

**LE CADRAN DE LA VISIBILITE DE LA SPHERE
MEDIATIQUE ET POLITIQUE SUR YOUTUBE**

Les effets de la recommandation algorithmique et manuelle

Bilel BENBOUZID
Emma GAUTHIER
Pedro RAMACIOTTI MORALES
Bertrand ROUDIER
Tommaso VENTURINI

Introduction

Ces dernières années ont été marquées par de très fortes controverses sur le système de recommandation de YouTube. Alors que du côté de la presse traditionnelle, on accuse l'algorithme d'être favorable aux contenus litigieux (désinformation, discours de haine, etc.), du côté des créateurs, on s'indigne, au contraire, d'une machine inégalitaire concentrant la recommandation sur une minorité de contenus qui présente un intérêt économique pour YouTube, au détriment de vidéos alternatives à caractère politique, au contenu original et souvent « amateur ».

Nous avons saisi l'occasion de ce débat pour mener une enquête quantitative sur ce qui se passe sur YouTube en matière de production et de recommandation de l'information. Quelle est la composition de l'espace médiatique sur YouTube ? Comment le système de recommandation d'une des plateformes les plus visités au niveau mondial diffuse, distribue et ordonne l'information ? Dans quelle mesure YouTube permet-t-il d'élargir la participation au débat public en rendant visible du contenu politique alternatif ?

Ces questions sont cruciales pour la compréhension plus large de la façon dont YouTube transforme la structure de l'industrie des médias. Nous nous concentrerons dans ce travail sur le cas du YouTube français, à travers l'analyse de la visibilité des différentes chaînes qui forment l'espace médiatique et politique sur YouTube. À partir de l'étude d'un échantillon quantitatif de chaînes aux contenus disparates, produits par des professionnels des médias, de la communication, des vidéastes amateurs ou pro amateurs dont le contenu médiatique est natif de YouTube, nous présentons ici des éléments sur la façon dont se structure la visibilité médiatique selon des niveaux d'engagement différents.

Nous présentons dans premier temps le matériel collecté, le travail de catégorisation des chaînes réalisé manuellement et la méthode utilisée pour dégager la manière dont se distribue la visibilité dans la sphère

médiatique et politique sur YouTube. Dans un second temps, nous analysons la structuration du réseau de recommandation algorithmique en observant la distribution des chaînes. Nous montrons ensuite que la visibilité sur YouTube se distribue selon quatre vitesses différentes : (1) l'hyper présence par la surproduction, (2) l'assimilation des codes de la culture YouTube, (3) l'entre soi des publics fan et (4) l'invisibilité totale des laissés pour compte de la révolution numérique. Nous discutons enfin comment cet espace public à quatre vitesses reflète une politique de modération du contenu opérée par YouTube suite aux critiques adressées à l'algorithme de recommandation, et ce que cela implique en termes de recomposition de l'espace public.

Popularité et radicalisation : la part de la recommandation algorithmique

Dans le domaine des études sur YouTube, on trouve deux approches différentes traitant du système de recommandation : d'une part, l'analyse de la contribution de la recommandation à la *popularité* des vidéos et, d'autre part, l'audit des mécanismes de filtrage algorithmique des contenus recommandés aux usagers, notamment en matière de radicalisation politique.

Popularité et recommandation algorithmique

La première approche observe la popularité (généralement caractérisée par le nombre de vues) comme un enjeu de modélisation statistique. Sur YouTube, la popularité présente une structure hétérogène : un petit nombre de vidéos concentre l'essentiel des vues. En 2007, l'une des premières études de la popularité sur YouTube montrent que la distribution des vues suit *une loi de puissance tronquée*, c'est-à-dire une accélération de la décroissance de la traîne, à partir d'un certain seuil (Cha et al., 2007). Cette coupure indique l'invisibilité d'un grand nombre de contenus qui pourrait s'expliquer par le délaissement par

système de recommandation. Il est donc envisagé d'agir sur le système de recommandation (avec un gain escompté de 45% du nombre de vues globale). Ainsi, lors des premières années de développement de la plateforme, le système de recommandation est d'abord envisagé comme un moyen d'augmenter l'audience globale et sauvant des oubliettes le plus possible de contenus marginaux.

Zhou et al. (2016) précisent ce type d'analyse en mobilisant les *tracked views* fournies par YouTube (les dix premières sources à l'origine des vues pour chaque vidéo, une information qui n'est plus disponible aujourd'hui) dans la perspective de mesurer la contribution du système de recommandation à la visibilité. Les auteurs concluent que la recommandation algorithmique est non seulement la source de visibilité la plus importante sur YouTube (soit 40-60% des vues, un score qui atteint désormais 70% selon les annonces officielles de la plateforme), mais aussi la moins clivant. Selon Zhou et al., la recommandation est même la source principale de vues pour les vidéos impopulaires. Ainsi, dans cette étude le système de recommandation apparaît comme un outil pour concentration de la popularité produite par les dynamiques humaines.

D'autres études ont analysé les rapports entre les systèmes de recommandation et les dynamiques des vues au niveau des vidéos à travers la modélisation de séries temporelles (Crane and Sornette, 2008). Figueiredo et ses collaborateurs (Figueiredo et al., 2014) ont, par exemple, comparé un échantillon de vidéos les plus populaires sur YouTube et un échantillon de vidéos aléatoire. Pour les vidéos de l'échantillon de vidéos populaires, les vues provenant de la recommandation algorithmique sont plus importantes. En revanche, pour le corpus de vidéos constitué aléatoirement, c'est l'algorithme de requête qui génère la plus grande part des vues. A l'issue de cette étude, il semble difficile d'associer l'algorithme de recommandation à un type particulier de signature. En général, les travaux qui ont recours aux *tracked views*, les données les plus granulaires, ne sont pas en mesure

de prouver un effet spécifique de la recommandation algorithmique sur les dynamiques de popularité.

Bulle de filtre et radicalisation

La deuxième approche se concentre sur l'analyse des effets du système de recommandation dans le contexte du débat sur les chambres d'écho médiatiques et les bulles de filtre qui leur sont associées. Dans une analyse théorique (qui repose sur des simulations), des chercheurs de Deepmind (Jiang et al., 2019) – donc étroitement liés à Google – ont montré comment les algorithmes de recommandation peuvent accélérer ou ralentir ces deux phénomènes différenciés : les chambres d'écho médiatiques renvoient à la polarisation des publics exposés de manière répétée à des contenus similaires, dans le cadre d'échange communautaires facilité par les réseaux sociaux ; les bulles de filtre quant à eux, sont générées par les algorithmes de recommandation qui maintiennent et accentuent cette exposition limitée à des opinions politiques similaires. En démêlant la chambre d'écho de l'effet de bulle de filtre, les chercheurs de Google montrent que plus le système de recommandation prédit les centres d'intérêt de l'utilisateur avec précision, plus il l'isole rapidement dans une bulle de filtre. Dans un contexte où le système de recommandation cherche à maximiser le temps de visionnage des vidéos (considéré comme une métrique de la qualité de la recommandation), la boucle de rétroaction entre l'utilisateur et le système de recommandation produit un rétrécissement inéluctable de la fenêtre d'exposition des vidéos. Pour limiter l'effet bulle de filtre, il faudrait recommander, selon les auteurs, des contenus qui ont moins de chance de susciter l'intérêt de l'utilisateur, mais ce serait aussi prendre le risque qu'ils quittent la plateforme.

Mais si cette recherche montre que les data scientists de Google (de YouTube aussi, donc) prennent au sérieux ce problème, il n'est pas possible de la transposer au cas spécifique de YouTube - il s'agit de simulations fondées sur des situations simplifiées et désincarnées. Lorsque les études traitent de situations réelles, généralement de la

sphère médiatique et politique sur YouTube, les conclusions sont plus incertaines, voire contradictoires. En effet, les rares observations quantitatives sur les mécanismes de recommandation des contenus extrémistes, reposant sur des catégorisations de chaînes produites à partir d'analyse de contenu, ne convergent pas. D'un côté, l'une des premières études réalisées par O'Callaghan et al. (2015) et al. montre que les utilisateurs qui accèdent à des vidéos d'extrême droite sur YouTube ont une probabilité élevée de se voir recommander d'autres contenus d'extrême droite, ce qui conduirait à une immersion dans une bulle idéologique en quelques clics seulement.

D'un autre côté, les recherches les plus récentes arrivent à des résultats moins tranchés sur les effets de « radicalisation » causés par la recommandation de l'algorithme de YouTube. En analysant les migrations des commentateurs entre les différents groupes de chaînes (les intellectuels réactionnaires – *intellectual dark web* ; les chaînes d'extrêmes droites moins radicales – alt-lite ; et les milieux d'extrême droite les plus radicaux – alt-right), Ribeiro et al. (Ribeiro et al., 2019) montrent un effet de radicalisation des utilisateurs sur YouTube, mais sans observer de tendance similaire en matière de recommandation algorithmique. Plus encore, Ledwich et Zaitsev (2019) ont montré que l'algorithme de YouTube sous-recommanderait les contenus extrémistes, favorisant les médias traditionnels, voire même les chaînes de gauche ou politiquement neutres. Il est probable, mais impossible à prouver faute de données accessibles, que cette divergence de résultats tiennent moins à des différences de méthode entre les études qu'à une évolution récente de l'algorithme de recommandation qui suit un changement du positionnement de YouTube dans le système médiatique. Alors que dans ses premières années, un des principaux objectifs de la plateforme était d'étendre son audience en attirant des publics marginaux mal desservi par l'offre de la télévision généraliste ; YouTube viserait aujourd'hui à consolider sa position dominante en prenant une orientation plus *mainstream*.

