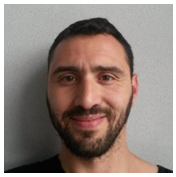
# Les multiples agendas médiatiques des Gilets jaunes sur YouTube

## Exploration d'un corpus de vidéos avec les *topics models*



Bilel BENBOUZID[1]

Hervé GUÉRINS[2]

**TITLE**

The multiple media agendas of the Yellow Vests on YouTube – Exploration of a corpus of videos with the topic models

**RÉSUMÉ**

Cet article vise à rendre compte des résultats d'une analyse quantitative de contenu autour du traitement médiatique des Gilets jaunes, sur un corpus de sous-titres de vidéos YouTube. Pour en assurer l'exploration, la méthode des *topics models* a été mobilisée. Cet article montre que les sous-titres des vidéos de YouTube forment un matériau sur lequel on peut faire enquête pour analyser l'espace médiatique. Si on observe une surreprésentation de la violence contestataire et spectaculaire dans les médias traditionnels, les chaînes de vulgarisation politique portées par des youtubeurs engagés ont joué un rôle clef dans le traitement médiatique du mouvement en y apportant des thèmes de fond liés à l'action publique. Nous montrons aussi que si les profils thématiques des chaînes Gilets jaunes et de contre-information ont quelques points communs, cela ne porte pas pour autant sur des topics conspirationnistes. De plus, ces deux catégories de chaîne présentent aussi des divergences importantes, notamment l'intérêt spécifique des chaînes Gilets jaunes pour les sujets de citoyenneté, un sujet rarement traité par les chaînes de contre-information. Enfin, nous montrons que le Grand Débat ne s'est pas simplement imposé comme un simple topic supplémentaire, il a modifié la structure de l'espace médiatique en se substituant aux discours critiques portés à l'encontre de la démocratie représentative. En guise de conclusion, nous proposons un regard réflexif sur l'analyse quantitative de contenu comme « statactivisme » dans les débats autour des biais de représentation médiatique des mouvements sociaux.

***Mots-clés*** **:** *agenda médiatique, analyse de contenu, Gilets jaunes, topics model.*

**ABSTRACT**

This article reports the results of a quantitative content analysis on a corpus of YouTube video subtitles related to the media treatment of the yellow vests movement in France. Through mobilising the topic models method, this article shows that the subtitles of YouTube videos are a material that can be investigated in analysing the media space. If we observe an over-representation of protest and spectacular violence in the traditional media, the "political popularization" channels carried by native YouTubers have played a key role in the media treatment of the movement by bringing to it substantive themes linked to public action. We also show that if the thematic profiles of the Yellow Vests and "counter-information" channels have some points in common, this does not necessarily relate to conspiratorial topics. In addition, these two channel categories also show significant differences, in particular the specific interest of Yellow Vests channels for citizenship issues, a subject rarely covered by counter-information channels. Finally, we show that the Great Debate did not simply impose itself as a simple additional topic, it modified the structure of the media space by replacing the critical speeches made against representative democracy. To conclude, we offer a reflective look at quantitative content analysis as a "statistic activism" in the debates around the media representation biases of social movements.

***Keywords***: *media agenda, content analysis, Yellow Vests, topics model.*

1. Maître de conférences en sociologie, LISIS – CNRS/Université Gustave Eiffel/Inrae, bilel.benbouzid@univ-eiffel.fr
2. Senior data scientist, IBM France

## 1. Introduction

Le mouvement des Gilets jaunes est apparu comme une forme paroxystique des reconfigurations de l'écosystème médiatique (Cardon et Granjon, 2010 ; Ferron, 2019 ; Granjon, 2014a, 2018) par au moins trois aspects : une autonomie revendiquée des Gilets jaunes vis-à-vis des médias traditionnels (Le Bart, 2020) qui rompt avec les situations observées classiquement de militantisme de *prime time*, c'est-à-dire de coopérations/conflits entre activiste et journaliste dans l'accès à l'espace médiatique pour la représentation des groupes mobilisés (Ferron, 2019) ; un soutien rapide et spontané d'une série de médias autonomes et alternatifs natifs du web qui ont permis de servir des intérêts sociaux correspondant à ceux que les Gilets jaunes entendaient défendre ; une capacité à s'emparer des réseaux sociaux et à y construire un dispositif d'échange horizontal, pour non seulement coordonner les actions sur le terrain des manifestations, sensibiliser à des situations d'injustice et faire circuler des cadres d'interprétation des problèmes, mais aussi, et surtout, pour débattre de la couverture médiatique accordée au mouvement, les Gilets jaunes se réfugiant ainsi sur Facebook dans une bulle protectrice de filtre de l'information.

Comment analyser la médiatisation du mouvement des Gilets jaunes dans cette nouvelle économie de la représentation médiatique (Granjon, 2014b) ? Pour répondre à cette question, nous avons mené une analyse quantitative de la couverture médiatique des Gilets jaunes sur la plateforme YouTube. Cette plateforme nous intéresse, car les canaux médiatiques de la plupart des acteurs qui composent l'espace public s'étendent désormais vers elle : presse écrite, radio, télévision, média alternatifs, partis politiques, associations militantes, citoyens ordinaires, etc. – autant d'acteurs présents sur YouTube qui sont mis sur le même plan dans l'espace médiatique par la possibilité de créer une chaîne et publier du contenu. Dans ce contexte, comment se forme l'agenda médiatique (McCombs and Shaw, 1972) selon les types d'acteurs qui composent l'espace médiatique et politique sur YouTube ? Et quelles ont été les évolutions de cet agenda ?

Si Facebook est l'espace numérique de la contestation des Gilets jaunes, YouTube est celui d'où l'on peut observer le mieux la lutte médiatique entre une pluralité de médias pour imposer un sens légitime à donner à l'événement. Tel est le « pari » à l'origine de cette enquête – penser YouTube comme un espace médiatique sur lequel peut s'opérer une analyse quantitative de contenu. En effet, les discours produits sur YouTube peuvent être agrégés pour représenter l'écosystème médiatique dans son ensemble en un corpus homogène, dans le sens de contenu partageant un support de communication commun, en l'occurrence la vidéo (explorée comme une source textuelle à partir des sous-titres). YouTube apporte ainsi une *commodité* d'analyse du traitement médiatique.

En explorant un corpus YouTube, notre enquête présente ainsi un intérêt vis-à-vis des quelques analyses quantitatives de contenus réalisées sur le vif et visant à analyser les couvertures médiatiques des Gilets jaunes (Sebbah *et al.*, 2018 ; Sebbah *et al.*, 2019). Pour intégrer les reconfigurations de l'écosystème médiatique, les chercheurs ont tenté de rendre compte des tensions entre la parole citoyenne des Gilets jaunes sur les réseaux sociaux et l'information journalistique. Ces études ont donc comparé des contenus produits sur différents espaces, mais l'hétérogénéité des corpus collectés implique de dissocier les explorations selon les espaces médiatiques car les contenus restent incommensurables – des journaux télévisés et des tweets dans une enquête de l'INA (Poels et Lefort, 2019) et des articles de presse papier et des discussions sur Facebook dans une série de rapports du LERASS mentionnés plus haut (voir la synthèse dans Souillard *et al.*, 2020). Cette dissociation rend difficile la comparaison des couvertures médiatiques. En revanche, YouTube offre la possibilité de comparer des contenus produits par des types d'acteurs différents qui se sont engagés sur la plateforme dans une lutte pour la représentation médiatique.

Cet article[3] vise à rendre compte des résultats d'une enquête quantitative sur un corpus de vidéos YouTube clairement délimité. Pour en assurer l'exploration, nous avons utilisé les *topics models*. Cette méthode développée dans le domaine de la linguistique computationnelle depuis plus d'une dizaine d'années est devenue classique dans le domaine de l'analyse quantitative de contenu en SHS (Cointet et Parasie, 2018 ; Evans and Aceves, 2016 ; Greene and Cross, 2017 ; Lindstedt, 2019 ; Wesslen, 2018).

Après avoir décrit dans une première partie la construction du corpus, et la méthode des *topics models*, nous montrons dans une deuxième partie les principaux résultats obtenus. Dans cette deuxième partie *Résultats*, nous interprétons d'abord de façon interne les topics en représentant de deux manières différentes les liens inter-topics du point de vue des termes qui les définissent (un dendrogramme et un graphe). Une fois les topics interprétés, une série d'analyses statistiques sont appliquées afin de les croiser avec des catégories de chaînes YouTube d'une part, et les actes du mouvement d'autre part. Dans une dernière partie *Discussion et conclusion*, nous analysons les différents agendas médiatiques du mouvement des Gilets jaunes selon les acteurs qui composent l'espace médiatique ainsi que l'évolution du cadrage médiatique du problème des Gilets jaunes au fil des actes. Enfin, en guise d'ouverture vers de futures recherches, nous concluons sur les enjeux de l'analyse quantitative de contenu comme *statactivisme* (Bruno et Didier, 2014) dans les débats autour des biais de représentation médiatique des mouvements sociaux.

## 2. Données et méthode

### 2.1 Corpus

Nous avons représenté l'espace médiatique et politique produit sur YouTube sans limiter notre corpus aux seules chaînes de production journalistique de l'information. Ce critère d'ouverture tient à la spécificité de l'espace numérique : dans un contexte où la distinction entre médias traditionnels et « médias sociaux » est de moins en moins évidente, comment les différents formats de « communication » sont-ils liés les uns aux autres et qui parvient le mieux à définir l'agenda médiatique ? Pour être en mesure de répondre à cette question, nous avons étendu notre corpus à tous les types de chaînes produisant des opinions, des analyses et des décryptages ; nous avons aussi inclus les chaînes explicitement liées à des entités économiques, politiques, syndicales et associatives ; enfin, nous avons intégré les organisations de services publics qui gèrent sur YouTube leurs relations publiques.

Pour répondre à ce critère de pluralité des formats de communication, nous avons construit des listes de chaînes à partir d'une typologie d'acteurs qu'il nous semblait important de pouvoir situer dans l'espace de YouTube : les médias professionnels ; les chaînes de youtubeurs notoires tournés vers la politique ; les associations militantes ; les députés ; les chaînes de candidats aux élections européennes ; les chaînes de partis politiques ; les chaînes créées à l'occasion des Gilets jaunes ; les chaînes d'associations tournées vers des causes publiques ; les chaînes de grandes institutions publiques ou privées – nous sommes ainsi parvenus à dix types d'acteurs dont nous avons cherché la présence sur YouTube. Sur cette base de chaînes construite *a priori*, nous avons mobilisé différentes procédures de collecte (pour l'essentiel en suivant le réseau des chaînes qui se recommandent mutuellement sur YouTube et en consultant la base de données de Wizdéo[4]). Nous avons ainsi atteint un effectif total de 1400 chaînes. Bien qu'imparfaite, incomplète et discutable, la méthodologie utilisée pour construire ce corpus met en évidence les multiples acteurs en compétition et, dans le même temps, représente les différents types de chaînes publiant du contenu à caractère politique et médiatique en France sur YouTube.

3. Une version de travail plus longue de cet article est accessible en ligne sur le portail HAL à cette adresse : https://hal.archives-ouvertes.fr/view/index/docid/3064932. On peut se reporter à cette version pour une présentation détaillée des différentes méthodes de calcul des topics et la manière dont nous les avons mises en œuvre dans cette enquête.
4. Wizdéo est un réseau multi-chaînes français dont l'une des spécialités est de collecter et préparer des données relatives aux chaînes YouTube.

Ainsi, il faut interpréter notre corpus comme un échantillon non aléatoire et raisonné qui permet de représenter YouTube comme un espace de lutte discursive autour de la fixation de l'agenda médiatique. À première vue, les acteurs s'affrontent dans cette lutte à armes égales : YouTube met tous les comptes sur un même plan, peu importe la nature du contenu produit ou diffusé, tous les comptes sont des chaînes. L'étude de l'écosystème médiatique et politique sur YouTube permet ainsi de mêler, dans un corpus homogène, médias traditionnels, activistes, militants, organisations, vlogueurs anonymes – tous engagés dans une lutte discursive sur la manière de dire la réalité.

Pour rendre compte des chaînes qui composent l'espace public et médiatique sur YouTube, nous dégageons des catégories de chaînes, en cherchant des régularités ou des regroupements de chaînes partageant des caractéristiques communes. Le tableau 1 ci-dessous présente chaque catégorie ainsi que des exemples de chaînes prototypiques pour chacune d'entre elles[5].

