_M_Jacques, j_paulfournier, j_pVigier, jacky_deromedi, jacquinedubois, Jacques_GILLOT, jacquesbigot, JacquesBompard, JacquesLeNay, JacquesSavatier, Jacquin_Olivier, JAviragnet, JBMagner, JC_Carle, JCattin2017, JBOUCHET, JCColasRoy, jcgaudin, JCGrelier, JCLarsonneur, jctaugourdeau, jctissot42, Jdetemmerman, jean_bizet, Jean_LucFUGIT, JeanArthuis, Jean-ClaudeLUCHE, jeandesessard, jeanlassalle, jeanlouistouren, jeanluclagleize, JeanMarieFIEVET, Jean-MarieJansse, JeanMichelMIS, JeanPierreGrand, JeanPierrePont, JeanSOL__, JeanTerlier, Jerome-Bascher, jeromebignon, Jeromedurain, jeromelambert16, jeromenury, JF_Acquaviva, jf_husson, jf_longeot, JFCesarini, JFEliaou, JFMBAYE, jfportarrieu, JGrosperin, jimmyphun, JJ_Gaultier, jjbridey94260, jferraraFB, JKrabal, JL_Reitzer, JLabbeSenat, jlbricout, jldupont14, JLeclabart, jffichet, JLLagourgue, jlmelenchon, jlthieriot, JM_Todeschini, JMaireofficiel, JMaquetDeputee, JMarilossian, jmclement86, JMGabouty, JMHoullegatte, jmsermier, jnzulesi, jnbarrot, jncardoux, jnguerini, Joaquim_PUEYO, JocelyneGuidez, joel_guerriau, Joelbigot49, JoelGiraud05, JolyPatrice, JosianeCostes, josypoueyto, jp_bansard, jp_caffet, JP_Dufregne, jp_mattei, JP_Prince_41, JP_Sueur, JPBosino, jpdecool, jpleleux, jpmasseret, JPPlancade, jpraffarin, jrcazeneuve, JRHugonet, juanico, JulienAubert84, Julien-Bargeton, JulienBorowczyk, juliendive, JVanlerenberghe, JVPlace, Kamardine_M, KarineClaireaux, Kerlogot22, KhatTabiF, L_Harribey, L_Lafon, L_poniatowski, L_saintpaul, LabaronneDaniel, Laborde_Senat, lachaudb, LaetitiaAvia, LalandeBernard, LamureE, LanaTetuanui, laqhila, LaugierMichel, lauredlr, LaurenceCohen94, laurencegayte, LaurenceVanceu, lauriannerossi, laurossignol, LauStmartin, lauzzanamichel, LavardeC, LDarcos, LDumontDeputee, lecocqcharlotte, LeguilleBalloy, LeilaAichi, LenaickADAM, LeVigoureuxF, LevrierMartin, LilianeTanguyAN, lise_magnier, LMaillart, loicdombrevail, loichervepublic, LoicKervran18, louis_aliot, LouisNegreCSM, luccarvounas, LudovicMDS, ludovicpajot, Lvichnievsky, m_blonadin, m_bouvard, M_MeynierM, M_Orphelin, MainaSage, manuelvalls, Marc-Delatte, marclefur, MarieAngeMagne, marielebec78, marietamarelle, marinebrenier, marion_lenne, MARSAUDSandra, MartadeCidrac, MartialSaddier, martine_filleul, MartineWonner, massonjeanlouis, mathdarnaud, mathildepanot, MaudPetit_AN94, MauriceANTISTE, MauriceLeroy, mauricevincent_, max_brisson, MaximeMinot, MaxMathiasin, mberthet2014, mcblandin, MDemessine, MDeprezAudebert, MdeVaucouleurs, menard2017, MESNIERThomas, meuniermichelle, Meyer_Habib, Mezard-Jacques, mfperoldumont87, mguevenoux, mhdesesgaulx, michel_berson, Michel_Larive, michel_raison, michelamiel, michelcanevet, MichelDelebarre, michelecrouzet, MichelePeyron, michelfanget, michelforissier, michellegreaume, MichelMagras, michelsavin38, MichelVASPART, MichelVialay, MichelZumkeller,