Matériel et Méthode

Comment représenter l'espace médiatique et politique sur YouTube ?

Dans cette enquête, nous avons souhaité représenter l'espace médiatique et politique produit sur YouTube pour la France sans limiter notre corpus aux seules chaînes de production d'information journalistique. 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. Nous avons donc étendu notre corpus à tous les types de chaînes produisant des opinions, des analyses et des décryptages et participent ainsi au débat public français ; 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 assurer à une telle hétérogénéité, nous avons construit des listes de chaînes *a priori* selon des types d'acteurs dont il nous semblait important de 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 parvenu à dix types d'acteurs dont nous avons cherché la présence sur YouTube. Pour chacune de ces types, nous avons manuellement dressé des listes de chaînes. Une fois agrégées, ces chaînes nous ont alors permis d'étendre notre corpus en profitant de l'effet boule de neige du réseau de recommandation entre chaînes. Par exemple, les chaînes régionales de France Télévision sont toutes liées les unes aux autres ou encore les chaînes de vulgarisation scientifique ont tendance à former un réseau

de chaînes « amies ». Nous avons répété plusieurs fois cette opération qui s'apparente à une méthode par chaînage. Nous avons opéré plusieurs vagues de collecte en ajoutant systématiquement les nouvelles chaînes. Au fil du *crawling*, l'apport en chaînes nouvelles correspondant au thème défini s'est progressivement raréfié, jusqu'à s'épuiser aux alentours de 800 chaînes.

Cette procédure de collecte reste néanmoins biaisée envers les individus les plus coopératifs sur YouTube. De nombreuses chaînes n'indiquent aucune chaîne amie et ne sont pas recommandées par d'autres. Pour limiter ce biais, nous avons cherché de nouvelles chaînes dans la base de données de Wizedéo (une entreprise spécialisée dans le marketing YouTube) en consultant les listes de chaînes similaires (c'est-à-dire ayant des commentateurs communs) associées à chacune de nos chaînes. Cette procédure nous a permis d'atteindre un effectif total de 1.400 chaînes.

Mais de quoi cette liste de 1.400 chaînes est-elle la représentation ? Il ne s'agit pas d'un échantillon parfaitement représentatif. En effet, si pour certaines catégories il est apparu simple de constituer une liste quasi-exhaustive (chaînes des députés, des ministères, des médias traditionnels), la limite du corpus est beaucoup plus difficile à arrêter lorsqu'il s'agit de petites associations militantes ou de vlogueurs individuels publiant des vidéos pour faire valoir leur opinion sur l'actualité. La traîne de la popularité sur YouTube étant particulièrement longue, nous avons décidé de critères pour limiter le corpus : les chaînes doivent atteindre au moins 500 abonnés pour être intégrées au corpus, être actives durant la période de collecte (les 6 derniers mois) et contribuer à des discussions dans le débat public. Les chaînes de moins de 500 abonnés intégrées au corpus correspondent à des institutions de grande notoriété. Bien que suscitant très peu d'intérêt sur YouTube leur intégration au corpus est déterminante pour comprendre les configurations renouvelées de l'espace public via YouTube

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. Si des chaînes importantes et influentes peuvent être absentes du corpus et si des engagements politiques peuvent ne pas être représentés, cette liste permet néanmoins d'avoir un aperçu global du système YouTube. Ainsi, il faut interpréter notre corpus comme un échantillon non aléatoire et raisonné des chaînes qui diffusent du contenu à caractère médiatique et politique (problèmes sociaux, moraux, scientifiques, économiques, écologiques, historiques etc.) Dans cet espace, YouTube met tous les comptes sur un même plan : peu importe la nature du contenu ou le statut du diffuseur. Médias traditionnels, activistes, militants, organisations, vlogueurs anonymes, sur YouTube tous les comptes sont des chaînes.

Classer les chaînes

Ainsi, l'unité d'analyse privilégiée dans cette étude est la chaîne (plutôt que la vidéo). Pour rendre compte de cette population de chaînes, nous avons dans un premier temps considéré les catégories de chaînes proposées par YouTube. Malheureusement, celles-ci peu fiables et manquent de granularité (la catégorie actualité et politique correspond à la quasi-totalité de notre corpus). Nous avons donc cherché à rendre compte des différentes formes de chaînes de notre corpus, par un travail de catégorisation en deux étapes.

Usages et statuts des chaînes

Dans un premier temps, nous avons appréhendé les chaînes par deux critères distincts : celui format des vidéos diffusés ; et celui du statut du diffuseur, c'est-à-dire la situation de l'entité qui anime la chaîne dans la société. Ce deuxième critère (le statut) ne pose pas de problème particulier : nous avons catégorisé les chaînes selon six statuts :

individus, partis politiques, médias professionnels, médias YouTube, organisations privées, et acteurs publics (cf. le tableau 2).

La question de la catégorisation des usages est plus complexe. Nous nous sommes d'abord demandé si le contenu est produit exclusivement pour YouTube. En posant cette question on obtient deux groupes de chaînes facilement identifiables : celles natives de YouTube (produisant du contenu inédit) et celles dont le contenu partagé est préexistant à l'activité de la chaîne. Les chaînes natives de YouTube se partagent en deux sous-groupes : des chaînes de monologue face caméra (des vlogs) et d'autres types au contenu plus diversifié que nous rassemblons dans un même groupe. Les chaînes qui ne sont pas natives de YouTube et qui ne produisent pas de contenu inédit peuvent aussi être subdivisées en deux sous-groupes : du replay et du montage vidéos. Nous avons ensuite distingué dans ces deux sous-groupes ce qui relevait de la télévision ou de contenu provenant d'espace du web, YouTube y compris (cf. le tableau 1 pour les précisions).

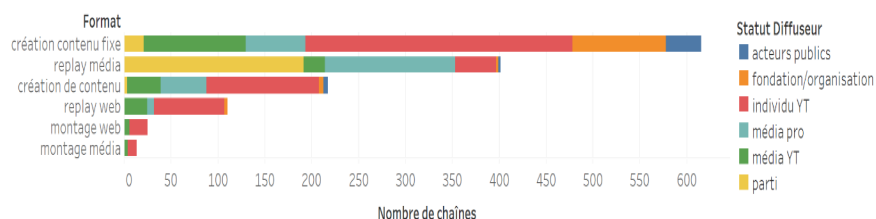


Figure 1 : Répartition des statuts des diffuseurs en fonction des formats

Ce sont les chaînes de « création de contenu fixe » qui représentent le plus grand effectif de notre corpus (45%) et qui obtiennent en moyenne le plus de vues (8.684 vues). Se mettre en scène face à la caméra, derrière un bureau, dans son jardin ou au volant de sa voiture est un style classique des « vidéastes » sur YouTube et est adapté à l'ensemble des genres (beauté, gaming, humour, décryptage, analyse, opinion etc...). Dans cette catégorie, nous avons aussi ajouté tous types de plan fixe ou en mouvement dans des décors qui ne changent pas. Ces plans filment des individus prenant la parole - de la conférence à la discussion experte autour d'une table, en passant par les plateaux de radio. Ce type

de format est le plus souvent produit par ce que nous avons appelé des individus YouTube (265 chaînes), mais c'est aussi le format privilégié des médias YouTube des organisations/fondations et des acteurs publics.

Les chaînes de replay brut (qui repostent du contenu diffusé à la télévision ou à la radio) représentent le deuxième plus grand effectif de notre corpus (28%), mais ce sont elles qui ont publié le plus de vidéos sur la période étudiée (131.475 vidéos). Si on additionne à ces chaînes de replay brut TV et radio, les chaînes de montage TV et radio, on observe que la télévision et la radio occupent près de 30% de l'espace politique et médiatique sur YouTube (en terme de chaînes). Si les médias professionnels (Arte, BFM, Sud Radio etc) utilisent le replay sur YouTube comme un moyen d'élargir leur audience, notons qu'il ne représente pas le type de diffuseur le plus important dans cette catégorie d'acteur. En rediffusant sur YouTube leurs discours ou interviews transmis dans les médias, les partis politiques et les militants associés sont ceux qui sont le plus représentés dans la catégorie replay TV ou radio. Quelques individus publient aussi du replay (44 chaînes), en rendant accessible des documentaires, des reportages, des émissions voire même le JT du 20h. En cherchant à attirer le maximum de passages d'internautes afin de générer des revenus publicitaires en ligne, ces chaînes captent une part non négligeable de l'attention. C'est pourquoi nous les avons intégrés au corpus. Notons que les comptes de ces chaînes font néanmoins régulièrement l'objet de clôture pour atteinte aux droits d'auteur.

La « création de contenu autre » arrive en troisième position (15%). Elle correspond à tous les autres types de contenu natif de YouTube. On trouve dans cette catégorie de nombreuses chaînes d'individus YouTube (120 chaînes) qui, la plupart du temps, réalisent des vidéos mobiles pour mettre l'accent sur un événement (le cas des vidéos de manifestation des gilets jaunes). Les médias YouTube diffusent quant à eux des reportages d'« enquête », d'analyse ou de contextes sociaux (26 chaînes).

Le « replay web » occupe la quatrième place avec 110 vidéos. Elles correspondent pour l'essentiel à des chaînes d'individu YouTube (75 chaînes) ou des médias YouTube qui sont pour l'essentiel des médias militants et alternatifs, souvent de journalisme amateur (des chaînes de gilets jaunes par exemple par le biais de *live* Facebook, plateforme sur laquelle ont été créés beaucoup de contenus natifs de la part des figures du mouvement). Les chaînes de montage web, qui occupent, une place marginale dans le corpus et sont essentiellement produites par des individus. Ces chaînes de montage de contenu web parviennent à accumuler près de 3 500 000 vues, ce qui néanmoins est relativement faible comparé aux 700 millions de vues cumulées par le Replay TV.