**Tableau 1** – *Description des catégories de chaînes codées manuellement avec quelques exemples de chaînes exemplaires*

| Catégories | Description | Chaînes exemplaires |
|---|---|---|
| Politique | Chaînes dont l'engagement politique est directement lié à un parti politique. La chaîne peut représenter une personne ou un parti. | - Jean-Luc Mélenchon<br>- Groupe Républicains Assemblée nationale<br>- LREM |
| Contre-information | Deux types de contre-information : 1) les chaînes mettant l'accent sur le rôle principal de la manipulation cachée par quelques personnes puissantes ou des groupes en particulier (les sionistes, les féministes, les gays, Big Pharma etc.) ; 2) les chaînes utilisant la couverture de l'actualité à des fins de déstabilisation politique. | - ERTV Rhône-Alpes<br>- Boulevard Voltaire<br>- RT France<br>- Sputnik France |
| Gilets jaunes | Chaînes spécialement créées à la faveur du mouvement des Gilets jaunes ou qui ont orienté leur ligne éditoriale en faveur du mouvement, se revendiquant par-là Gilets jaunes. | - Gilets Jaunes Commercy<br>- Éveil Global Conscience Gilets Jaunes<br>- Isadora Duncan |
| Médiation | Chaînes spécialisées dans le décryptage, la vulgarisation, pouvant néanmoins être attachées à une critique sociale. | - Hugo Décrypte<br>- Xerfi Canal<br>- Osons causer |
| Médias alternatifs | Chaînes de médias d'opinion et d'analyse politique de l'actualité. | - Mediapart<br>- Le Média<br>- Thinkerview |
| Médias mainstream | Chaînes des journalistes professionnels qui délivrent un contenu d'information grand public. On retrouve à la fois les médias télévisuels, radio, la presse écrite, les émissions de TV, mais aussi les *pure players*. | - France Inter<br>- L'Obs<br>- Arte<br>- Brut |

5. La liste complète des chaînes relevant de chaque catégorie est consultable en annexe dans la version longue.

| | | |
|---|---|---|
| Information locale | Chaînes d'information « locale », associées à un territoire géographique délimité (villes, régions) | - Télé Lyon Métropole<br>- France 3 Pays de la Loire<br>- Journal Citoyen Haute-Marne |

Une fois la liste de chaînes constituée et catégorisée, il a fallu délimiter le corpus de vidéos. Nous disposions au départ de près de 350000 vidéos produites par cette liste de chaînes sur la période s'étendant de juin 2018 à juillet 2019[6]. Nous avons voulu restreindre ce corpus aux seules vidéos présentant un intérêt, direct ou indirect, quant à la problématique du traitement du mouvement des Gilets jaunes. De plus, nous avons réduit les vidéos de notre corpus aux seuls segments textuels pertinents pour notre enquête. Pour ce faire nous avons mobilisé différentes techniques, tant statistiques que de traitement du langage naturel, pour sélectionner les vidéos pertinentes à notre analyse. De plus, nous avons non seulement sélectionné les vidéos qui mentionnent explicitement les Gilets jaunes dans le titre, la description, les tags ou les sous-titres, mais aussi celles dont les textes des sous-titres montrent une proximité avec la thématique des Gilets jaunes, en calculant un score de similitude avec un texte de référence traitant des Gilets jaunes[7]. On obtient un nombre final d'un peu plus de 26000 vidéos.

### 2.2 Les topics générés

Pour explorer ce corpus, nous nous sommes tournés vers une méthode d'analyse inductive des données linguistiques : les *topic models*, des algorithmes capables d'opérer des calculs sur une représentation numérique des documents textuels sans exiger de modèle *a priori* (Blei, 2012). Au lieu d'établir *a posteriori* une catégorisation des contenus, l'analyste se laisse en quelque sorte guider par l'identification algorithmique de classes latentes d'occurrence des termes lexicaux (mots, groupes de mots) dans les documents. Le calcul des *topics models* repose sur la distribution des termes lexicaux dans la collection des documents. Il produit des listes de *topics* constitués de termes qui co-occurrent dans les documents selon différents *patterns*. On peut considérer un topic comme un sujet (ou thème) abordé dans les documents, mais il faut aussi savoir que certains de ces topics sont plus des marqueurs stylistiques que des sujets de discussion, d'où l'emploi du terme neutre (en français) de « topic ». Chaque document peut contenir plusieurs topics, et chaque topic a une cohérence interne qui dépend du niveau de coprésence des termes les plus représentatifs du topic dans les documents et par rapport aux autres topics.

Dans cette étude, nous avons choisi de mettre en œuvre la méthode NMF (Arora *et al.*, 2012 ; Lee and Seung, 1999) qui est une approche issue de l'algèbre linéaire[8]. La problématique classique d'une analyse de *topic model* est celle de la détermination du nombre idéal de topics en fonction de diverses métriques de qualité (Mimno *et al.*, 2011 ; Röder *et al.*, 2015), en déterminant le « sweet spot » parvenant au meilleur compromis entre les différentes exigences de qualité (Greene *et al.*, 2014). Si un nombre autour de 50 serait *a priori* le nombre de topics idéal selon les métriques calculées, nous avons choisi d'en conserver 72 car l'évaluation visuelle a permis de constater qu'à ce niveau les topics demeuraient suffisamment cohérents et apportaient un niveau de granularité plus riche pour les analyses ultérieures. Il est courant dans l'évaluation du nombre de topics que les évaluations humaines, donc visuelles, ne corroborent pas tout à fait celles des algorithmes. La signification des 72 topics peut être rapidement appréhendée en listant pour chacun d'entre eux les 10 termes qui y sont les plus représentés (voir le tableau 2).

6. Si les premières manifestations des Gilets jaunes apparaissent en octobre 2018, nous avons souhaité étendre notre corpus à juin 2018 dans la perspective d'un repérage de topics préexistants au mouvement.
7. Ce texte a été constitué d'extraits de l'article de Wikipedia France sur les Gilets jaunes, et complété par un florilège de citations sur ces événements, disponible sur un site aussi lié à Wikipedia.
8. On distingue deux grandes catégories d'approches algorithmiques pour l'analyse de topics : les approches probabilistes générativistes, où chaque document est considéré comme généré à partir de probabilités de topics sur les documents, et de termes sur les topics, chacunes avec leur propre distribution ; et les approches issues de l'algèbre linéaire, en procédant par factorisation de matrices.

**Tableau 2** – *Liste des 72 topics*

| |
|---|
| **topic 0 - action :** faire, essayer, aller, sorte, besoin, pouvoir, déjà, passer, train, mettre |
| **topic 1 - mesures :** mesure, annoncer, ministre, annonce, président_république, prendre, édouard_philippe, réponse, milliards_euros, répondre |
| **topic 2 - manifestations 1 :** manifestant, forces_de_l'ordre, gaz_lacrymogène, affrontement, crs, disperser, place_de_la_république, calme, tension, situation |
| **topic 3 - projet :** projet, association, ici, travail, place, mettre, travailler, également, permettre, action |
| **topic 4 - taxation carburants** : taxe, voiture, diesel, carburant, essence, augmenter, véhicule, taxer, transition_écologique, taxe_carbone |
| **topic 5 -** discours politique : politique, chose, croire, moment, société, pouvoir, manière, évidemment, forme, finalement |
| **topic 6 - grand débat 1 :** débat, sujet, organiser, débattre, grand_débat_national, topic, lettre, participer, proposition, président_république |
| **topic 7 - Macron !! :** macron, macro, france, bon, monsieur_macron, merdia, castaner, sarkozy, voter, vouloir |
| **topic 8 - droit & loi :** loi, texte, droit, article, député, constitution, liberté, manifester, sénat, assemblée_nationale |
| **topic 9 - partis politiques :** gauche, droite, parti, extrême_droite, socialiste, parti_socialiste, benoît_hamon, républicain, libéral, génération |
| **topic 10 - finances** : argent, banque, système, payer, état, milliard, dette, riche, financier, économie |
| **topic 11 - gilets jaunes :** gilet_jaune, jaune, gilet, rond-point, rond_point, soutenir, eric_drouet, crise, rencontrer, début_du_mouvement |
| **topic 12 - police 1 :** policier, police, quartier, commissariat, flic, suicide, violences_policières, enquête, police_nationale, effectif |
| **topic 13 - police 2 :** collègue, hiérarchie, fonctionnaire, police_nationale, maintien_de_l'ordre, monsieur_le_ministre, suicide, ordre, sécurité, service |
| **topic 14 - familier :** oui, bon, savoir, vrai, aimer, sûr, petit, bonjour, accord, mal |
| **topic 15 - écologie** : écologie, climat, écologique, transition_écologique, action, écologiste, planète, environnement, climatique, europe |
| **topic 16 - débat public :** monsieur, monsieur_macron, avoir, venir, écouter, être, entendre, croire, savoir, demander |
| **topic 17 - très familier :** truc, ouais, mec, merde, putain, savoir, live, bon, coup, salut |
| **topic 18 - discours action :** falloir, croire, sûr, solution, problème, important, savoir, mettre, moment, arrêter |
| **topic 19 - infos générales** : matin, rtl, hier, bonjour, hier_soir, heure, strasbourg, ministre, édouard_philippe, midi |
| **topic 20 - syndicat** : syndicat, cgt, grève, premier_mai, syndical, convergence, salarié, travailleur, syndicaliste, cfdt |
| **topic 21 - violences (condamnées)** : violence, violent, condamner, voir, légitime, image, violences_policières, justifier, réponse, république |
| **topic 22 - médias :** journaliste, média, presse, information, journal, médiatique, image, bfm, france, article |
| **topic 23 - en direct :** demain, soir, bonsoir, heure, attendre, venir, journée, évidemment, monde, entendre |
| **topic 24 - mobilisation** : mobilisation, mobiliser, acte, semaine, continuer, rassemblement, chiffre, poursuivre, ministère_intérieur, week-end |
| **topic 25 - antisémitisme :** antisémitisme, juif, antisémite, haine, acte_antisémite, alain_finkielkraut, acte, antisionisme, sioniste, racisme |
| **topic 26 - manifestations 2** : manifestation, manifester, déclarer, interdire, manif, liberté, organisateur, organiser, lieu, arrêter |
| **topic 27 - blocages :** bloquer, blocage, camion, ici, péage, rond-point, autoroute, rond_point, action, automobiliste |
| **topic 28 - mouvement GJ** : mouvement, soutenir, leader, soutien, france, parti_politique, organiser, structurer, continuer, mouvement_social |
| **topic 29 - action gouvernement :** gouvernement, ministre, politique, édouard_philippe, porte_parole, exécutif, opposition, majorité, mettre, benjamin_griveaux |
| **topic 30 - revenus salaires :** euro, 100, payer, monnaie, smic, prime, euro_mois, 200, toucher, mois |
| **topic 31 - modalités :** effectivement, justement, finalement, cas, peut-être, évidemment, également, vrai, rappeler, dire |
| **topic 32 - à Paris :** paris, parisien, ville, capitale, france, province, quartier, partout, arrondissement, bordeaux |
| **topic 33 - féminisme :** femme, homme, féministe, droit_des_femmes, victime, viol, combat, féminisme, mari, lutte |
| **topic 34 - commerce & GJ :** commerçant, commerce, noël, magasin, chiffre_d'affaires, centre_ville, client, boutique, fermer, ville |
| **topic 35 - lycéens étudiants :** lycéen, lycée, jeune, étudiant, réforme, élève, établissement, professeur, enseignant, université |
| **topic 36 - retraites :** retraite, retraité, réforme, travailler, an, pension, 62, fonctionnaire, réforme_des_retraites, fonction_publique |
| **topic 37 - grand débat 2 :** grand_débat, grand_débat_national, proposition, contribution, participer, réunion, réponse, grand, attendre, fin |
| **topic 38 - agriculture :** produit, agriculteur, prix, consommateur, producteur, agriculture, acheter, magasin, client, vendre |
| **topic 39 - réseaux sociaux :** vidéo, youtube, facebook, chaîne, commentaire, réseau_social, live, internet, twitter, message |
| **topic 40 - fiscalité :** impôt, payer, riche, taxe, fiscal, fiscalité, impôt_sur_revenu, revenu, baisser, milliard |
| **topic 41 - peuple & pouvoir :** peuple, pouvoir, peuple_français, révolution, peuple_de_france, démocratie, élite, pays, système, france |
| **topic 42 - cortège manifestation** : cortège, place, ici, heure, rue, place_de_la_république, rassemblement, rejoindre, calme, marche |
| **topic 43 - affaire Benalla :** elysée, alexandre_benalla, affaire, monsieur_benalla, sénat, passeport, benalla, mediapart, affaire_benalla, justice |
| **topic 44 - Corse :** corse, nationaliste, île, dialogue, bastia, région, continent, ajaccio, élu, visite |
| **topic 45 - présidence république** : emmanuel_macron, président, chef_d'état, président_république, crise, elysée, quinquennat, politique, communication, françois_hollande |
| **topic 46 - général 1 :** personne, vraiment, essayer, coup, passer, chose, monde, forcément, justement, rapport |
| **topic 47 - général débats :** question, poser, répondre, pose, réponse, sujet, savoir, cas, évidemment, exemple |
| **topic 48 - démocratie** : citoyen, démocratie, élu, démocratique, politique, proposition, pouvoir, institution, justement, référendum_initiative_citoyenne |
| **topic 49 - violence policière 1** : blessé, arme, grenade, utiliser, maintien_de_l'ordre, blessure, oeil, lbd, blesser, flashball |
| **topic 50 - élections européennes :** liste, voter, européen, élections_européennes, rassemblement_national, élection, candidat, campagne, parti, europe |
| **topic 51 - violence policière 2 :** gazer, taper, retraité, police, aimer, crs, mal, manifeste, plaire, continuer |
| **topic 52 - casseurs :** casseur, black_block, casser, forces_de_l'ordre, maintien_de_l'ordre, christophe_castaner, ministre_intérieur, dispositif, interpeller, casse |
| **topic 53 - RIC :** référendum, constitution, voter, référendum_initiative_citoyenne, vote, suisse, démocratie, élection, parlement, référendum_initiative_populaire |
| **topic 54 - géné. live YT** : ami, live, constitution, petit, monde, aller, constituer, bon, salut, ici |
| **topic 55 - familles :** enfant, an, famille, parent, vie, école, vivre, jour, venir, mère |
| **topic 56 - expression sentiments :** colère, exprimer, entendre, comprendre, pays, croire, profond, dialogue, manifester, réponse |
| **topic 57 - politique locale** : maire, commune, élu, ville, habitant, territoire, département, mairie, bordeaux, métropole |
| **topic 58 - nation française** : français, france, président_république, pays, croire, républicain, voir, réalité, sujet, évidemment |
| **topic 59 - général 2 :** chose, vouloir, accord, dire, savoir, aller, avoir, passer, mettre, prendre |
| **topic 60 - macro économie :** pourcent, chiffre, sondage, baisse, croissance, an, 10, 20, 2018, hausse |
| **topic 61 - 1er mai :** hôpital, premier_mai, christophe_castaner, intrusion, patient, ministre_intérieur, médecin, infirmier, soignant, service_réanimation |
| **topic 62 - pol. sécurité routière :** radar, route, 80_kilomètres_heure, vitesse, sécurité_routière, automobiliste, accident, chiffre, département, voiture |
| **topic 63 - pouvoir achat :** pouvoir_d'achat, salaire, augmenter, smic, retraité, salarié, augmentation, prime_d'activité, revenu, travail |
| **topic 64 - champs-elysées :** champs_elysées, place_de_l'étoile, avenue, crs, arc_de_triomphe, image, champ, avenue_champs_elysées, ici, rassemblement |
| **topic 65 - revendication :** revendication, entendre, référendum_initiative_citoyenne, porter, justement, revendiquer, exprimer, pouvoir_d'achat, répondre, demande |
| **topic 66 - actes GJ** : samedi, manifester, semaine, week-end, appel, samedi_prochain, rue, vendredi, acte, appeler |
| **topic 67 - France insoumise** : jean_luc_mélenchon, france_insoumise, politique, perquisition, marine_le_pen, mélenchon, député, insoumis, rassemblement_national, assemblée_nationale |
| **topic 68 - description manifestation :** voir, vraiment, passer, regarder, ici, image, train, petit, aller, venir |
| **topic 69 - boxeur :** cagnotte, boxeur, christophe_dettinger, gendarme, soutien, leetchi, soutenir, forces_de_l'ordre, frapper, blessé |
| **topic 70 - justice & GJ :** justice, avocat, prison, procès, juge, tribunal, condamner, juger, dossier, judiciaire |
| **topic 71 - armée :** militaire, armée, guerre, arme, soldat, mort, pays, général, france, opération |