mickaelnogal, MIRALLESMP, MIZZONJeanMari1, MKaramanli72, MLP_officiel, mmerciersenat, mnlienemann, Moetai1, MonicaMichel_Of, Moniquelborra, moreaujb23, MORELPIERRE, MorhetRichaud, Mouiller_P, mounir, MounirBelhamiti, mressiguiet, MRobert_11, MTabarot, mustapha_laabid, n_delattre, N_Forissier, NadegeLefebvre_, nadiaessayan, NadiaHai78, NadiaRamassamy, Nadine_Grelet, NaimaMoutchou, NassimahDindar, NatachaBouchart, NathalieBassire, nbricq, nicole_sanquer, NicoleDubre17, NicoleDuranton, NicoleLePeih, NicoleTrisse, noel_sylviane, npouzyreff78, NRauscent89, ODassault, odelphine, OHenno, OlgaGivernet, OliveGaillard30, oliviagregoire, olivier_serva, Olivier-Cadic, olivierdam1, OlivierFalorni, oliviermarleix, olivierpaccaud, olivierveran, Ossoncatherine, OuzoulisP, p_a_raphan, p_abate, p_chaize, P_Kaltenbach, p_medevielle, PA_Anglade, pacolombani, PacomeRupin, PascalBoisLaREM, pascalbrindeau, pascaleBORIES, pascaleboyer05, PascaleGruny02, patrice_anato, PatriceVerchere, patrick_masclat, patrickhettel, PatrickKanner, PatrickMignola, PatrickVignal, Pau_Langevin, paul_molac, PaulaForteza, PBertaGard, pdharreville, pereaalain, perrinegoulet, PGallerneau, Ph_Adnot, phbolo, PhDominati, phdumont, phgosselin, Philippe_Marini, philippedallier, philippefolliot, PhilippeGomesCE, PhilippeHuppe, PhilippeMichelK, PhilippePemezec, philmadrelle, pierr_person, pierre_jackie, pierrecharon, PierreCuypers77, pietraszewski_I, PJarlier, platombe, plaurent_pcf, poletti_b, potierdominique, PrudhommeLoic, PSavoldelli94, pybournazel, rachelmazurse1, RachidTemal, RaconCathy, RaphaelGerard17, RAPINJF, RAYMOND_COUDERC, RaymondVall, RemiFeraud, remypointereau, renepaulsavary, RichardFerrand, RichardLioger, RixainMP, Rkaroutchi, RKokouendoJ, RobertHueOff, robinreda, rogermadec, Roland_Ries, RolandCourteau, rolandescure, RomainGrau, Romeiro1L, ronan_legleut, RonanDantec, roserenxavier, rouxjy, RRebeyrotte, RSCactu, s_mazars, S_Trompille, S_Vermeillet, SabineRubinFI93, SabineThillaye, Sacha-Houlie, saidahamada, SamiaGhali, SandrineLeFeur, SandrineMorch, sarahelhairy, sczebonne, SCazenove, Schillinger2, sclaireaux, Sebastien_Nadot, SebastienHuyghe, sebastienjumel, sebchenu, SebLeclerc14, SegouinV, Sempastoushp, senateur61, senateurcabanel, SenateurDTheoph, senateurJGM, SenateurLozach, SenateurMilon, SenateurRYung, SenatriceADK, SenatriceJasmin, SenatriceThomas, sereineenmarche, SergeLetchimy, severine_gipson, SGoyChavent, sirasylla76, SKerbarh, SLeFoll, SmeurantL, solere92, Sonia_Krimi, SoniadlpCaen, sonjoachim, SoPanonacle, sophie_auconie, SophieErrante, SophieJOISSAINS, SophieMette1, sophieprimas, STaillePolian, Stan-Guerini, stelladupont, Stephane_Ravier, stephane1peu, StephaneArtano, StephaneBuchou, stephanie_rist, StephanieAtger, StephPiednoir, StephRiocreux37, StTRAVERT, sutour, Sy_robert, sylvainmaillard, SylvainWaserman, SylviaPinel, SylvieCharriere, SylvieTolmont, tgassilloud, thibault_bazin, ThierryBenoit35, ThierryMichels, ThillAgnes, trastour2017, trobert974, trudigoz, TuffnellLREM17, TurquoisNicolas, Typhanie_Degois, ugobernalicis, V_Delahaye, V_Faure_Muntian, V_Hammerer, V_Riotton, valerie_oppelt, Valerie_Rabault, valerieboyer13, valerieletard, ValeriePetit_EM, VALLINIAndre, vbazinmalgras, verdierjouclas, verienparent, VeroLouwagie, VGB83, VGUILLOTIN, VictorinLurel, Vidal7602, VigierPhilippe, VillaniCedric, VincentDescoeur, VincentEBLE, VincentLedoux59, VincentThiebaut, ViryStephane, VLacroute, VRolland73, Vthomas_63, warsmann, WatrinDominique, XavierBatut, XavierPalu2017, YaelBRAUNPIVET, ydaniel_depute, YDaudigny, ydecourson, yfavenec, YHaury44, YVaugrenard, yves_blein, YvesBouloux, YvesDetraigne, yvesjego, YvesPDB, yvesrome, zivkapark.