Formes d'expression et de communication

Si ce système de catégorisation à deux niveaux donne un premier aperçu du corpus, il lui manque une dimension sémantique. C'est pourquoi nous avons développé une troisième catégorisation permettant de mieux rendre compte de la pluralité des formes de communication politiques et médiatiques. Plutôt qu'une catégorisation par des critères objectifs explicite comme nous l'avons fait pour les usages et les statuts, nous avons procédé par induction à partir de chaînes qui nous semblaient exemplaires de certaine forme d'expression sur YouTube. Nous avons dû renoncer à une exigence de cohérence absolue afin d'éviter d'accumuler les partitions, ce qui aurait conduit d'une part à une multiplication des chaînes frontières susceptibles à elles seules de former des catégories et, d'autre part, à un accroissement de catégories aux effectifs excessivement faibles. Nous sommes progressivement arrivés à douze regroupements de chaînes exemplaires autour desquels nous avons classé manuellement les chaînes qui nous semblaient les plus proches.

La catégorie « Discours militants » est la plus conséquente de notre corpus (19%) et regroupe une grande diversité de chaînes. Ce sont les chaînes dont l'engagement politique ou la cause est défendue de manière explicite, officielle et légitimée, par des partis politiques, des mouvements alternatifs et personnes engagées. Dans cette catégorie se

retrouvent alors Jean-Luc Mélenchon, l'association de lutte contre les violences faites aux animaux L214, le Groupe Républicain de l'Assemblée Nationale. Nous avons inclus tous les députés présents sur YouTube dans cette catégorie, même si leur nombre d'abonnés est très souvent faible (moins d'une centaine d'abonnés pour la plupart des députés).

La deuxième catégorie, « Contenus litigieux » (18%), nommée « Complotiste, confusionniste, pamphlétiste » sur les graphes, représente une grande diversité de contenus et de moyens investis. Ici on trouve des chaînes « pièges à clic » qui diffusent des fausses informations à tendance complotistes qui mettent l'accent sur le rôle principal de la manipulation cachée par quelques personnes puissantes ou des groupes en particulier (les francs-maçons, notamment). On y trouve aussi des chaînes polémistes qui appartiennent à la sphère du mouvement « dissident », entretenant la confusion quant à leurs appartenances politiques extrémistes, par exemple Dieudonné, le réseau Egalités et Réconciliation de Rhône Alpes (ERTV Rhône Alpes), Panamza, VoltaireNetTV, L'Heure de se Réveiller, Teddy Boy RSA, Le Raptor, autant de chaînes qui traitent l'actualité politique par le biais de la polémique. Ces chaînes ont pour point commun de diffuser un discours flou, parfois belliqueux et insultant. Nous avons aussi intégré dans cette catégorie des chaînes professionnelles comme Russia Today, Sputnik et CCTV. Si ces chaînes ont des moyens professionnels de média web mainstream et produisent une information de qualité équivalente à celle des chaînes d'information en continue, elles maintiennent une position plus ambiguë sur le traitement de l'actualité et sur l'indépendance envers les gouvernements étrangers qui les financent.

La troisième catégorie correspond aux chaînes « Gilets Jaunes ». Nous avons décidé d'intégrer dans cette catégorie autant les chaînes créées à la faveur du mouvement des gilets jaunes que des chaînes dont la politique éditoriale a été modifiée à l'occasion du mouvement. On retrouve alors à la fois des chaînes qui sont spécialisée dans le

traitement des manifestations (HORS-ZONE Press), des chaînes de discussions/soutien/réaction du mouvement à la manière de vlog (Ramous), des chaînes de recensement des contenus (Gilets Jaunes Commercy), mais également des chaînes qui ont transformé leur contenu pour correspondre à l'actualité, et probablement dans une stratégie de visibilité de la chaîne (par exemple des chaînes de gaming).

La quatrième catégorie est la « Médiation, vulgarisation et le décryptage » (9%). Il s'agit de chaînes spécialisées dans le décryptage de l'actualité, la vulgarisation souvent scientifique ou politique, pouvant néanmoins être attachées à une critique sociale, adoptant le registre de la dénonciation. On retrouve souvent des individus YouTube qui créent du contenu natif à YouTube comme Hugo Decrypte, Osons Causer spécialisés dans le décryptage politique, Aude What the Fake qui s'attachent à débunker des fake news. Quelques émissions de médias traditionnels peuvent appartenir à cette catégorie. Il s'agit pour l'essentiel d'émission qui imitent le format des vidéos YouTube.

La cinquième catégorie est la « Communication de Relation Publique » (7%). Elle regroupe toutes les chaînes promouvant l'action d'organisation, de fondation et de représentation du gouvernement et de l'action publique en général. On retrouve dans cette catégorie les Pompiers de Paris, le Gouvernement, l'Armée de l'Air, mais aussi la Mairie de Paris, entre autres.

La sixième catégorie sont les « Médias traditionnels » (6%). On retrouve toutes les chaînes des journalistes professionnels qui délivrent un contenu d'information grand public : médias télévisuels, radio, presse écrite, les émissions de tv, mais aussi les médias pureplayers. Il s'agit donc à la fois de contenus reproduits venant d'autres médias (comme France Inter, Arte, On N'est Pas Couché), mais aussi des formats natifs du web comme Brut et Kombini. Certains de ces médias ont également une stratégie numérique très développée, comme France 24 et France Inter qui diffusent en direct *live* leurs émissions toute la journée, puis tronçonnent ses émissions en petites vidéos qu'ils postent au long de la journée. On repère donc une grande activité et un grand

nombre de publication, ainsi qu'une stratégie de *live* semble-t-il encouragée par YouTube. En effet, quand on regarde le flux des « Tendances » de la catégorie « News », on retrouve mis en avant les *live* de ces chaînes médiatiques traditionnelles.

La septième catégorie regroupe les chaînes d'« Informations locales » (5%). Ces chaînes sont nombreuses, et associées à un territoire géographique délimité (villes, régions) et rendent compte des informations locales, les actualités aussi bien que les événements organisés s'y déroulant. C'est un mélange entre de l'information d'actualité et de la promotion institutionnelle des territoires. On retrouve notamment le Grand Lyon TV, les chaînes régionales de France 3 etc.

La huitième catégorie regroupe les chaînes « Humour, musique, satire » (5%). Nous nous sommes limités aux chaînes publiant du contenu musical, humoristique et satirique dont la ligne éditoriale est entièrement ou partiellement axée vers des sujets politiques – par exemple, La Cagoule et ces parodies musicales pour les Gilets Jaunes, Fabien Olicart et son décryptage de la gestuelle d'Emmanuel Macron qui a été l'une des plus populaires de notre corpus, ou encore Verino, un humoriste qui traite de l'actualité en public. Si elles sont moins productives que les médias traditionnelles, ces chaînes produisent des contenus très vus et très partagés sur tous les réseaux sociaux, et peuvent avoir un impact important sur la compréhension et la lecture de l'actualité.

La neuvième catégorie concerne les « Informations sélectionnées, collectionnées, détournées, manipulées » (4%). Ces chaînes publient du contenu vidéo extrait des médias traditionnels ou d'autres chaînes YouTube. Le principe de ces chaînes est de sélectionner du contenu, souvent polémique, sensationnel ou étrange, et de le republier. Si certaines chaînes se contentent d'une simple republication sans aucune intervention sur le contenu original, d'autres réalisent un montage (le format du zapping qui est très populaire sur YouTube, comme ZapduBuzz) ou détournent le contenu en ajoutant des commentaires,

comme un format de chronique (Le Pixel Mort). Une grande diversité de contenus donc soit repostés, soit modifiés coexistent dans cette catégorie, certaines chaînes sont par ailleurs supprimées au fil des mois pour atteinte au droit d'auteur.

On retrouve dans la dixième catégorie les « Conférences » (2%). Ce style de vidéos est très spécifique et montre l'attrait des internautes pour l'apprentissage et les longs formats de discussions et tables-rondes sur YouTube. Elles se différencient des chaînes de vulgarisation dans le ton et le format des vidéos, la plupart du temps il s'agit de captations de conférences réalisées « in real live », et non pas d'un contenu natif pensé pour YouTube, s'adressant aux internautes. Ces chaînes peuvent être celles d'Université comme celle de Dauphine qui retranscrivent leurs conférences, ou celles d'individus comme Franck Lepage qui est très suivi par sa communauté sur YouTube.

La onzième catégorie regroupe les chaînes dites d'« Enquête et investigations » (2%). Ces chaînes partagent en commun leur méthode de l'enquête ou l'investigation comme principe de production de contenu. On retrouve des vidéos de debunkage (Defakator), mais aussi des chaînes d'émissions d'investigations comme Médiapart, Fakirpress, ou de manière plus traditionnelles Cash Investigation, ou encore Documentaires HD qui repostent des documentaires télévisuels d'investigation. Ces formats des contenus sont souvent longs et sensationnalistes et gardent l'attention de leurs spectateurs, et ont un nombre de vue très élevé.

La dernière catégorie est celle des « Informations spécialisées » (1,9%). Ce sont des chaînes diverses de niche, spécialisée ou experte, comme l'Ina Actu, Boursorama pour l'économie ou L'Équipe pour le sport.