## 3. Résultats

Dans la littérature en sciences sociales, les résultats des *topics models* servent généralement à des tests d'hypothèses destinés à vérifier des hypothèses *a priori* sur des relations entre les

topics et des variables (DiMaggio *et al*., 2013 ; Tsur *et al.*, 2015). Dans cet article, nous avons suivi une autre approche, plus exploratoire, visant à rechercher et découvrir des relations entre les topics entre eux et des variables externes, sans hypothèse a priori (Tukey, 1977). Nous avons privilégié trois techniques analytiques : l'arbre des divergences inter-topics (un dendrogramme), le réseau de similarité de topics (un graphe) et les profils de topics par catégorie de chaînes et par acte des manifestations (des matrices de corrélation).

Avant de commenter ce que ces trois techniques analytiques donnent à voir, il nous faut rappeler la posture analytique avec laquelle nous envisageons cette lecture du corpus. Nous nous inspirons ici des travaux de Franco Moretti sur l'étude quantitative de corpus littéraires : « *Lorsque nous étudions 200 000 romans au lieu de 200*, dit Moretti, *nous ne nous contentons pas de faire la même chose à une échelle 1 000 fois plus importante : nous les étudions de façon différente. Ce changement d'échelle modifie notre relation avec notre objet d'étude, et modifie de fait jusqu'à l'objet lui-même* » (Moretti, 2016). Cette approche de la lecture distante implique de traiter les corpus comme des « objets artificiels », la réalité sociale s'observant *« au sein même des abstractions » produites par les méthodes computationnelles*.

Dans les pages qui suivent, il n'y aura pas de retour au texte, dans le sens d'une démonstration par la citation verbatim des contenus pour illustrer le sens des topics[9]. Bien entendu, nous avons consulté un nombre important de vidéos lorsque nous avons voulu comprendre les topics. Mais le sens du corpus que nous souhaitons produire passera seulement par les abstractions computationnelles, c'est-à-dire les topics et la série de visualisations permettant de les explorer, de découvrir des relations entre les variables et de rendre compte d'une signification d'ensemble du traitement médiatique des Gilets jaunes.

### 3.1 L'arbre des divergences inter-topics

La représentation sous forme d'arbre hiérarchique ou dendrogramme, produit à partir d'un algorithme de classification hiérarchique, constitue dans cette enquête une première étape pour apporter une visualisation d'ensemble des topics. En construisant des classes de topics, le dendrogramme rend compte de la similarité entre topics et de leur niveau commun de regroupement.

On peut observer sur le dendrogramme (figure 1) deux principaux blocs qui s'agrègent en dernier : d'une part, une classe que nous appelons « actions publiques », composée de sujets de fond du débat public et des multiples acteurs et institutions qui y participent et d'autre part, une classe qui indique les « conflits directs », notamment tout ce qui a trait aux manifestations.

9. De notre point de vue, de nombreuses études en humanité numérique commettent l'erreur de traiter de manière computationnelle des corpus, mais en interprétant les résultats comme des objets qualitatifs.

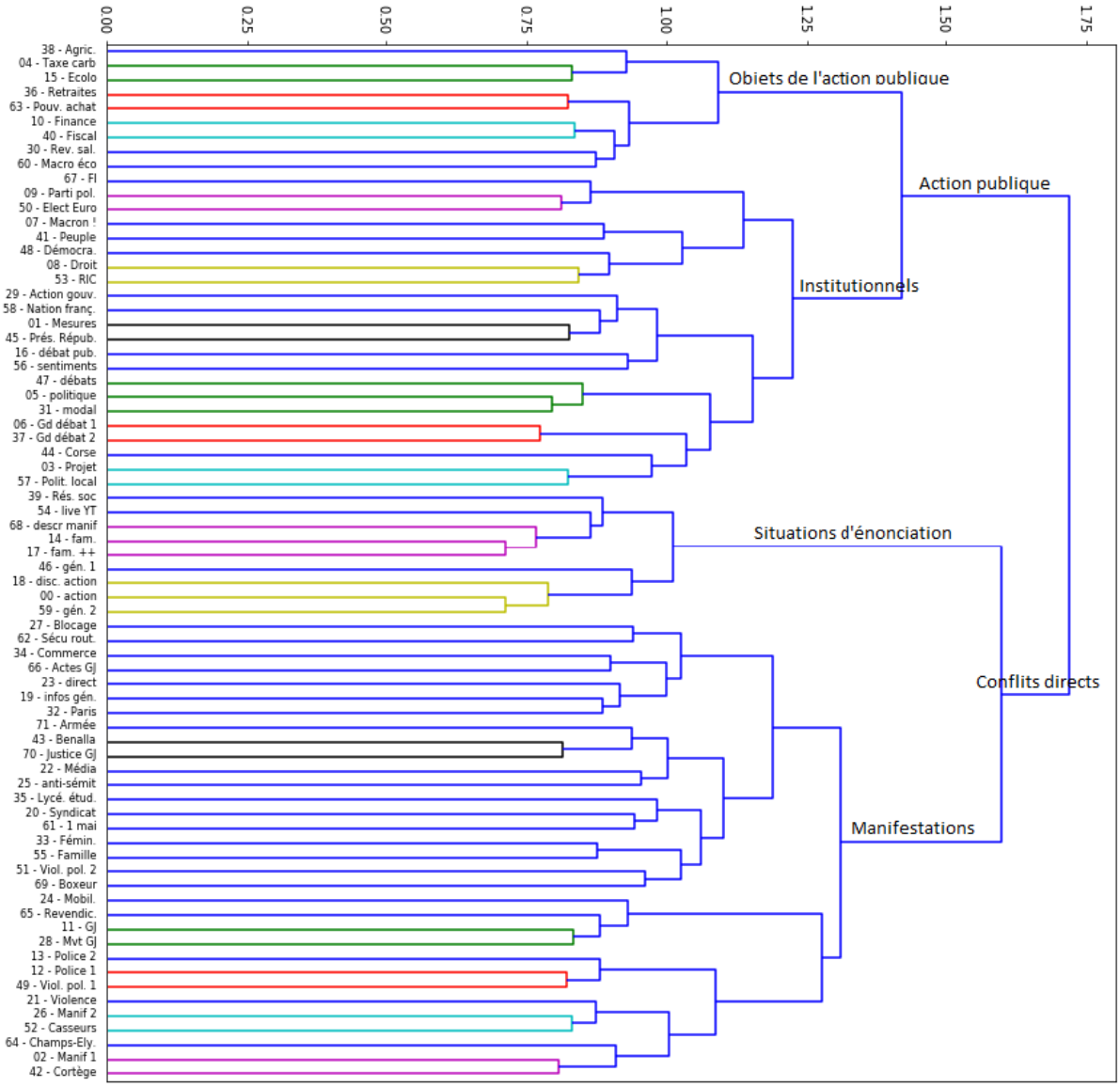



**Figure 1** – *Regroupement hiérarchique ascendant (agglomératif), en utilisant comme distance la valeur 1 - similarité (corrélation de Pearson). La méthode d'agglomération utilisée est celle de Ward, qui permet de mieux identifier des petits regroupements et de ne pas coller arbitrairement entre eux les regroupements. Le seuil de regroupement pour la coloration des branches inférieures du dendrogramme a été pris suffisamment bas pour mettre en évidence la dizaine de groupements les plus pertinents. On lit par exemple que les topics « taxe carburant » et « écologie », très proches, se sont agrégés dès la première itération, puis ces deux topics s'agrègent dans une deuxième itération avec l'agriculture, etc.*

Pour comprendre ces deux principales classes (actions publiques et conflits directs), descendons d'un nœud dans la hiérarchie de la classification, et examinons pour chacune d'elle le tracé de l'arbre en dégageant maintenant les principales classes qui les composent. On observe ainsi que les deux classes que nous venons de présenter sont elles-mêmes décomposées en deux sous-classes. Ainsi l'espace médiatique sur YouTube peut être décomposé en quatre grandes classes de discours :

- **La classe des topics d'objets de l'action publique.** On observe d'abord en partant du haut de la liste des topics, une première classe composée uniquement de topics liés à des enjeux sociaux, environnementaux et écologiques : *T38 Agriculture*, *T04 Taxe carburant*, *T15 Écologie*,

*T36 Retraites*, *T63 Pouvoir d'achat*, *T10 Finance*, *T40 Fiscalité*, *T30 Revenus/Salaires*, *T60 Macro-Économie*. Il s'agit de topics qui renvoient à des objets de l'action publique.

- **La classe des topics institutionnels.** On relève ensuite une classe des topics des institutions qui font et accueillent le débat public : *T67 France Insoumise, T09 Parti Politique, T05 Élections Européennes, T07 Macron, T41 Peuple, T48 Démocratie, T08 Droit, T53 RIC, T29 Action Gouvernementale, T58 Nation Française, T01 Mesure, T45 Président de la République, T16 Débat Public, T47 Débat, T05 Politique, T06 Grand Débat 1, T37 Grand Débat 2, T03 Projet, T57 Politique Locale* et *T44 Corse*. Deux topics viennent rompre la cohérence apparente : les topics *T31 Modal*, renvoyant à des marqueurs de la modalité dans les discussions, et *T56 Sentiments*, composé d'éléments de qualification de la colère des Gilets jaunes.