B. Domaines les plus cités par les 892 groupes Facebook

Domaines	Nombre de mentions	Nombre de groupes	Nombre d'URLs uniques
youtube video	474390	887	100533
FranceTVInfo	81518	861	13634
Russia Today en Français	53226	811	5020
Le Parisien (Aujourd'hui en France)	52203	831	10406
Le Monde	38889	808	4962
Le Figaro	37017	814	6309
20minutes.fr	35272	810	6452
BFMTV	30664	804	5423
Revolutionpermanente.fr	28797	780	2044
Libération	28010	769	3399
HuffingtonPost.fr	27553	770	5005
www.mesopinions.com	26768	809	2614
Mediapart, le club	26596	762	2331
Mediapart	26271	725	2378
fr.Sputniknews.com	24362	695	4748
France Bleu	24076	787	4946
Ouest France	23570	765	5293
L'Obs	21147	762	2773
www.change.org	17899	770	3371
Actu.fr	17065	707	4540
Marianne	16209	692	1431
L'Express	14897	686	2910
L'Express	14897	686	2910
Europe1.fr	14335	704	2701
Le Point	14059	691	2717
RTL	13923	707	2517
Gj-Magazine.com	13281	595	2144
Sud Ouest	12369	634	2828
Planetes360.fr	11552	623	1803
Le Dauphiné	11013	624	2523
Capital	10906	656	1390
La Voix du Nord	10734	606	3242
www.leetchi.com	9952	679	2015
La Dépêche du Midi	9755	610	2339
Reporterre.net	9563	638	813
La Provence	8997	566	2829
actu.Orange.fr	8892	560	3183
MSN fr	8308	494	3482
FDeSouche	7145	261	4455
Midi Libre	6885	543	1622
Actu17.fr	6823	591	1138
YétiBlog.org	6803	592	479
L'Humanité	6632	539	1010
Bastamag.net	6611	550	536
chng.it	6396	572	2436
Union populaire républicaine (UPR)	6278	460	659
Franceinter.fr	5969	592	1103
vimeo.com	5969	592	720
www.dailymotion.com	5931	623	1404
www.article3.fr	5885	551	60
Resistance Républicaine	5596	444	1761
Le Progrès	5430	458	1595
Valeurs Actuelles	5352	500	1422
Public Sénat	5071	536	527
CNews	4869	539	1385
RMC	4792	582	1095
Le Journal du Dimanche (JDD)	4697	538	881
Bvoltage.fr	4572	409	1713
Lundi.am	4262	508	301
Les Echos	4186	500	1126
Ripostelaique.com	4103	401	1259
Vincent Lapierre (LeMediaPourTous.fr)	4027	529	177
Lesmoutonsrebelles.com	4008	485	560
Franceculture.fr	3997	465	725
L'Est Républicain	3994	404	1279
Var-Matin	3983	390	910
Challenges	3959	478	884
France Soir	3899	523	383