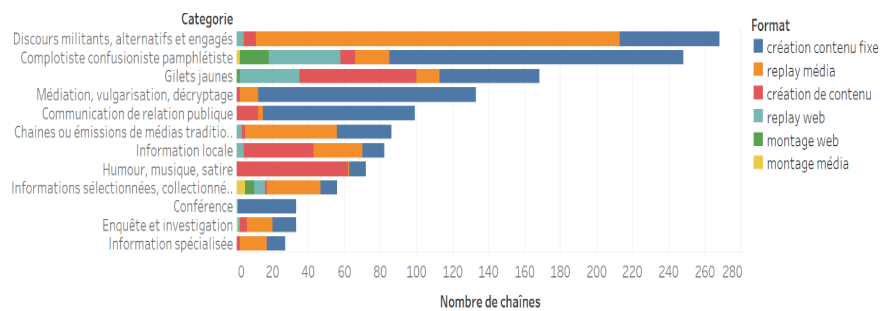


Figure 2 : Répartition du format des vidéos en fonction de la catégorie des chaînes du corpus

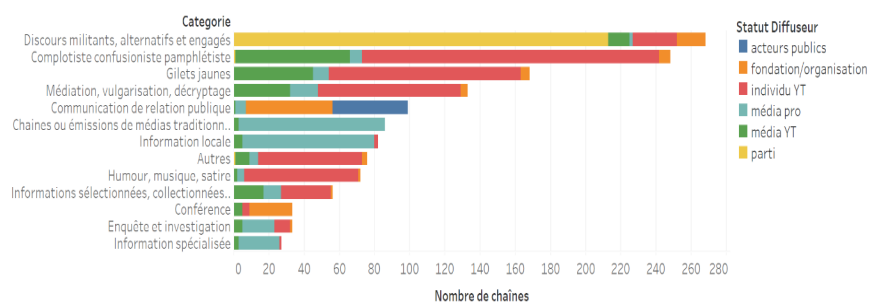


Figure 3 : Répartition du statut des diffuseurs en fonction de la catégorie des chaînes du corpus

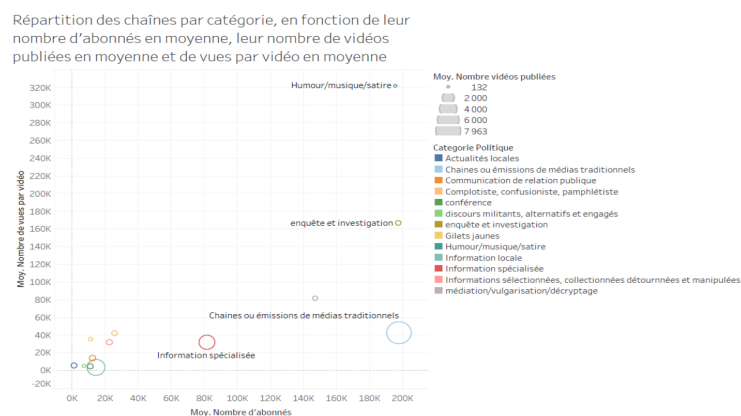


Figure 4 : Projection des catégories de chaînes en fonction du nombre moyen d'abonnés, de vues et de vidéos

Alors que les formes d'engagement politique traditionnel utilisent YouTube pour retransmettre du contenu produit par d'autres médias, les catégories Contenus litigieux, Gilets Jaunes et Vulgarisation produisent du contenu inédit et destiné principalement à YouTube. Si les médias traditionnels atteignent un nombre d'abonnés similaire à celui des chaînes d'humour et de médiation, ils sont largement plus actifs que les chaînes des autres catégories. En revanche, les chaînes d'humour, d'enquête ou de médiation ont un nombre de vues moyen par vidéo largement supérieur à celui des médias traditionnels. Ces derniers se situent approximativement au même niveau en moyenne que les chaînes de complotistes et de Gilets Jaunes en termes de nombre de vues moyen par vidéo.

Productivité, abonnement inter-chaînes et recommandation algorithmique

À cette triple catégorisation, nous avons ajouté le nombre d'abonnés, le nombre total de vues, de likes et de dislikes, le nombre total de vidéos et la date de création de la chaîne. A ces indicateurs, nous pouvons ajouter une autre métrique, construite à partir du graphe des chaînes qui se recommandent les unes les autres : le nombre de chaînes de notre corpus ayant recommandé la chaîne en le désignant comme « amies » ou en s'y abonnant (dans le vocabulaire de l'analyse de réseaux, le degré entrant du nœud dans le graphe). Cette sélection de chaînes opérée par la chaîne elle-même fournit une métrique de « visibilité interne », c'est-à-dire produite par les YouTubeurs eux-mêmes.

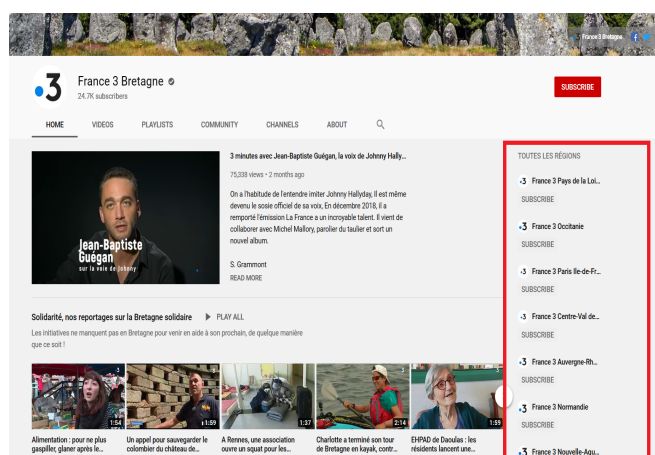


Illustration 1 : Capture d'écran d'une chaîne YouTube: Le rectangle rouge correspond à ces chaînes « featured », dites chaînes amies renseignées par les chaînes elles-mêmes, ici France 3 Bretagne. On note pour France 3 Régions une stratégie globale de visibilité des chaînes locales sur tout le territoire qui s'entre-recommandent.

Outre ces recommandations entre chaînes, nous avons collecté aussi des informations relatives aux recommandations algorithmiques. Sur YouTube les recommandations algorithmiques sont proposées à différents endroits : notamment sur la page d'accueil, dans la section des tendances, après chaque visionnage et à droite de chaque vidéo visionnée. Nous nous sommes intéressés à ce dernier mode de recommandation car il représente la majorité des recommandations suivies. YouTube distingue deux types de recommandations : les vidéos personnalisées que l'on reconnaît par l'indication « recommandées pour vous » (calculées en fonction du profilage de l'utilisateur) et celles qui ne le sont pas. Ces dernières, que YouTube appelle « cold start », correspondent à ce que la plateforme recommande à un utilisateur qui n'a pas activité de compte. C'est dans cette configuration de *cold start* que nous avons dressé le réseau dressé le réseau de chaînes liées par vidéos recommandées. Les chaînes associées aux 20 premières vidéos recommandées en relation au dernier vidéo publié par une chaîne ont été extraites une fois par jour du 1er mai au 3 juin 2019. Ainsi,

soulignons-le, le niveau de recommandation calculée est relatif à notre corpus.

Il est d'abord intéressant d'examiner dans quelle mesure ces indicateurs évoluent de façon conjointe. L'analyse des corrélations entre les variables dont nous disposons, représentée ici sous la forme d'une projection de l'analyse en composantes principales (ACP), est sans équivoque (figure 2)¹.

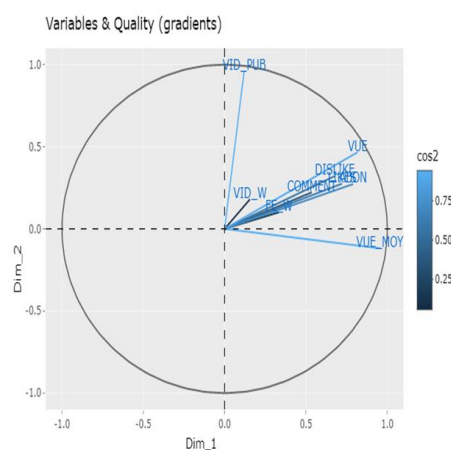


Figure 5 : Cercle des corrélations de l'Analyse en Composante Principales pour les deux premières dimensions.

Six indicateurs sont très fortement corrélés entre eux : le nombre total de vues, de likes, de dislikes et de commentaires, le nombre d'abonnés et le degré entrants du réseau de chaînes amies. En revanche, les recommandations algorithmiques sont faiblement corrélées au nombre de vue moyen des vidéos (une quasi-orthogonalité sur la représentation). Une précaution doit être soulignée ici : les métriques de recommandation algorithmique et les autres métriques n'ont pas été

¹ Nous avons intégré ces métriques dans l'ACP comme variables « actives » (la contribution à la construction des axes factoriels), tandis que les catégories de chaînes (les formes d'expression) sont seulement illustratives (positionnées par rapport aux différents axes factoriels sans intervenir dans leur formation).

construites de la même façon. Ces dernières mesurent des indicateurs relatifs aux chaînes depuis leur création, tandis que la mesure de recommandation algorithmique correspond à une période plus courte et estime des recommandations de chaînes à partir d'un nombre limité de vidéos associées. L'effet est néanmoins suffisamment net pour considérer deux dimensions relativement distinctes de l'accès à la visibilité : certaines chaînes sont visibles, mais sans bénéficier du soutien de l'algorithme, alors que d'autres semblent en dépendre entièrement.