- **La classe des topics de situation d'énonciation.** En partant du haut de la grande classe des conflits directs, on trouve différents topics indiquant des marqueurs d'oralité : *T39 Réseaux sociaux, T54 Live YT, T68 Description Manif, T14 Familier, T17 Fam ++*. Ces marqueurs d'oralité renvoient à des situations spécifiques de prise de parole. Toujours dans la même classe, on trouve deux autres topics qui ne renvoient pas nécessairement à des marqueurs d'oralité, mais qui proposent des repères d'énonciations : le topic *T00 Action* renvoie à des repères verbaux relatifs à l'agir, comme « faire », « essayer », « réussir », « décider », « obliger », « pouvoir » (suivi de l'infinitif d'un verbe d'action), autant de verbes qui apparaissent comme des marqueurs d'une attente sociale ; le topic *T18 Disc. Action* indique des modalités d'énonciation propres à des énoncés impératifs ou exclamatifs, pour l'essentiel autour d'exhortations à l'action et au changement : « falloir », « électrochoc » et « nouvelle politique » sont les mots et expressions les plus saillants de ce topic[10].

- **La classe des topics autour des manifestations.** Enfin, la classe contenant le plus grand nombre de topics traite des manifestations et des différents éléments qui peuvent leur être associés : *T27 Blocage, T62 Sécurité Routière, T34 Commerce, T66 Actes GJ, T32 à Paris, T43 Benalla, T70 Justice GJ, T22 Média, T25 Anti-sémitisme, T35 Lycée Étudiants, T20 Syndicat, T61 1er mai, T69 Boxeur, T24 Mobilisation, T65 Revendication, T11 GJ, T28 Mouvement GJ, T12 Police 1, T13 Police 2, T21 Violence, T49 Violence Policière 1, T51 Violence policière 2, T26 Manif 2, T52 Casseurs, T64 Champs Élysées, T02 Manifestation 1, T42 Cortège* et *T71 Armée*, ce dernier topic renvoyant pour l'essentiel à l'usage des sentinelles dans la protection des équipements publics. Si les topics *T33 Féminisme* et *T55 Famille* appartiennent à cette classe, c'est parce qu'ils renvoient respectivement, pour une grande part, aux manifestations des femmes Gilets jaunes et aux problèmes rencontrés par les familles pour les gardes d'enfant les samedis. Enfin deux topics de cette classe, *T23 Direct* et *T19 Info générales*, auraient toute leur place dans la classe des topics de situations d'énonciation, mais il s'agit de situations d'énonciation propres aux manifestations.

En résumant les proximités entre topics de façon hiérarchique, le dendrogramme nous permet un premier ordonnancement du corpus en quatre classes qui donnent un aperçu d'ensemble de la topologie de l'espace médiatique sur YouTube. Mais le dendrogramme ne permet l'analyse des topics qu'à travers leurs regroupements successifs. C'est pourquoi nous nous sommes orientés vers une autre méthode, celle de l'analyse de graphes qui représentent la multiplicité des liens entre topics à partir de la coexistence des termes qui les composent.

10. Notons aussi que c'est dans cette classe que l'on trouve deux topics dont il est impossible de discerner un sens spécifique, ce pourquoi ils sont qualifiés de généraux : *T46 General 1* et *T59 Général 2*.

### 3.2 Les systèmes thématiques

En effet, en représentant les contenus de notre corpus comme un arbre de divergence de topics, nous avons permis une lecture simple et efficace des topics, mais nous avons trahi la réalité de la composante textuelle des topics : les mots et expressions occupent des positions multiples à la fois dans les documents et les topics. Il nous faut maintenant trouver le moyen de représenter les interconnexions entre les termes et les documents et des termes entre eux. Pour ce faire, la représentation graphique la plus simple que nous pouvons mobiliser est le réseau des topics similaires.

Dans ce cas, les similitudes entre topics peuvent s'observer en considérant leurs co-occurrences au niveau des documents du corpus ainsi que dans le vocabulaire des termes qu'ils mobilisent. Dès lors qu'on affecte à ces similitudes des mesures numériques, on peut également considérer toute combinaison de ces mesures pour tenir compte à la fois des distributions des topics sur les documents et des distributions des mots sur les topics. Une mesure de similarité entre deux topics se calcule en considérant les vecteurs associés aux deux topics, tant sur la matrice « documents x topics » que sur la matrice « topics x termes ». Nous avons utilisé le très classique coefficient de corrélation de Pearson, et ainsi construit une liste de K*(K-1)/2 similarités entre les topics, en considérant à pondération égale tant les similarités au niveau des documents que celles au niveau des termes.

Pour la représentation graphique du réseau des corrélations entre topics, plutôt que d'avoir un graphe illisible (1390 connexions pour 72 topics), nous n'avons conservé que les connexions dont le poids est supérieur à un seuil retenu de façon à obtenir le nombre de connexions minimales pour qu'aucun topic ne soit isolé. On obtient ainsi un réseau plus lisible de 451 connexions (figure 2), éliminant ainsi les structures les moins importantes du réseau de départ.

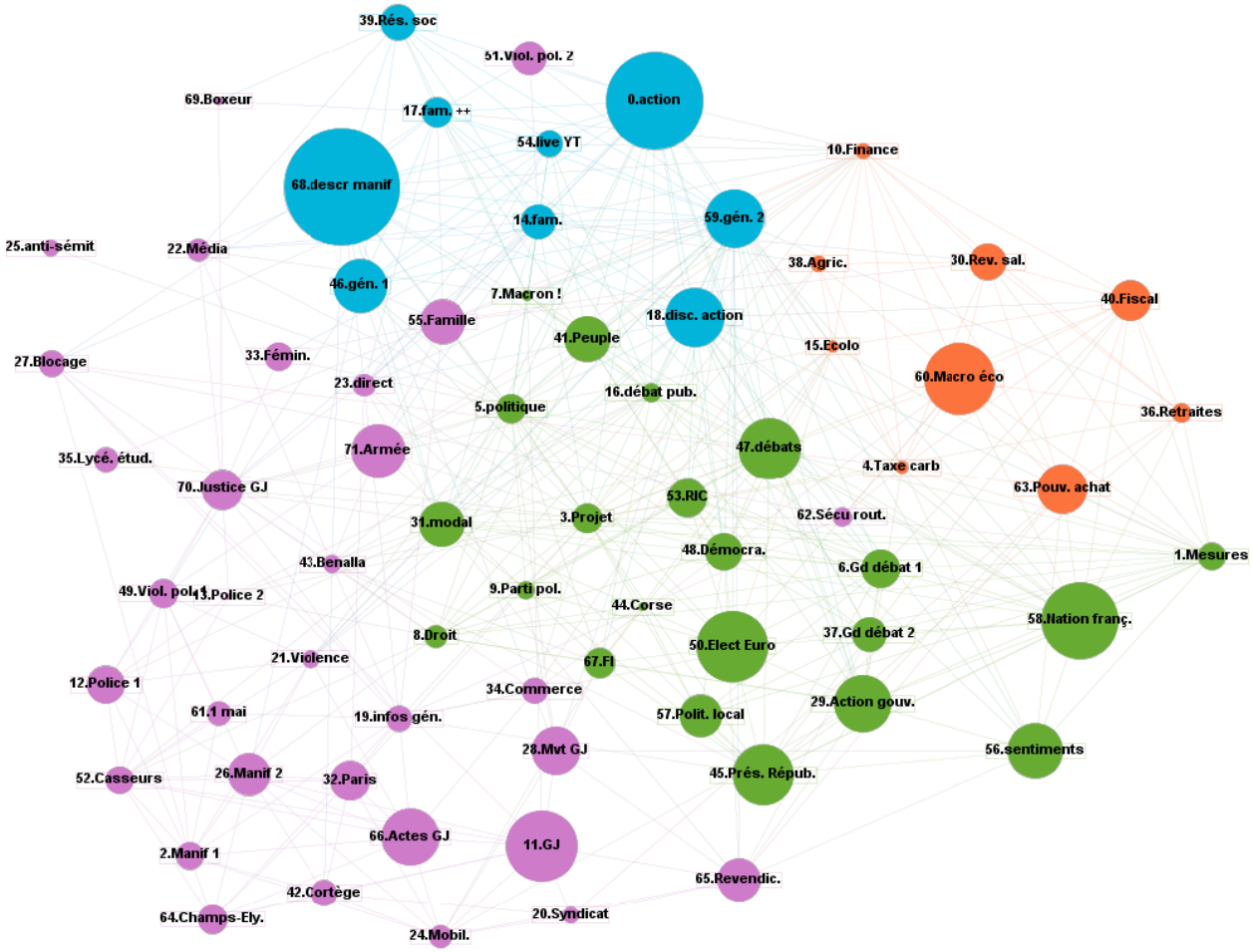



**Figure 2** – *Réseau de topics similaires. Les liens conservés ont un poids > 0,06. Les couleurs correspondent aux classes calculées avec le dendrogramme. En bleu, les topics de situations d'énonciation, en orange les topics des objets de l'action publique, en vert les topics institutionnels et en violet les topics autour des manifestations. L'algorithme de spatialisation utilisé est Force Atlas 2 avec Gephi. La taille des nœuds correspond aux poids des topics dans l'ensemble du corpus.*

Qu'est-ce que cette visualisation des topics en réseau peut nous apprendre de plus sur le contenu de notre corpus ? Elle apporte une nouvelle synthèse synoptique à partir de laquelle nous pouvons mieux prendre en compte la nature réticulaire des topics et leurs rapports avec d'autres topics appartenant à des classes différentes dans le dendrogramme. Bien que ce réseau invite à de nombreuses observations, nous nous concentrons seulement sur les questions laissées ouvertes par le dendrogramme dans la classe des topics institutionnels, pour rappel les topics *T31 Modal*, renvoyant à des marqueurs de la modalité dans les discussions, et *T56 Sentiments*, composé d'éléments de qualification de la colère des Gilets jaunes.

Pour tenter de comprendre *T31 Modal*, confrontons-le à un autre topic stylistique qui indique des marqueurs d'oralité : *T54 Live YT*. Retraçons pour chacun de ces deux nœuds les chemins de connexions vers les autres topics.

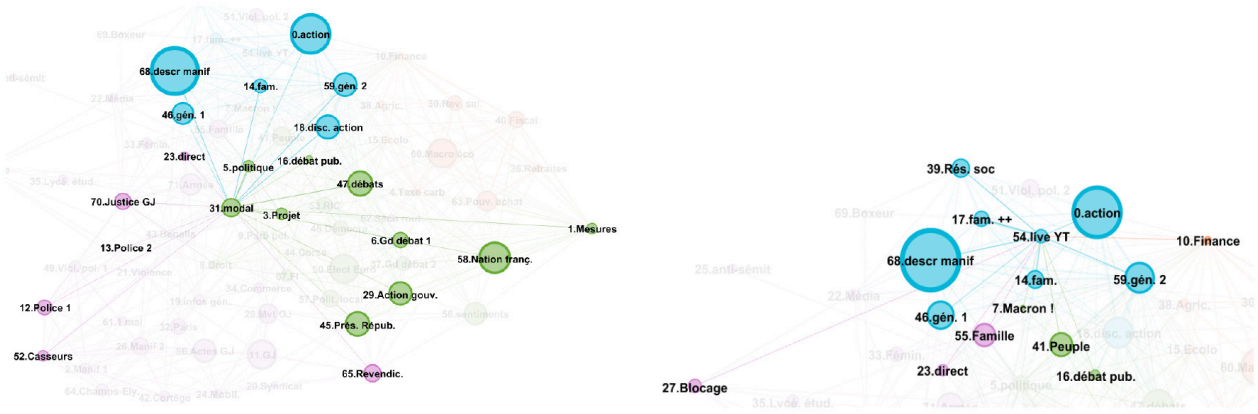



**Figure 3** – *Capture d'écran du voisinage direct des topics* T31 Modal *(à gauche) et* T54 Live YT *(à droite)*

On le voit, les topics *T31 Modal* et *T54 Live YT* (figure 3) produisent des systèmes thématiques très différents : le T31 Modal, plus officiel, se concentre autour des thèmes notoires autour des manifestations (la justice, la police, les casseurs) et incarne la parole du Président de la République (*T45*) ; l'autre, plus informel (connexion forte avec le *T17 Fam++* qui indique des marqueurs de la vulgarité) se connecte sur la famille, la finance et incarne la parole du peuple (*T41*). On comprend mieux dès lors pourquoi le topic *T31 Modal* prend place dans la classe des topics institutionnels : il est spécifique aux contenus officiels.