L'ACP dégage quatre composantes principales :

- (1) La première composante montre une corrélation forte entre le nombre de vues moyennes par vidéo, le nombre de vues total accumulées par la chaîne, l'engagement (commentaires, likes, dislikes) et les abonnements. Cet axe peut être considéré comme l'axe de l'engagement et du visionnage effectif. Nous l'appelons l'axe du *degré de publicité*.
- (2) La deuxième composante représente les cadences de production des chaînes, c'est-à-dire la quantité de vidéos partagées par jour par les chaînes. Cet axe est considéré comme un axe de mise en visibilité : produire, et surtout surproduire, est une manière d'occuper l'espace et, partant, de se rendre visible.
- (3) La troisième composante renvoie au degré de centralité du réseau de recommandation interchaîne, il représente ainsi l'axe de mise en visibilité par les chaînes elles-mêmes. Cet axe sera peu exploité dans l'interprétation de l'ACP.
- (4) La quatrième composante est aussi un axe de mise en visibilité. Elle nous informe sur le niveau de recommandation algorithmique des chaînes sur la période du mois de mai 2019. Notons que cet axe n'oppose pas une forte et une faible soutien algorithmique, mais plutôt, en haut, la recommandation algorithmique « *cold start* » opposée, en bas, aux autres modes de recommandation, par exemple la personnalisation ou l'alerte produite par les abonnements, ou bien

encore les requêtes sur les moteurs de recherche Google et YouTube etc.

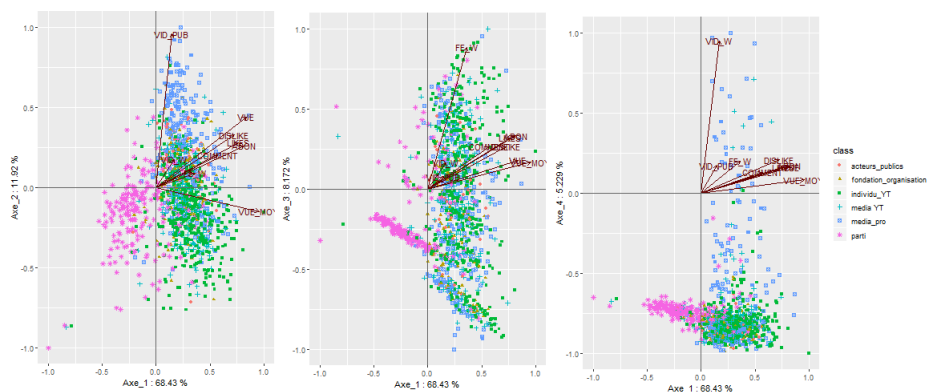


Figure 6 : Projection des chaînes à partir de l'ACP selon trois configurations des plans factoriels. L'axe 1 vertical représente donc toujours le degré de publicité. L'axe horizontal représente respectivement de gauche à droite la production, les abonnements entre chaîne et la recommandation algorithmique.

Ainsi, l'espace des positions des chaînes sur YouTube se distribue selon le degré de publicité (l'axe horizontal) et trois modes de mise en visibilité : la production, les abonnements entre chaîne et la recommandation algorithmique *cold start*. En projetant le statut des chaînes, on observe des positions caractéristiques de certaines catégories de chaîne : les partis politiques apparaissent pour la plupart en bas à gauche, donc comme des acteurs peu productifs, suscitant peu d'abonnement interchaînes et sous recommandés ; les médias professionnels se retrouvent souvent en haut à droite : ils sont donc hyper présents car à leur capacité productive et sont les chaînes les mieux recommandés. Les individus YouTube quant eux atteignent un bon niveau de visibilité, mais avec un niveau de production et un soutien algorithmique loin derrière celui des médias traditionnels.

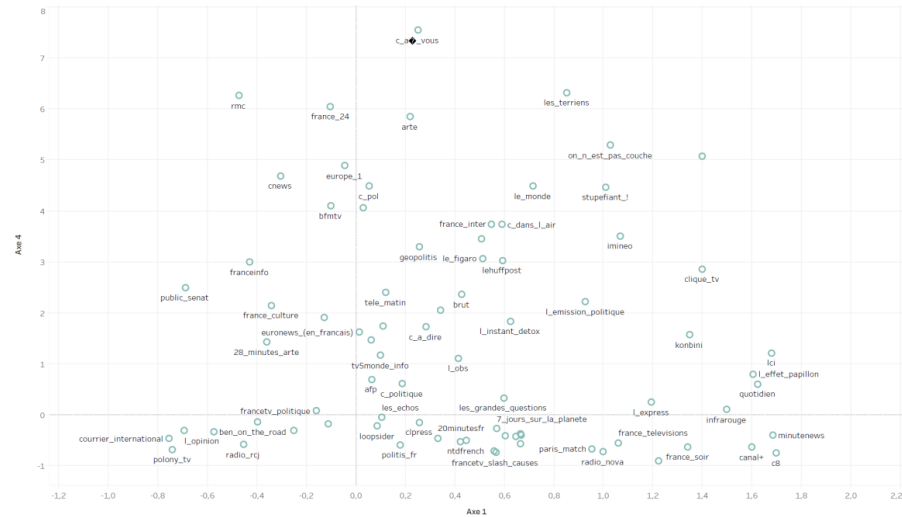
Zoom sur la recommandation algorithmique : le cadran de l'espace public numérique.

Puisque ce papier s'intéresse avant tout à la relation entre publicité et recommandation algorithmique, nous allons par la suite nous concentrer sur l'espace de droite de la figure 6 qui croise l'axe 1 à l'axe 4. Dans la section suivante, nous allons analyser les différentes régions de ce cadran publicité/recommandation, en observant quelles sont les catégories plus représentées dans chacune.

Haute recommandation, basse publicité : la stratégie de l'occupation de l'espace de YouTube.

En haut à gauche du cadran visibilité/recommandation, on trouve des chaînes très recommandées par l'algorithme *cold start*, mais au degré de publicité faible. Les chaînes les plus caractéristiques de cette situation sont celles appartenant à la catégorie « médias traditionnels », comme France 24, Europe 1, RMC, CNEWS, BFM, France info, Public Senat, Arte, son émission 28 minutes, TV5 monde et son émission TV5 monde info, Sud Radio, l'émission On a tellement de chose à se dire sur RT, Brut ; L'Obs ; C l'hebdo ; C-à-dire ; Geopolitis ; C'est à vous. D'une manière moins significative, d'autres catégories de chaînes s'inscrivent dans cette configuration comme Boursorama, INA Politique, Société et Talk Shows, La grande librairie, La Quotidienne (Information spécialisée) ; Martinique La 1ere, France 3 Paris Île de France, Nouvelle Aquitaine, Grand Est (médias locaux) ; Mediapart, le Media et Citoyens, Citoyennes! (Enquête), Adrien Quatennens, Tatiana Ventose, Francois Ruffin, Jean-Luc Mélançon, Mathilde Panot, Les Veilleurs (militants) ; Regards (Vulgarisation) ; Russia Today France, Sputnik France, Nouvelles VN7, Planetes360, Grand Angle, Dissident Officiel, Le média pour Tous, (Contenu litigieux).

Chaînes médias traditionnels



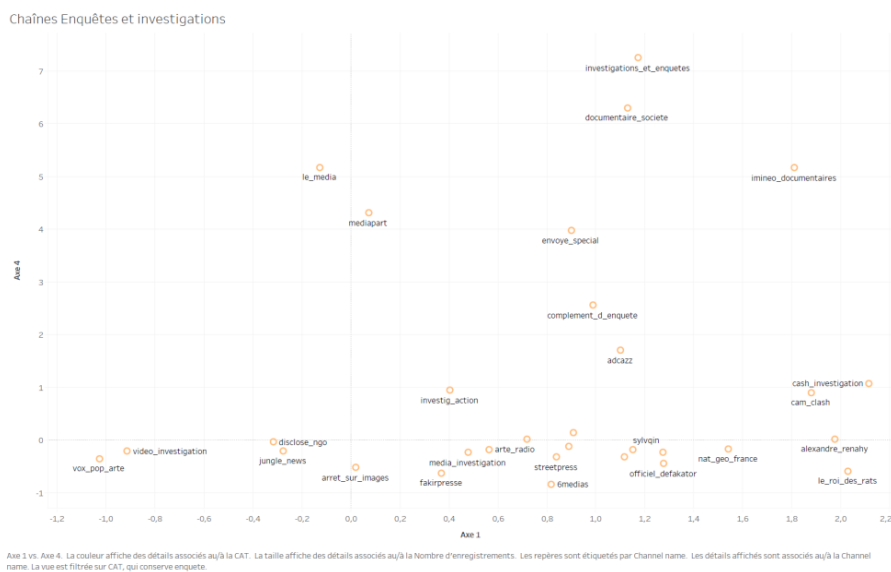
Axe 1 vs. Axe 4. La couleur affiche des détails associés au CAT. La taille affiche des détails associés au nombre d'enregistrements. Les repères sont étiquetés par Channel name. Les détails affichés sont associés au Channel name. La vue est filtrée sur CAT, qui conserve media_trad.

Si cet espace est caractéristique des chaînes de la catégorie des médias traditionnels, notons que les chaînes des autres catégories qui occupent cette position sont tenues par des professionnels des médias ou en republient les contenus. Par exemple, dans la catégorie Contenu litigieux, la chaîne Dissident Officiel publie du replay qui dit se concentrer sur des moments « politiquement incorrects » de la télévision ; dans la catégorie enquête, la chaîne Citoyens, Citoyennes ! est aussi une chaîne qui publie en replay des reportages de la télévision ; François Ruffin anime sa chaîne en tant que député de la Somme proche de la France Insoumise, mais c'est un journaliste professionnel, fondateur et rédacteur en chef du journal Fakir, longtemps reporter pour l'émission « Là-bas si j'y suis » diffusée sur France Inter.

Quel que soit leur catégorie, ces chaînes partagent la caractéristique de produire plusieurs dizaines de vidéos par jour, voire des centaines pour les chaînes de Radio ou l'AFP. Si les scores de vues individuelles des vidéos sont loin derrière ceux des vidéos positionnées à droite sur l'axe du degré de publicité, l'accumulation des vues de ces chaînes finit par atteindre un score élevé. Parmi ces chaînes, on observe des pratiques très spécifiques : au-delà de poster des dizaines de vidéos par jour, on

retrouve une tendance à diffuser de l'information en direct. C'est le cas de France24, de France Inter, d'Europe 1, de RT France, de Mediapart et bien d'autres encore. Dès lors, ces chaînes apparaissent régulièrement dans l'onglet Tendances de YouTube, dans la catégorie « Actualités », où l'on retrouve des vidéos en *live*. Ces vidéos *live* servent à la fois à obtenir une recommandation élevée de la chaîne, mais aussi à tronçonner ces *lives* en plusieurs dizaines de vidéos (par émission, par exemple). Ainsi, bien qu'elles atteignent un faible degré de publicité, la visibilité de ces chaînes est importante : elle procède d'un couplage entre une occupation massive de l'espace de YouTube avec un haut niveau de recommandation algorithmique.

Haute recommandation, haute publicité : le succès de l'info-divertissement.



A droite, il s'agit des chaînes fortement recommandée et caractérisées par un haut degré de publicité. Les chaînes les plus caractéristiques de cette position sont les chaînes de la catégorie Enquête qui diffusent des reportages longs d'investigation : le YouTubeur Adcazz qui propose des explorations urbaines critiques ; Cam clash, l'émission de France 4 qui

[illegible]

TV5 monde. On trouve aussi dans cette catégorie des chaînes de presse papier comme Le Monde, Le Huffingtonpost et le Figaro et des médias *pure players* comme Konbini et Brut. Quelques chaînes de contenu litigieux occupent aussi cette espace, mais notons qu'elles sont produites par des professionnels de la télévision ou des médias comme l'humoriste Dieudonné ou la chaîne de l'essayiste Pierre Jovanovic, longtemps journaliste au Quotidien de Paris, un journal polémiste. Alain Soral, qui occupe cet espace avec la chaîne ERTV, a débuté sa carrière de polémiste en tant qu'écrivain sur les plateaux de télévision des émissions de *talk show* les plus populaires. On peut faire le même constat d'une présence de professionnels de la télévision pour les chaînes de vulgarisation qui occupent cette zone, avec par exemple la chaîne Data Gueule de France 4. Et lorsqu'on trouve des chaînes de vulgarisation natives de YouTube dans cette zone, il s'agit de youtubeurs mainstream comme Hugo Décrypte, Fabien Olicard et Thinkerview dont les contenus suscitent un intérêt dans le monde des médias traditionnels. L'accès algorithmique à la visibilité sur YouTube est donc indissociable d'une certaine légitimité professionnelle et, partant, de l'autorité des chaînes.

Quel que soit leur catégorie, toutes ces chaînes – fortement recommandées par l'algorithme et d'un haut degré de publicité – partagent les caractéristiques de proposer du contenu accrocheur et de l'info-divertissement, un type de contenu qui trouve facilement son public sur YouTube. C'est donc aussi la part de la télévision la plus polémique qui bénéficie du soutien de l'algorithme de recommandation.

Ainsi, de la partie haute de l'espace des positions des chaînes, se dégage un mode d'accès à la visibilité qui s'apparentent à une forme de *colonisation* de l'espace numérique par les professionnels des médias (télévision, radio et presse papier), à grand renfort de recommandation algorithmique – un accès à la visibilité soit par une hyper-présence (cadence de publication, contenu long de qualité), soit par la production d'un contenu qui s'assimile à YouTube et en assimile les codes.

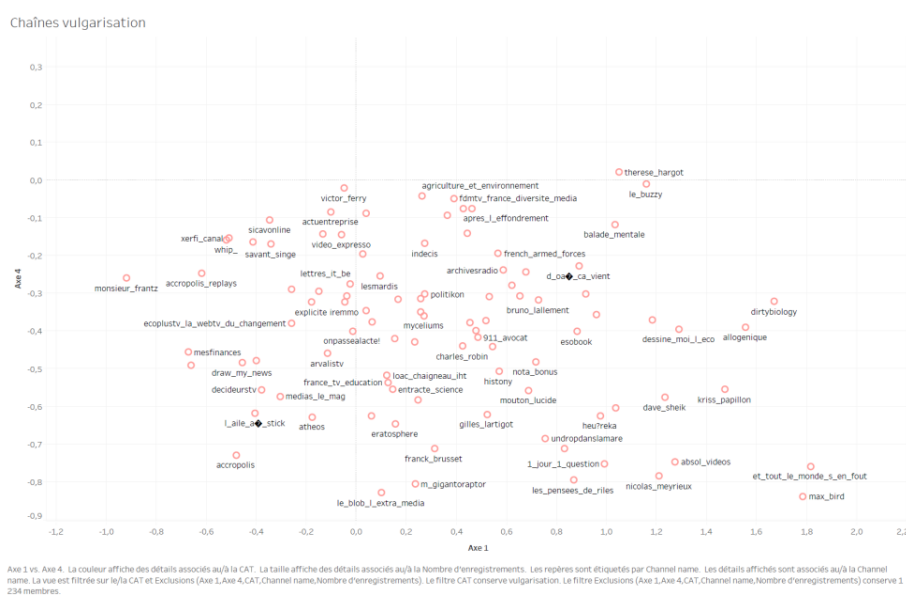
Mais pourquoi les médias professionnels sont-ils plus recommandés que les autres catégories de chaînes ? Est-ce le fruit de l'apprentissage machine dans le système de recommandation ou d'une intervention directe sur l'algorithme afin de valoriser les médias professionnels ? On peut faire l'hypothèse que ces deux processus agissent de manière concomitante.

D'un côté, la visibilité générée par l'apprentissage machine pourrait s'expliquer par les multiples avantages indirects qui découlent des fortes cadences de production et des dynamiques de couvertures médiatiques propre aux médias traditionnels : un niveau de fraîcheur élevé des vidéos (une variable qui pèse dans les choix du système de recommandation) ; et des agendas médiatiques mimétiques entre médias professionnels produisant des liens de recommandation sémantiques entre les chaînes de la même catégorie (par la similarité sémantique, les fortes recommandations de certains augmentent par effet de ruissellement les recommandations des autres).

D'un autre côté, l'hypothèse d'une visibilité générée par une action directe de YouTube est aussi plausible : YouTube crée pour une minorité de chaîne le P-score, un système de note qui repose sur des critères de qualité comme ... Il est évident que ces critères sont favorables aux médias professionnels. En effet, le dévoilement récent de ces notes a montré que ce système de classement est largement à l'avantage des chaînes de télévision, du moins de tout un ensemble d'acteurs dont la notoriété s'est construite en dehors de YouTube. Si le P-score est l'instrument du programme *Google Preferred* qui permet à YouTube une sélection de chaînes qui représentent une valeur sûre pour les annonceurs, on peut se demander dans quelle mesure cette politique de monétisation interfère avec l'algorithme de recommandation. La question reste ouverte.

Basse recommandation, haute publicité : les autres supports de la visibilité.

Une troisième configuration, en bas à droite du cadran, indique un degré élevé de publicité, mais une recommandation *cold start* relativement faible. Les chaînes les plus caractéristiques de cette zone sont les chaînes de vulgarisation et de contenu litigieux.

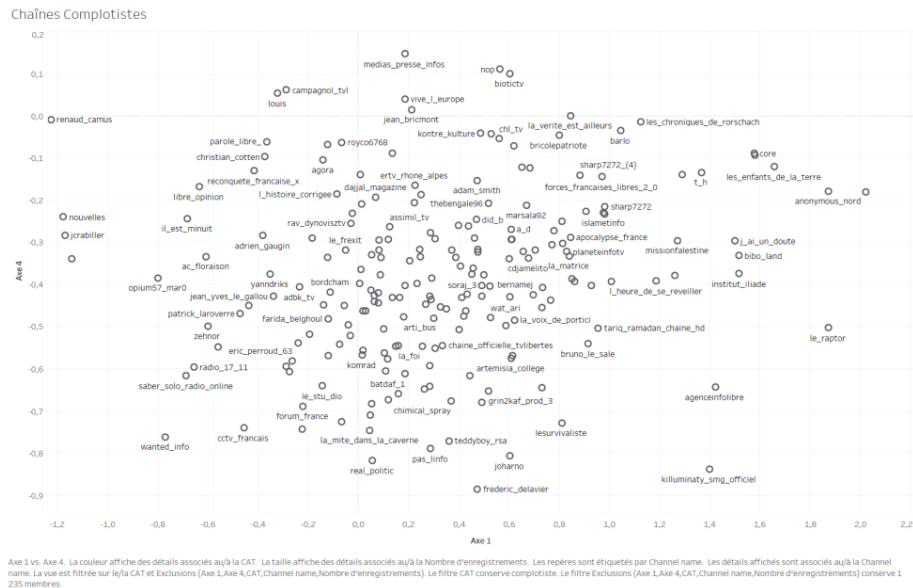


Pour la vulgarisation, on trouve par exemple Poisson Fécond, E Penser, Telecrayon, Citoyen Prévoyant, Therese Hargot, Le Buzzy, Balade Mentale, D'où ça vient, Draw my economy, Dirtybiology, Horizon Gull, L'astrolabe, Allogénique, Dessine moi l'éco, Esobook, Kriss papillon, Dave Sheik, Hygiène Mentale, Heureka, Undropdanslamare, Krissdenerf, Absol videos, 1 jour 1 question, Et tout le monde s'en fout, Nicolas Meyrieux, Les pensées de riles. Du côté des contenus litigieux, Blondewoman1, Le Raptor, Anonymous Nord, Les Enfants de la Terre, La vérité, Core, Institut Iliade, Bibo Land, J'ai un doute, Agence Infolibre, Killuminaty smg officiel, T H, Videoarchives3, Missionfaestine, Jeremysteres, L'heure de se réveiller, Les Chroniques de Rorschach, Barlo, Le Chat Patriote, Sharp7272, Quenelle Power,

Islam et info, Forces Francaises libres 2 0, Tariq Ramadan Chaîne_hd, L'informateur, Bruno le Sale.