Procédons de la même manière pour le topic *T56 Sentiments*, en le confrontant à d'autres topics relatifs aux raisons de la colère des Gilets jaunes, à savoir le *T65 Revendication* et le *T29 Action Gouvernementale* qui renvoient respectivement aux revendications des Gilets jaunes et aux solutions apportées par le gouvernement à la crise. En comparant ces trois réseaux (figure 4), on remarque que ces trois topics forment une clique (une triade où toutes les connexions possibles sont réalisées entre ces trois topics), si bien qu'on peut les considérer comme un ensemble formant le système thématique des raisons de la colère des Gilets jaunes. On observe ainsi que les revendications des Gilets jaunes sont pour l'essentiel liées au pouvoir d'achat (commun aux trois réseaux) et à la taxe carburant, laissant de côté les autres sujets de fond comme la macro-économie (*T60*), la finance (*T10*) et l'écologie (*T15*). Dès lors une nouvelle question mérite d'être posée : quel système thématique se forme autour de ces trois autres topics objets de l'action publique ?

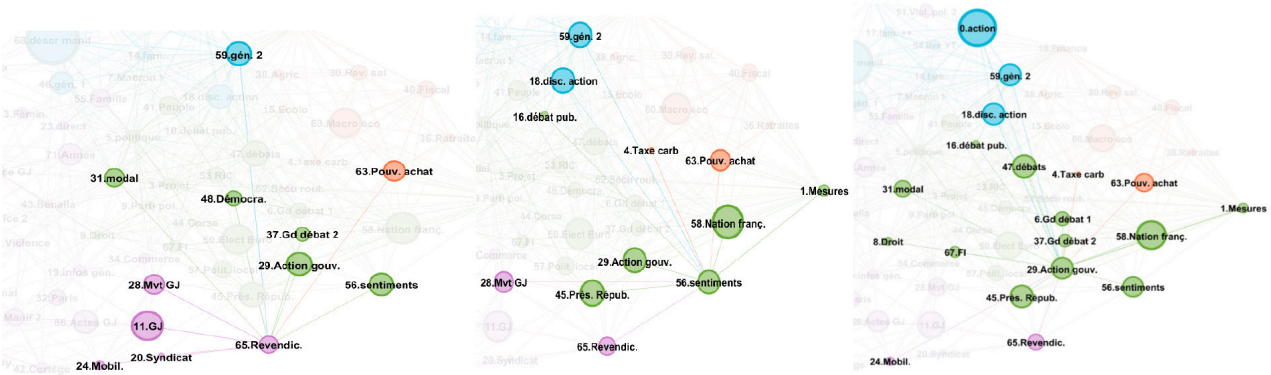



**Figure 4** – *Capture d'écran du voisinage direct, de gauche à droite, des topics* T65 Revendication, T56 Sentiments *et* T29 Action Gouvernementale

Pour répondre à cette question, on peut observer (figure 5) les autres topics appartenant à la

classe des topics de l'action publique dans le dendrogramme. On voit tout d'abord que le topic *T10 Finance* parvient à s'interconnecter à l'ensemble des topics objets de l'action publique. En partant de la finance, on perçoit bien les multiples enjeux interconnectés qui vont du climat au revenu des ménages. En revanche, les topics *T60 Macro-économie* et *T15 Écologie* forment des systèmes thématiques plus restreints. Plus encore, ils ont seulement en commun les topics *T4 Taxe Carburant* et *T38 Agriculture* et ils ne sont pas liés entre eux. Ce qui signifie qu'on parle rarement ensemble de ces deux thèmes qui restent pourtant étroitement connectés dans la réalité.

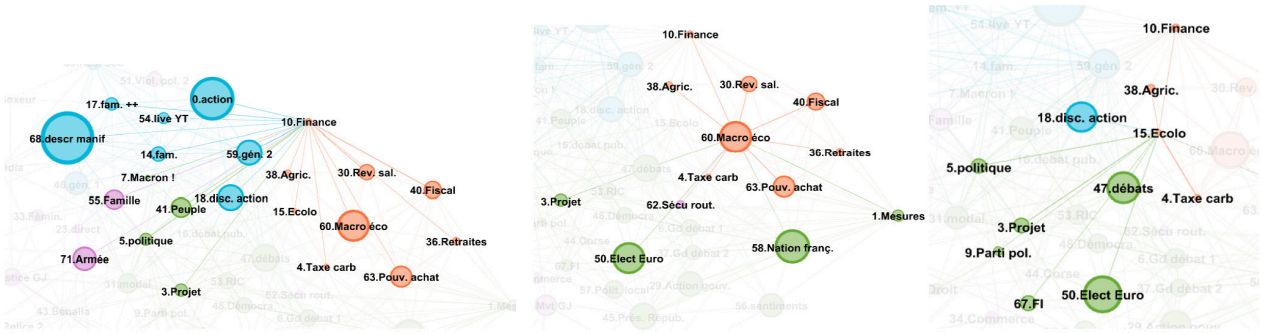



**Figure 5** – *Capture d'écran du voisinage direct des topics* T10 Finance *(à gauche),* T60 Macro-économie *(au centre) et* T15 Écologie *(à droite)*

L'abstraction produite par le réseau ne permet pas d'aller plus en détail dans l'interprétation de ces traitements différenciés des problèmes de fond, mais elle nous permet de montrer des séparations fortes entre des systèmes thématiques.

### 3.3 Les profils de topics selon les catégories

Dans cette première phase d'analyse des résultats, nous nous sommes concentrés sur la compréhension des topics grâce au dendrogramme et au réseau. Mais cette analyse interne des topics n'est pas suffisante pour rendre compte de la structuration de l'agenda médiatique. Il faut la compléter par une caractérisation externe des topics en cherchant les relations existantes entre les topics et les catégories de chaînes d'une part, et les actes du mouvement d'autre part.

Nous représentons les relations entre topics et catégories de chaînes en utilisant un procédé classique : des matrices de corrélation, en particulier des cartes de chaleur sur lesquelles peuvent être appliquées des classifications hiérarchiques selon le sens des corrélations, à la fois pour les entités en ligne et en colonne. La figure 6 montre cette représentation d'ensemble. Le dendrogramme vertical des catégories de chaînes indique une proximité des chaînes selon certains topics. On voit comment les catégories de chaînes forment deux classes distinctes : d'un côté les chaînes qui produisent du contenu de journalistes professionnels opérant au sein de médias traditionnels (catégories Information locale et Médias *mainstream*) et de l'autre une classe rassemblant les catégories de chaînes YouTube qui sont au cœur de l'élargissement et des reconfigurations de l'espace médiatique, cette seconde classe formant deux sous-classes : Gilets jaunes et Contre information d'une part, et Médiation, Médias alternatifs et analyses, et Politique, d'autre part.

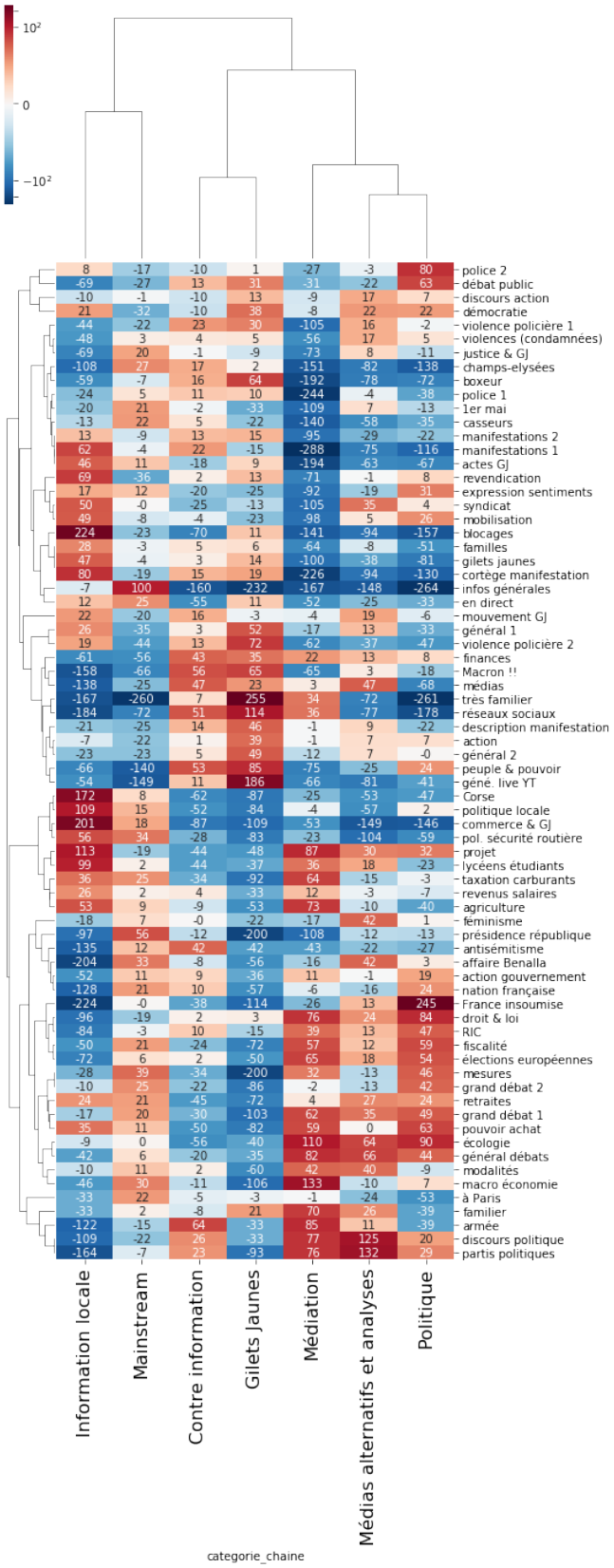



**Figure 6** – *Matrice de corrélation avec classement hiérarchique non supervisé des catégories de chaînes et des topics. Le score indiqué est une mesure d'une déviation odds ratio par rapport à une valeur attendue s'il y avait indépendance entre topics et catégorie de chaînes. Les scores sont visualisés en bleu pour une saillance négative, en rouge pour une saillance positive et en blanc pour un score de saillance nul, donc une absence de corrélation.*

Une première lecture de la carte de chaleur peut être faite en listant les topics dont les corrélations positives sont concomitamment élevées pour chacune des trois classes de catégories de chaînes que nous venons de présenter, qui forment, pour rappel, deux binômes et un trinôme de catégories de chaînes. Le binôme des médias locaux et *mainstream* se constitue par des topics dits institutionnels (*T44 Corse* et *T57 Politiques Locales*) et ceux relatifs aux manifestations (*T34 Commerçants* et *T62 Sécurité Routière*). Pour le deuxième binôme Gilets jaunes et Contre-information, les chaînes ont en commun une surreprésentation de cinq topics, soit le *T69 Boxeur* relatif aux manifestations, le *T54 Live YT* comme style d'énonciation, les *T41 Peuple* et *T07 Macron* dans la classe des topics institutionnels et le *T10 Finances* comme seul topic d'action publique. Enfin, le trinôme de catégories de chaînes « médias alternatifs », « médiation » et « politique » se caractérise en grande partie par une surreprésentation des topics dits institutionnels (*T47 Débat*, *T6 Grand débat 1*, *T50 Élections européennes*, *T53 RIC*, *T8 Droit et Loi*) et la surreprésentation du seul topic *T15 Écologie* parmi les topics de l'action publique.

S'il existe des proximités thématiques entre catégories de chaînes, on voit que chaque catégorie de chaînes a un profil de topics qui lui est propre. Les médias locaux ont porté un intérêt particulier à la couverture des manifestations en général – en témoigne la surreprésentation des topics sur les blocages, les cortèges, les commerçants, les revendications, les témoignages des Gilets jaunes sur les violences policières (corrélation positive avec *T51 Violence policière 2*). Ils ont aussi porté un intérêt particulier aux topics d'action publique (sauf l'écologie qui y est sous-représentée), en particulier l'agriculture.

Les médias mainstream ont un profil bien particulier : ils se caractérisent par un nombre important de topics aux scores de saillance presque nuls. Ce qui signifie que les médias traditionnels ne se distinguent pas par une « sous » ou « sur » représentation pour de nombreux topics. On observe néanmoins une sous-représentation pour les topics de style d'énonciation propres aux contenus de réseaux sociaux (souvent très familiers) et une surreprésentation pour *T19 Information Générale*. Ces médias se distinguent également par une surreprésentation, néanmoins faible, pour les topics de manifestation (*T70 Justice*, *T64 Champs-Élysées*, *T61 1er mai*, *T52 Casseurs*, *T23 Direct*, *T32 à Paris*) et pour le topic *T45 Président de la République*. Enfin, parmi les topics de l'action publique traités par les médias *mainstream*, c'est *T60 Macro-économie* qui a le score de saillance le plus élevé.