Toutes ces chaînes se caractérisent par un haut niveau de visibilité par vidéo, mais une capacité de production plus faible que celle des médias professionnels, ce qui s'explique en partie par le statut de ces chaînes (des YouTubeurs qui ont des ressources plus faibles que celles des médias professionnels) et la nature des contenus produits qui visent une certaine originalité (un rythme lent de production, donc).

Une question mérite d'être posée : si ces chaînes atteignent un niveau élevé de visibilité, mais ne sont pas soutenues par la recommandation algorithmique, comment obtiennent-elle autant de vues et d'engagement ? Deux hypothèses peuvent être faites : la première est que ces chaînes accèdent à la visibilité via un public d'abonnés, une communauté de fan très investie dans leur contenu ; la deuxième est qu'elles seraient rendues visibles via la recommandation personnalisée de l'algorithme. Ces deux hypothèses ne sont pas exclusives—l'abonnement pourrait susciter de la recommandation algorithmique étant donnée l'historique de visionnage et vice-versa. En général, ces chaînes ne sont pas favorisées par la recommandation de YouTube, mais plutôt par une « fanbase », ou une communauté de fan très investie dans leur contenu. Par exemple, Dirty Biology est une chaîne de vulgarisation scientifique célèbre qui a largement compté sur son audience pour fonder sa visibilité. Par ailleurs, le faible recommandation par l'algorithme peut aussi s'expliquer dans l'adéquation de notre corpus, en effet nous nous concentrons sur un YouTube politico-médiatique, ces chaînes de vulgarisation n'ont pas forcément un contenu directement lié à l'actualité, et sont donc peu recommandées dans ce cadre très précis.

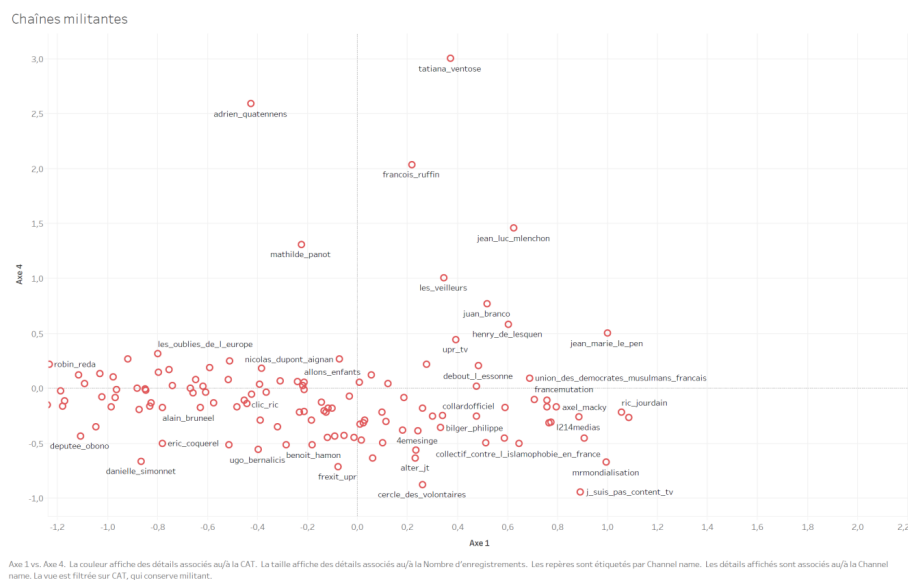


On observe ainsi que certains YouTubeurs publiant des contenus litigieux produisent relativement peu de vidéos en comparaison aux médias professionnels, mais parviennent à une visibilité élevée pour chaque vidéo publiée. Cette visibilité pourrait s'expliquer par la fidélité d'un public de fan qui, en retour, influencerait la recommandation personnalisée vers ce même public. Autrement-dit, ce troisième type de visibilité est probablement celui des chambres d'écho et des bulles de filtrage et repose principalement sur les abonnements et la personnalisation.

Basse recommandation, basse publicité : YouTube comme espace public de stockage.

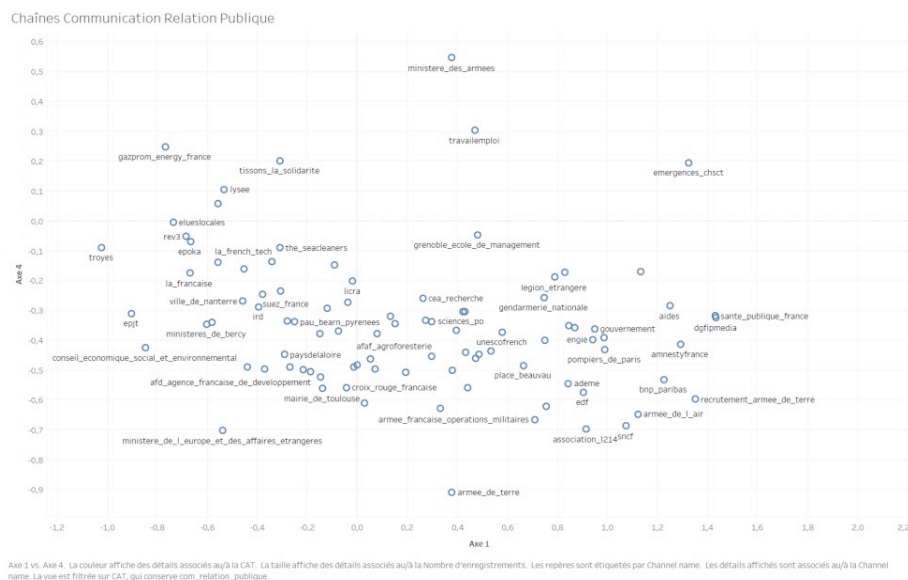
Une dernière configuration (en bas à gauche de l'espace des positions) indique des chaînes au degré de publicité faible, non soutenu par la recommandation algorithmique. Il s'agit de la zone la plus occupée de l'espace des positions : près de 60% des chaînes de notre corpus se retrouvent dans cette situation d'invisibilité. Toutes les catégories de

chaîne sont présentes, mais les plus caractéristiques de cette zone sont les conférences, la communication de relation publique et les militants.



Les chaînes de la catégorie « militant » sont composées de députés (ou d'autres type d'hommes et femmes politiques) et d'association. Si, comme nous l'avons vu plus haut, une poignée de chaînes ont intégré des méthodes qui leur permettent d'être visibles sur YouTube, la plupart ne s'en servent pas. Rappelons-le, pour exister sur YouTube, il faut se prêter au jeu de la plateforme et de ses usagers : c'est-à-dire, créer du contenu natif aux codes et aux attentes des spectateurs de YouTube. Ainsi, reposter des interventions fleuves à l'Assemblée Nationale ne semble pas correspondre à la demande. Aujourd'hui peu de représentants de la classe politique en France sont capables de maîtriser ces codes. Même les militants d'extrême droite, en particulier Marine et Jean-Marie Le Pen, qui déploient depuis le début du web une importante stratégie sur Internet pour contourner leur absence dans les médias traditionnels, ne constituent pas une force très ancrée sur YouTube. Les rares députés du Rassemblement National présent sur la plateforme se limitent, comme la plupart des députés, à rediffuser leurs discours institutionnels à l'Assemblée Nationale.

Cependant, les militants qui parviennent à assimiler les codes des YouTubeurs décuplent leur visibilité : on pense notamment à François Asselineau qui a durant les dernières élections Européennes été l'une des chaînes les plus visibles ou Jean-Luc Mélenchon et Juan Branco qui, parmi les militants, sont ceux qui ont plus bénéficié de la visibilité algorithmique tout en atteignant un degré élevé de visibilité. Ces chaînes reprennent bien les codes des YouTubeurs les plus populaires (publication d'une chronique toutes les semaines à heure fixe, créant ainsi un rendez-vous à leur communauté, vignettes attractives, discours dirigé vers les publics, chroniques d'actualité etc.).



Les chaînes de Communication de relation publique restent elles-aussi relativement invisibles. Comme pour les chaînes militantes, on comprend mieux les raisons de cette invisibilité en comparant les chaînes de communication de relation publique à bas et haut degré de publicité. Alors que le Ministère de la Justice relativement invisible se contente de publier pêle-mêle des vidéos d'évènements liés à tel ou tel déplacement du Ministre, des courtes vidéos de présentation de l'ENM ou quelques témoignages ; l'Armée de terre a su créer des chaînes spécifiques à chaque secteur, comme celle du « Recrutement armée de

terre » dont le contenu s'adresse directement à un public ciblé. Les ressorts de cette visibilité peuvent provenir de campagnes de publicité sur YouTube, ou encore de partenariats avec des YouTubeurs spécialisés (notamment le rôle de Tibo In Shape dans ces campagnes de recrutement) qui ont rendu visible à leur communauté la chaîne de Recrutement de l'armée de terre, des Pompiers etc. Mais ces chaînes de recrutement n'ont pas de lien avec les débats qui font l'espace public : les gouvernements locaux et nationaux n'interviennent pas sur les sujets polémiques qui font l'actualité, ce qui explique les désintérêts qu'elle représente pour le public de YouTube.