Les chaînes de contre-information se caractérisent par un non-traitement des topics d'action publique, c'est-à-dire des topics de fond du débat public, hormis celui de la finance (*T10*) qui atteint le score de saillance le plus élevé comparé à celui des autres catégories de chaînes. Si le dendrogramme sur la matrice indique une sous-classe commune pour les chaînes de contre-information et les chaînes de Gilets jaunes, on observe néanmoins que la contre-information présente des scores de saillance similaires avec les médias *mainstream* pour les topics autour des manifestations, notamment les topics *T25 Anti-sémitisme*, *T52 Casseurs* et *T64 Champs-Élysées*. Cette similitude tient sans doute à la présence de Russia Today France dans la catégorie des chaînes de contre-information qui, de fait, pourrait tout aussi bien appartenir à la catégorie des chaînes *mainstream*. Cette similitude tient aussi peut-être à la logique systématique de réaction de chaînes de contre-information par rapport aux médias *mainstream*. Notons enfin la spécificité de la contre-information : le score de saillance le plus élevé est celui du topic *T71 Armée* qui renvoie à des sujets variés (par exemple, des mises en perspective historiques du mouvement accompagnant des prophéties d'une guerre insurrectionnelle ou mondiale, etc.) ; viennent ensuite dans l'ordre décroissant *T07 Macron*, *T41 Peuple* et *T22 Médias*. On retrouve ici quelques ingrédients thématiques de la rhétorique contre-informationnelle.

La proximité du profil thématique des Gilets jaunes avec celui de la contre-information, telle qu'indiquée par le dendrogramme de la figure 6, doit être considérée avec précaution. En effet, si les chaînes de Gilets jaunes partagent avec la contre-information la caractéristique de

s'intéresser beaucoup à la finance, elles s'en distinguent en accordant un intérêt particulier pour la classe des topics institutionnels, notamment *T48 Démocratie* et *T16 Débat Public*, les Gilets jaunes ayant fait du débat sur la citoyenneté le cœur de leurs discussions.

De fait, les chaînes qui atteignent les scores de saillance les plus élevés sur les topics d'objet de l'action publique appartiennent aux catégories « médiation », « médias alternatifs et analyses », et « politique ». Parmi ces types de chaînes, la catégorie « médiation » se distingue néanmoins par les scores les plus saillants pour les topics *T60 Macro-économie*, *T15 Écologie*, *T38 Agriculture* et *T04 Taxation Carburant*. Soulignons que le topic *T71 Armée* atteint aussi pour cette catégorie de chaînes le score de saillance le plus élevé si on le compare à celui atteint par les autres catégories.

La catégorie « médias alternatifs et analyses » se caractérise par la saillance importante de *T05 Discours Politiques*, *T09 Partis Politiques*, *T43 Benalla*, *T33 Féminisme* et *T21 Violences Condamnées*. Notons des corrélations négatives avec les topics *T60 Macro-économie*, *T38 Agriculture*, *T30 Revenus et Salaires* et *T04 Taxation Carburant*, et une corrélation nulle avec le topic *T63 Pouvoir d'Achat*. Le topic *T11 Gilets jaunes*, qui renvoie à des discussions générales sur le sens à donner au mouvement, y est largement sous-représenté. Dès lors, tout porte à croire que ce qui a intéressé surtout les médias alternatifs durant la première année du mouvement relève surtout du scandale, de la dénonciation et des luttes partisanes.

Enfin, les chaînes politiques se distinguent par une surreprésentation de *T67 France Insoumise* qui tient à la surproduction de contenus du parti de Jean Luc Mélenchon et de ses députés. Dans une moindre mesure, les partis politiques sur YouTube ont concentré leur ligne éditoriale sur les topics institutionnels *T06 Grand Débat 1*, *T37 Grand Débat 2*, *T16 Débat Public* et *T56 Sentiment*s. Parmi les topics de l'action publique, seul *T63 Pouvoir d'Achat* est surreprésenté. Enfin, le seul topic relatif aux manifestations traité par les chaînes politiques est le *T13 Police 2* qui discute du manque de moyens donnés aux policiers pour encadrer les manifestations. Sans étonnement, les chaînes de partis politiques, et en particulier celles de la France Insoumise, se concentrent sur une critique des institutions (notamment le débat public) et de la gestion des manifestations, tout en s'autoproclamant porte-parole des Gilets jaunes (la saillance des *T56 Sentiments* et *T63 Pouvoir d'Achat* en est un révélateur). Autrement dit, les chaînes politiques n'ont pas eu d'attachement spécifique aux définitions des objets de fond du débat public, du moins en ce qui a concerné les Gilets jaunes.

### 3.4 Les profils de topics selon les actes du mouvement

Ces profils thématiques selon les catégories de chaînes rendent compte de manière globale de la spécificité des catégories de chaînes, mais il s'agit d'une représentation statique qui cache des dynamiques temporelles. Observe-t-on des topics caractéristiques de certaines périodes ? Comment les topics se distribuent-ils dans le temps[11] ? Pour répondre à ces questions, nous avons mobilisé une analyse de contingence afin de calculer les corrélations entre les topics et les actes de GJ. La méthode est similaire à celle mobilisée pour les catégories de chaînes, mais nous avons représenté la distribution des topics avec des histogrammes de scores de saillance, plutôt qu'avec des matrices. On produit ainsi une visualisation de la façon dont les topics sont répartis sur les vidéos selon les actes, et inversement. Cette méthode nous permet de mesurer l'intensité avec laquelle un topic est traité selon les actes. Les visualisations suivantes comparent l'évolution des scores de saillance pour quatre topics à la fois. Nous y avons systématiquement ajouté un graphe d'évolution des volumes associés à ces topics, soit la distribution des volumes de vidéos selon les topics dans le temps, ce qui apporte une information complémentaire.

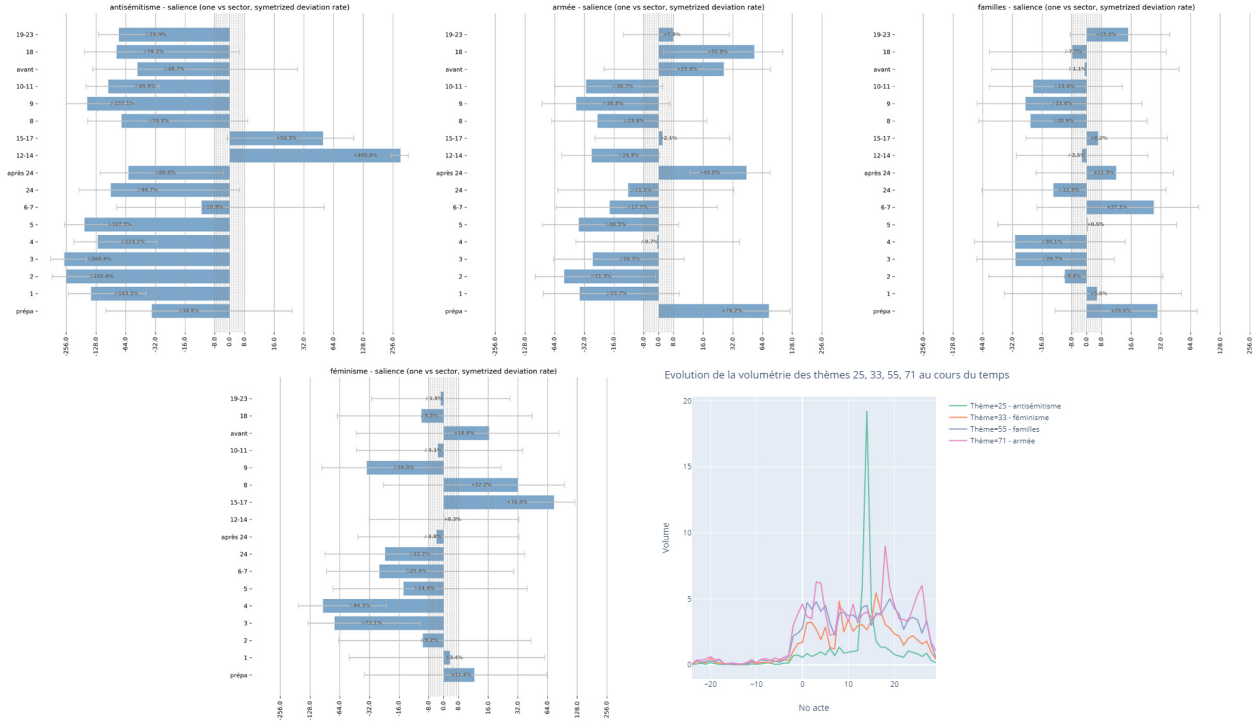



**Figure 7** – *Profil temporel selon les actes pour les topics (de haut en bas)* T25 Antisémitisme, T71 Armée, T55 Famille *et* T33 Féminisme*. La saillance du topic par acte est indiquée par un score de déviation symétrique négatif ou positif. Attention, les actes ne sont pas ordonnés dans le sens du temps sur les profils. En bas à droite, l'évolution de la volumétrie des quatre topics.*

Commençons par quatre topics relatifs aux manifestations (figure 7) – *T25 Antisémitisme*, *T71 Armée*, *T33 Féminisme* et *T55 Famille* – qui ont été les plus difficiles à interpréter parmi les topics de cette classe. *T25 Antisémitisme* se concentre autour de l'acte 15, suite aux injures antisémites dont a été victime le philosophe Alain Finkielkraut lors d'une manifestation. C'est donc un sujet épisodique, et non pas un thème structurant du mouvement. Le volume de ce topic est quasi nul en dehors de cette période et atteint un score de saillance important et une pointe exceptionnelle lors de cet évènement (c'est la pointe de loin la plus haute par rapport aux autres topics). *T71 Armée* quant à lui est plus constant dans le temps, avec une légère surreprésentation sur la période antérieure au mouvement (produit par une poignée de vidéos hétérogènes sans intérêt pour notre étude) et au moment de l'acte 18, cette dernière surreprésentation s'expliquant par l'annonce gouvernementale de la mobilisation des militaires pour protéger les bâtiments officiels lors des manifestations. Quant à *T33 Féminisme*, il présente un profil à plusieurs saillances. En amont du mouvement et lors du 1er acte, le topic est surreprésenté en raison du recouvrement des Gilets jaunes avec le mouvement « NousToutes ». Le thème réapparaît à partir de l'acte 8 suite à une manifestation dominicale de femmes Gilets jaunes. La pointe autour de l'acte 15 correspond aux discussions qui ont suivi la journée de la femme du 8 mars. Sur la période étudiée, l'évolution du volume de ce thème est plus ou moins croissante jusqu'au 8 mars, puis diminue avec l'essoufflement progressif des manifestations.

11. Pour un cas d'usage intéressant des topics models couplés à une analyse dynamique de mise à l'agenda médiatique, consulter (Pinto et al., 2019)

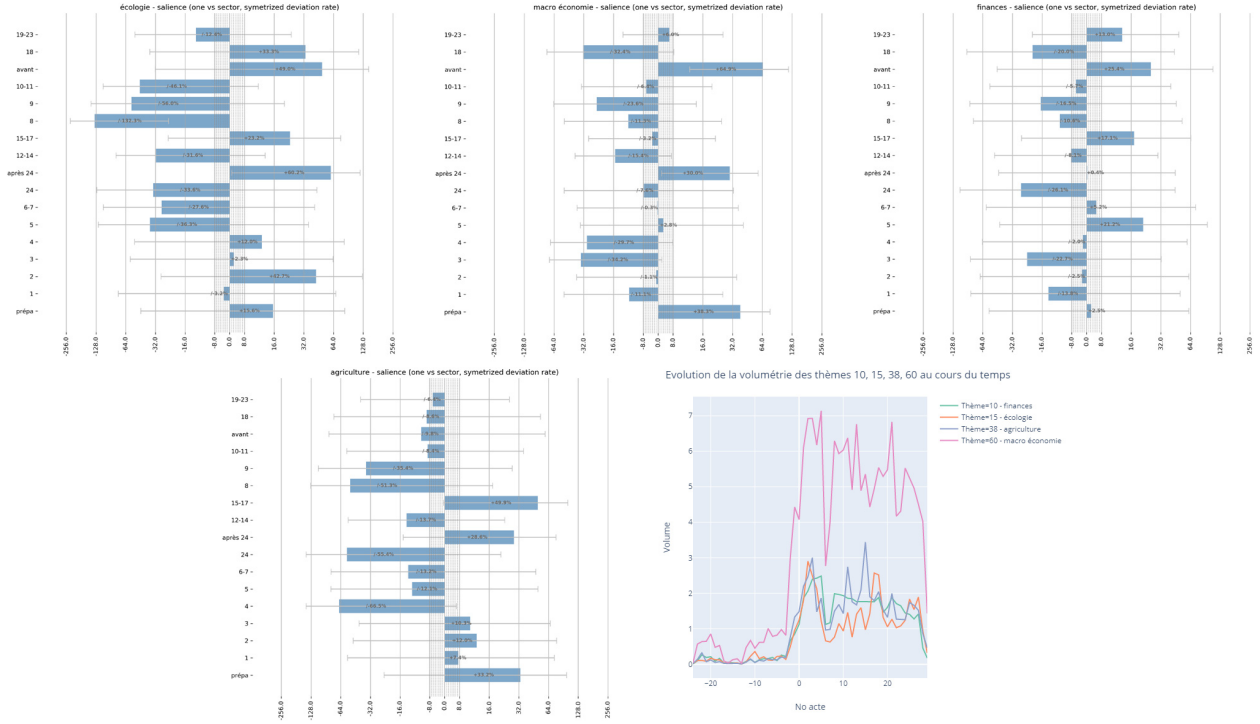



**Figure 8** – *Profil temporel selon les actes pour les topics (de haut en bas)* T15 Écologie, T60 Macro-économie, T10 Finance, T38 Agriculture*. La saillance du topic par acte est indiquée par un score de déviation symétrique négatif ou positif. Attention, les actes ne sont pas ordonnés dans le sens du temps sur les profils. En bas à droite, évolution de la volumétrie des quatre topics.*

Qu'en est-il des topics objet de l'action publique *T60 Macro-économie*, *T15 Écologie*, *T38 Agriculture* et *T10 Finance* (figure 8) ? Ont-ils des profils temporels moins épisodiques que ceux que nous venons de commenter ? Commençons par *T60 Macro-économie* qui apparaît en surreprésentation en amont, puis après l'acte 24. Cette surreprésentation ne signifie pas une absence de ce topic en dehors de ces deux périodes. En effet, nous observons que *T60 Macro-économie* est un topic représentant un volume de vidéos important et constant dès l'acte 1. La macro-économie est donc un topic quasi permanent dans l'espace public autour du mouvement, mais il est bien plus saillant dans les périodes où l'intensité du traitement médiatique du mouvement baisse. En revanche, *T15 Écologie* atteint ses scores de saillance maximum avant le mouvement et après l'acte 24, ce qui indique un certain effet du mouvement des Gilets jaunes sur l'intensité et la nature du débat autour de l'écologie. La saillance de *T15 Écologie* est positive autour des actes 2, 3 et 4 en rapport à la question automobile et du carburant, puis ressurgit sur la période allant de l'acte 15 à l'acte 18, lorsque l'écologie est rediscutée au moment de la clôture du Grand Débat. *T15 Écologie* renvoie à un volume de vidéos relativement faible. Le thème est de fait quasiment absent dans les phases de saillance négative. On peut considérer que l'écologie reste dans l'ensemble un thème peu structurant, porté à des moments spécifiques par les chaînes de vulgarisation, les médias alternatifs et les parties politiques (cf. la partie précédente). Enfin, notons que la période qui correspond aux dernières semaines avant la clôture du Grand Débat (actes 15, 16 et 17) est caractérisée par une saillance forte des trois topics, *Écologie*, *Agriculture* et *Finance*. Cette saillance peut s'expliquer par l'intensification des débats de fond due au Grand Débat.

Les quatre autres topics objets de l'action publique (figure 9) qui ont été aux origines du mouvement – *T04 Taxation Carburants*, *T63 Pouvoir d'Achat*, *T30 Revenus et Salaires* et, dans une moindre mesure, *T36 Retraites* – ont des profils temporels plus contrastés.

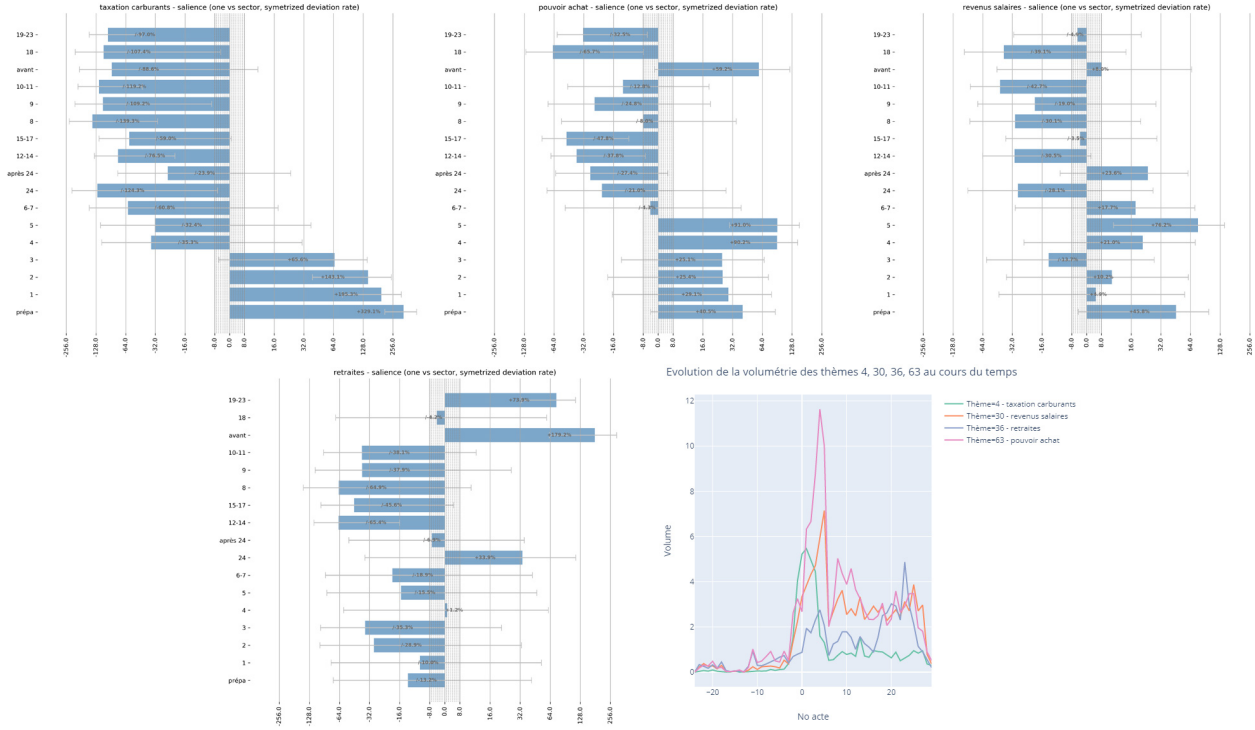



**Figure 9** – *Profil temporel selon les actes pour les topics (de haut en bas)* T04 Taxation Carburants*,* T63 Pouvoir d'achat*,* T30 Revenus et Salaires *et* T36 Retraites*. La saillance du topic par acte est indiquée par un score de déviation symétrique négatif ou positif. Attention, les actes ne sont pas ordonnés dans le sens du temps sur les profils. En bas à droite, évolution de la volumétrie des quatre topics.*

En effet, *T04 Taxation Carburants* a cessé d'être caractéristique des débats sur les Gilets jaunes dès le troisième acte, *T63 Pouvoir d'Achat* à partir du cinquième et *T30 Revenus et Salaires* à partir du septième. Notons aussi que, durant les périodes de saillance négative, ces trois topics restent néanmoins beaucoup plus importants sur le plan volumétrique que le topic de l'écologie (*T15*). Enfin, *T36 Retraites* est particulièrement surreprésenté avant le mouvement. Bien que présent durant les premiers actes, il occupe une place moins importante par rapport aux autres thèmes fondateurs du mouvement. La question des retraites prend une place conséquente autour de l'acte 19 à l'occasion de la clôture du Grand débat. D'une manière globale, le volume de *T36 Retraites* est quasi équivalent à celui de *T15 Écologie* en dehors de cette période de saillance. On peut observer ainsi une présence relativement importante de ce sujet dans le débat public, avant l'arrivée des contestations autour des projets de réforme du système de retraite.

Pour terminer notre exploration, observons les topics institutionnels (figure 10), en particulier, *T41 Peuple et Pouvoir*, *T48 Démocratie*, *T16 Débat public* et *T37 Grand Débat 2*. Soulignons d'emblée que ces thèmes sont absents dans la période qui précède les actes, ce qui signifie que le débat sur le débat public reste globalement un sujet marginal de l'agenda médiatique.
Un autre point manifeste de ces quatre profils temporels, c'est la saillance des topics décalée d'un acte : d'abord *T41 Peuple et Pouvoir* qui est le topic le plus contestataire, et qui émerge puis disparait progressivement entre l'acte 2 et l'acte 10 ; ensuite vient *T48 Démocratie*, d'une nature conceptuelle plus élevée, qui démarre autour de l'acte 5 et s'arrête aussi à l'acte 10 ; puis *T16 Débat Public*, qui décrit le déroulé des consultations d'Emmanuel Macron, débute autour des actes 6 et 7 et s'estompe à l'acte 18 ; et enfin *T37 Grand Débat 2*, à partir de l'acte 8 jusqu'à l'acte 23, qui renvoie à des éléments plus techniques de fonctionnement de la plateforme participative et de contenu du débat.

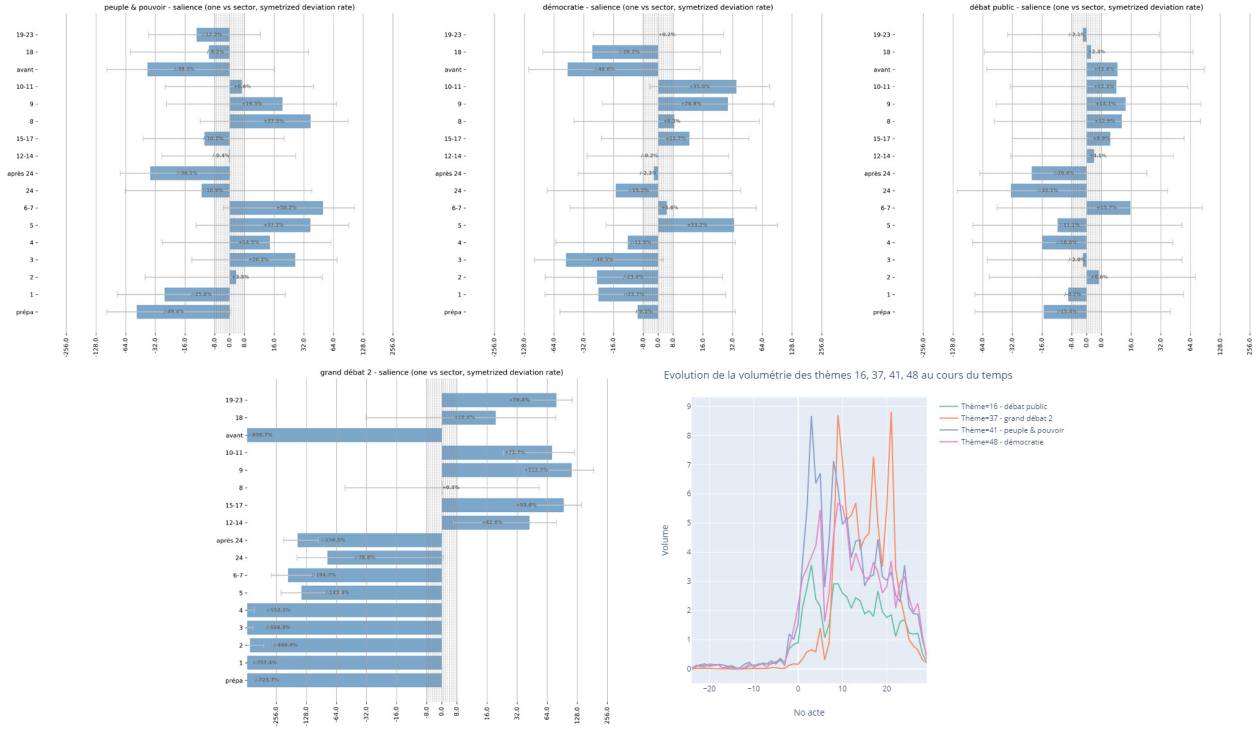



**Figure 10** – *Profil temporel selon les actes pour les topics (de haut en bas)* T41 Peuple et Pouvoir, T48 Démocratie, T16 Débat Public *et* T37 Grand Débat*. La saillance d'un topic par acte est indiquée par un score de déviation symétrique négatif ou positif. Attention, les actes ne sont pas ordonnés dans le sens du temps sur les profils. En bas à droite, évolution de la volumétrie des quatre topics.*

Si ces quatre topics traitent de questions de citoyenneté, ils ont chacun un statut différent : alors que *T41 Peuple et Pouvoir* et *T48 Démocratie* correspondent à des discussions critiques, voire conceptuelles, autour du système politique et de la démocratie représentative, les topics *T16 Débat Public* et *T37 Grand Débat 2* portent sur le déroulement de l'initiative participative mise en place par le gouvernement. Le « Grand Débat » aurait-il eu un effet progressivement annihilateur sur les topics critiques du système politique et d'une manière plus générale, un effet de cadrage sur l'espace médiatique des Gilets jaunes ? C'est sans doute une des questions majeures soulevées par cette exploration, et qu'il nous faut entre autres discuter.

## 4. Discussion et conclusion

Telles sont donc les informations que le dendrogramme, le réseau, la matrice de corrélation et les profils temporels nous ont données à voir. Pour intéressant que soient ces résultats, ont-ils apporté des connaissances nouvelles sur le traitement médiatique des Gilets jaunes ? La réponse est évidemment mitigée. Les 72 topics, et les quatre manières de les représenter, ont pu corroborer ce que les spécialistes des médias ont déjà montré sur le traitement médiatique des mouvements sociaux en général, à savoir une surreprésentation de la violence contestataire et spectaculaire dans les médias traditionnels (Baisnée *et al.*, 2021) ; des médias alternatifs proposant des informations différentes de celles des médias dominants, mais dépendantes de l'agenda militant. Jusque-là, rien de nouveau dans notre étude. Mais de même que l'analyse par les *topics models* corrobore par certains aspects les études existantes sur les rapports entre médias, plateformes numériques et mouvements sociaux, on peut dire que réciproquement ces recherches prouvent la fiabilité des *topic models* et la pertinence de partir de YouTube comme terrain d'enquête. L'un des apports de cet article est d'avoir montré que les sous-titres des vidéos de YouTube forment un matériau sur lequel on peut faire enquête pour analyser l'espace médiatique.

Cette exploration apporte néanmoins quelques apports originaux sur l'étude des rapports des Gilets jaunes aux médias. Tout d'abord, nous montrons que si les profils thématiques des chaînes Gilets jaunes et de contre-information ont quelques points communs, cela ne porte pas pour autant sur des topics conspirationnistes. De plus, ces deux catégories de chaînes présentent aussi des divergences importantes, notamment l'intérêt spécifique des chaînes Gilets jaunes pour les sujets de citoyenneté, un sujet rarement traité par les chaînes de contre-information. Ces résultats sont de fait cohérents avec ceux issus du *YELLOWPOL Project* du médialab de Sciences Po. En effet, à partir d'un corpus massif de discussions extraites des pages Facebook de différents groupes Gilets jaunes, (Froio *et al.*, 2020) montrent la primauté des questions de citoyenneté par rapport à celles plus classiquement abordées par l'extrême droite. Ainsi, nos résultats invitent à nuancer les analyses des journalistes qui ont cadré l'analyse médiatique du mouvement des Gilets jaunes autour des accointances douteuses des acteurs du mouvement parmi les médias conspirationnistes et xénophobes (voir par exemple les dénonciations de Bornstein (2019)). Cette nuance est d'autant plus importante que la plus grande proximité thématique des chaînes de contre-information se situe du côté des médias *mainstream*. Cette observation mérite des explorations plus approfondies sur les rapports d'interdépendance de ces deux dernières catégories de chaînes dans les situations de crise.

Enfin, s'il y a une autre nouveauté dans cet article au regard des travaux existants sur le traitement médiatique des Gilets jaunes (Souillard *et al.*, 2020), elle se situe au niveau des topics des objets de l'action publique et des topics institutionnels. Pour les topics relatifs aux objets de l'action publique, on peut retenir trois résultats importants : ces topics sont caractéristiques des chaînes de médiation, ce qui signifie que la vulgarisation politique a joué un rôle clef dans le traitement médiatique du mouvement en y apportant des éléments de fond ; l'écologie et la macro-économie sont déconnectés et l'écologie est restée marginale relativement à la macro-économie ; et si les thèmes fondateurs du mouvement (pouvoir d'achat, taxation des carburants, revenus et salaires, et retraite) ont diminué progressivement en intensité, ils occupent un volume largement supérieur à celui de l'écologie tout le long de la période étudiée.

Au sujet des topics institutionnels, il nous semble particulièrement important de retenir les topics autour des rapports des citoyens à leur système politique. Les Gilets jaunes se sont en grande partie mobilisés autour de revendications sur la démocratie. Ces revendications, visibles dans notre corpus, ont aussi évolué au fil des actes. Nous l'avons montré, le Grand Débat ne s'est pas imposé comme un simple topic supplémentaire : il a modifié la structure de l'espace médiatique en se substituant aux discours critiques à l'encontre de la démocratie représentative. S'il y a une réussite du gouvernement dans la gestion du mouvement des Gilets jaunes, c'est celle d'avoir recadré l'espace médiatique en créant un espace de dialogue. Tout porte à croire que le principal effet politique du Grand Débat est médiatique : en créant ce dispositif, le gouvernement est parvenu, sans doute sans l'avoir planifié, à avoir une prise sur l'espace médiatique et l'agenda qui lui est associé[12].

Ainsi, cet article ouvre une perspective novatrice : explorer YouTube à partir d'une analyse non supervisée peut être un moyen puissant de totalisation statistique pour faire émerger un point de vue général sur l'espace médiatique. Mais si la production de contenus médiatiques, autour d'un mouvement social notamment, est un enjeu de quantification, elle est rarement appréhendée comme un objet de statistique. Pourtant, l'analyse de la production médiatique procède d'une forme de « statactivisme » (Bruno and Didier, 2014). Les méthodes de l'analyse quantitative de contenu qui classent et comptent la production médiatique ont été mobilisées durant le mouvement des Gilets jaunes comme pratiques statistiques pour critiquer et s'émanciper de la réalité médiatique construite par les médias.

12. On pourrait aussi prêter au gouvernement – mais nous n'en avons pas la preuve – un certain machiavélisme, en transposant la célèbre formule de Clémenceau : « Si vous voulez enterrer un problème, nommez [non plus] une commission », mais en l'occurrence, dans notre cas, « un grand débat » national.

Il existe donc deux réalités médiatiques[13] : celle produite par les médias en enquêtant sur le monde social et celle produite par l'analyse de contenu qui en mesurant la production des médias fige une « représentation critique du traitement médiatique ».

C'est cette deuxième réalité médiatique que nous avons construite dans cet article. De ce point de vue, l'analyse quantitative de contenu médiatique par les *topics models* peut être discutée du point de vue de la sociologie de la quantification : l'analyse des topics revient à mettre en place une nouvelle construction via le déploiement d'une infrastructure de classification des contenus. Si cette réalité médiatique (la représentation critique du traitement médiatique) n'est pas accessible directement, mais par l'intermédiaire d'algorithmes, il devient urgent de faire de l'analyse quantitative de contenu un instrument officiel, systématique et harmonisé pour faire de l'espace médiatique un objet de commune mesure. YouTube et les *topics models* sont de bons candidats pour construire cet espace d'équivalence et le langage commun permettant de débattre des agendas médiatiques et de leurs effets sur l'espace public.

## Références

Arora S., R. Ge, A. Moitra (2012), « Learning topic models – going beyond SVD », *IEEE 53rd annual symposium on foundations of computer science*, pp. 1-10.

Baisnée O., A. Cavée, C. Gousset et J. Nollet (2021), « La "violence" des Gilets jaunes : quand la fait-diversification fait diversion. Les routines journalistiques à l'épreuve des manifestations à Toulouse (novembre 2018-juin 2019) », *Sur le journalisme* (à paraître).

Blei D. (2012), « Probabilistic Topic Models », *Communications of the ACM*, vol. 55, n° 4, pp. 77-84.

Boltanski L. (2009), *De la critique : précis de sociologie de l'émancipation*, Paris, Gallimard.

Bornstein R. (2019), « En immersion numérique avec les "gilets jaunes" », *Le Débat*, n° 204, pp. 38-51.

Bruno I. et E. Didier (2014), *Statactivisme,* Paris, La Découverte.

Cardon D. et F. Granjon (2010), *Médiactivistes*, Paris, Presses de Sciences Po.

Cointet J.-P. et S. Parasie (2018), « Ce que le big data fait à l'analyse sociologique des textes », *Revue française de sociologie*, vol. 59, n° 3, pp. 533-557.

DiMaggio P., M. Nag, and D. Blei (2013), « Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding », *Poetics*, vol. 41, n° 6, pp. 570-606.

Evans J. A. and P. Aceves (2016), « Machine Translation: Mining Text for Social Theory », *Annual Review of Sociology*, vol. 42, pp. 21-50.

Ferron B. (2019), « Mouvements sociaux : le jeu médiatique en vaut-il la chandelle ? », *The Conversation*, http://theconversation.com/mouvements-sociaux-le-jeu-mediatique-en-vaut-il-la-chandelle-128139 (consulté le 10/02/2021).

13. En s'inspirant de Boltanski, on distingue ici le « monde médiatique », que l'on peut saisir, et la « réalité médiatique », construite par l'analyste qui se donne pour mission d'explorer le traitement médiatique (Boltanski, 2009).

Froio C., P. R. Morales, J.-Ph. Cointet, and O. F. Metin (2020), « It's not radical right populism! The Yellow Vests in France », UiO: C-REX-Center for Research on Extremism, https://www.sv.uio.no/c-rex/english/news-and-events/right-now/2020/its-not-radical-right-populism.html (consulté le 10/02/2021).

Granjon F. (2014a), *Médias dominants, mouvements sociaux et mobilisations informationnelles. Histoire des mouvements sociaux en France*, Paris, La Découverte.

Granjon F. (2014b), « Citoyenneté, médias et TIC », *Réseaux*, vol. 2-3, n° 184-185, pp. 95-124.

Granjon F. (2018), « Mouvements sociaux, espaces publics et usages d'internet », *Pouvoirs*, vol. 1, n° 164, pp. 31-47.

Greene D. and J. P. Cross (2017), « Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach », *Political Analysis*, vol. 25, n° 1, pp. 77-94.

Greene D, D. O'Callaghan, and P. Cunningham (2014), « How Many Topics? Stability Analysis for Topic Models », *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*, vol. 8724, pp. 498-513.

Le Bart Ch. (2020), *Petite sociologie des Gilets jaunes. La contestation en mode post-institutionnel*, Rennes, Presses Universitaires de Rennes.

Lee D. and H. S. Seung (1999), « Learning the parts of objects by non-negative matrix factorization », *Nature*, vol. 401, n° 6755, pp. 788-791.

Lindstedt N. C. (2019), « Structural Topic Modeling For Social Scientists: A Brief Case Study with Social Movement Studies Literature (2005-2017) », *Social Currents*, vol. 6, n° 4, pp. 307-318.

McCombs M. E. and D. L. Shaw (1972), « The Agenda-Setting Function of Mass Media », *Public Opinion Quarterly*, vol. 36, n° 2, pp. 176-187.

Mimno D., H. M. Wallach, E. Talley, M. Leenders, and A. McCallum (2011), « Optimizing semantic coherence in topic models », *Proceedings of the 2011 conference on empirical methods in natural language processing*, pp. 262-272.

Moretti F. (2016), *La Littérature au laboratoire*, Paris, Les Éditions d'Ithaque.

Pinto S., F. Albanese, C. Dorso, and P. Balenzuela (2019), « Quantifying time-dependent Media Agenda and public opinion by topic modeling », *Physica A: Statistical Mechanics and its Applications*, vol. 524, pp. 614-624.

Poels G. et V. Lefort (2019), « "Gilets jaunes" : une médiatisation d'une ampleur inédite », La Revue des Médias, http://larevuedesmedias.ina.fr/gilets-jaunes-mediatisation-chaines-info-twitter (consulté le 10/02/2021).

Röder M., A. Both, and A. Hinneburg (2015), « Exploring the space of topic coherence measures », *Proceedings of the eighth ACM international conference on web search and data mining*, pp. 399-408.

Sebbah B., L. Loubère, N. Souillard, L. Thiong-Kay et N. Smyrnaios (2018), « Les Gilets jaunes se font une place dans les médias et l'agenda politique », Rapport de recherche du LERASS, https://hal-amu.archives-ouvertes.fr/hal-02120478 (consulté le 10/02/2021).

Sebbah B., L. Loubère, N. Souillard, J. Renard et N. Smyrnaios (2019), « La dilution des Gilets jaunes dans l'agenda médiatique et politique », Rapport de recherche du LERASS, https://www.histoiredesmedias.com/Etude-La-dilution-des-Gilets.html (consulté le 10/02/2021).

Tsur O., D. Calacci, D. Lazer (2015), « A frame of mind: using statistical models for detection of framing and agenda setting campaigns », *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, vol. 1, pp. 1629-1638.

Wesslen R. (2018), « Computer-Assisted Text Analysis for Social Science: Topic Models and Beyond », arXiv, https://arxiv.org/abs/1803.11045 (consulté le 10/02/2021).