On retrouve ce même phénomène pour les chaînes de conférence : les nombreuses universités conçoivent YouTube comme un espace de stockage et d'archive de leur actualité interne. Deux chaînes de conférence sortent du lot : celles du rendez-vous du CERA et de l'Université de Nantes : le premier fait de la conférence sa spécialité, alors que la seconde cherche à créer se donne pour « ambition de créer le débat, faire connaître les sciences et d'éclairer le grand public sur les grands enjeux de notre société ». Là encore, ces deux chaînes produisent une offre de contenu qui rencontre une demande sur YouTube.

Aux dynamiques de colonisation, de l'info-divertissement et de la fidélisation de fans, vient s'ajouter une quatrième : la relégation, c'est-à-dire la conservation de certaines chaînes dans une position à très basse visibilité. Sur YouTube, se rendre visible implique un investissement de forme, mais la plus grande partie des chaînes de notre corpus ne s'en donne pas les moyens. N'étant pas produites pour fidéliser un public de fan, ces chaînes ne captent pas l'attention des usagers et en retour sont rétrogradées par l'algorithme de recommandation.

Diversité et visibilité : comment l'algorithme de recommandation ordonne-t-il l'espace médiatique

Pour approfondir l'analyse exploratoire, on peut enfin comparer les deux réseaux de mise en visibilité (recommandation entre chaînes et recommandation algorithmique) et particulièrement la diversité ou l'homogénéité qu'elle encouragent. On peut ainsi comparer deux manières de rendre visible des chaînes, l'une produite par une machine, l'autre par les créateurs de YouTube.

Pour cette comparaison, nous employons des marches aléatoires de longueur prédéfinie (K), qui simulent le comportement d'un utilisateur de YouTube se déplaçant d'une chaîne à l'autre les chemins créés par les liens d'abonnement entre ceux-là, ou alors par la recommandation algorithmique. Nous simulons, pour les deux réseaux, une centaine de marches pour chaque valeurs de K , de 0 jusqu'à 40 ($K=0$ correspond à un utilisateur qui regarde une vidéo, alors que pour K correspond à un utilisateur qui fait du *binge watching* ou « visionnage en rafale »). Nous calculons la distribution des différentes catégories dans la liste de chaînes traversées par les 100 marcheurs pour chaque réseau et chaque K . Enfin nous calculons la diversité de cette distribution pour obtenir un indicateur de bulle de filtre c'est-à-dire de faible diversité. Parmi les nombreuses mesures de diversité, nous avons choisi la perplexité ($2^{\text{EntropieShannon}}$)².

² La perplexité présente l'intérêt d'être facilement interprétable. Par exemple, si pour des marches aléatoires de longueur 39 la Perplexité des distributions d'une catégorie donnée est de 2.5, cela signifie que la consommation simulée (pour 100 marcheurs et $K=39$) est "aussi diverse" que celle d'un utilisateur qui consomme de façon équilibré et uniforme 2.5 catégories données de chaîne différentes

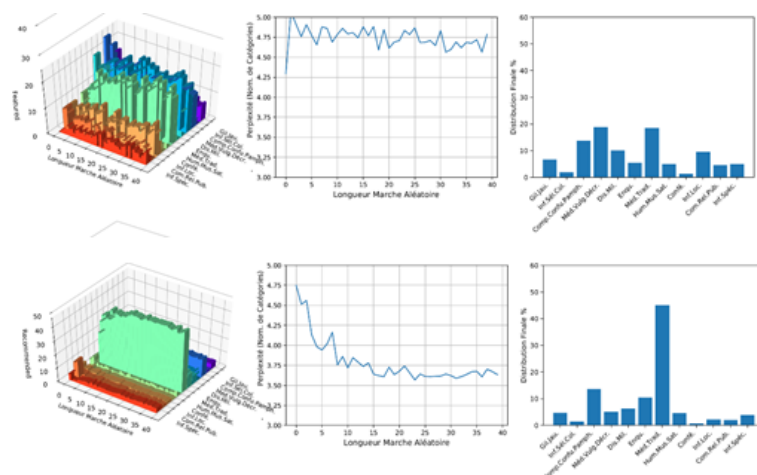


Figure 6 Distribution de la diversité de décile de degré de publicité par un marcheur aléatoire se déplaçant selon quatre réseaux de recommandations différents : (en haut) recommandation entre chaînes, (en bas) recommandation algorithmique . A Gauche, a distribution des formes d'expressivité (en %) des 100 marcheurs simulés, pour chaque K. C'est l'évolution (à K croissante) de la distribution des catégories atteints dans les marches aléatoires. Au centre, pour chaque K, on prend la distribution d'angle politique des 100 marcheurs, et on calcule la Perplexité. Cette courbe montre le changement de diversité selon la quantité de vidéos regardées par un utilisateur. Enfin à droite, on montre, pour clarté, la dernière distribution (celle pour K=40) selon les deux types de catégorie. Elle donne une idée de la consommation asymptotique dans une situation de *binge watching*.

On relève d'abord qu'un utilisateur a beaucoup plus de chance de tomber sur des chaînes à plus haut degré de publicité lorsqu'il emprunte les chemins de la recommandation algorithmique (fig. 6, représentation du bas). L'algorithme a donc une plus forte tendance à présenter à l'utilisateur des chaînes qui cumulent un nombre important de vues que la recommandation inter-chaîne. La recommandation algorithmique semble donc enfermer l'utilisateur dans un bulle, mais plutôt la bulle des contenus très visibles (et donc en grand partie provenant des médias traditionnels) que celle des contenus litigieux comme le soutiennent de nombreux critiques. Cette bulle est composée principalement des chaînes de média traditionnel. Notons enfin, que la

diversité est plus importante dans le système de recommandation produit par les humains – elle est bien plus homogène que celle produite par l’algorithme.

Conclusion

De nombreuses questions restent ouvertes à l’issue de cette enquête exploratoire. Cependant, la caractérisation du cadran de la visibilité sur YouTube apporte des éléments de contribution au débat sur les transformations de l’« espace public connecté » (Cardon, 2019).

Le résultat principal de cette recherche est que le système de recommandation (en *cold start*) réserve un accueil plus favorable aux médias traditionnels qu’aux chaînes natives de YouTube, en particulier les chaînes produisant du contenu litigieux. Ces résultats contrastent avec la critique adressée généralement à l’algorithme de YouTube par de nombreux activistes sans les nier pour autant. Il faut plutôt envisager cette faveur algorithmique accordée aux médias traditionnels comme un phénomène récent : YouTube a su intégrer la critique en apportant des modifications au système de recommandation (YouTube Official Blog, 2019). Mais ces transformations ont un impact sur l’ordonnancement global de l’information : l’objectif initial de rendre public des productions vidéos amateurs – l’appel au *broadcast yourself* – est progressivement compromis par l’impératif économique et moral de diriger le public vers du contenu préalablement filtré par des gatekeepers.

Si des chaînes natives de YouTube parviennent à un niveau élevé de visibilité, ce n’est pas grâce au soutien de la recommandation en *cold start*. Nous avons constaté que, si les vidéos générées par les chaînes de vulgarisation et de contenus litigieux atteignent un score de vues et un engagement largement supérieur à celui des médias traditionnels, c’est surtout grâce à la fidélité de leur public de fans. En l’état des données accessible, tout porte à croire que la visibilité de ces chaînes soit produite par les abonnements et/ou par la dimension personnalisée de

la recommandation, deux modes d'accès à la visibilité qui interagissent et se renforcent mutuellement.

Enfin, pour l'essentiel, les acteurs de la sphère publique traditionnelle ne parviennent pas à rendre visible leur contenu, notamment les députés, les associations militantes, les chaînes de gilets jaunes créées à l'occasion du mouvement et les chaînes institutionnelles de l'état. Dans ces catégories, seules les chaînes qui ont su adopter les codes de la création propre à YouTube parviennent à remonter la longue traîne de la visibilité.

YouTube apparaît comme un champ de lutte pour la visibilité dans lequel l'espace public traditionnel (la presse écrite, la radio et la télévision) co-existe avec un espace public étendu à des formes de prise de parole alternatives. Si la prise en main indissociablement politique et technique récemment opérée par YouTube favorise les sources d'informations traditionnelles, cela n'est pas sans conséquences pour le débat sur la régulation des plateformes. À la demande croissante pour une vérifications gouvernementales des systèmes algorithmiques au nom de la lutte contre la désinformation et les propos haineux (comme le montre la dernière expérimentation du CSA sur YouTube), il faut aujourd'hui ajouter la demande d'indicateurs pour évaluer dans quelle mesure les systèmes de recommandation nuisent à la diversité du débat public. Pour développer ce système de veille, il faudra que YouTube crée une API spécifique accordant un accès privilégié aux données pour les autorités de contrôle et la recherche publique.

Bibliographie :

Cardon D (2019) *Culture numérique*. SCIENCES PO.

Cha M, Kwak H, Rodriguez P, et al. (2007) I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System. In: *Proceedings of the 7th*

ACM SIGCOMM Conference on Internet Measurement, New York, NY, USA, 2007, pp. 1–14. IMC '07. ACM. DOI: 10.1145/1298306.1298309.

Crane R and Sornette D (2008) Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences* 105(41): 15649–15653. DOI: 10.1073/pnas.0803685105.

Figueiredo F, Almeida JM, Gonçalves MA, et al. (2014) On the Dynamics of Social Media Popularity: A YouTube Case Study. *ACM Trans. Internet Technol.* 14(4): 24:1–24:23. DOI: 10.1145/2665065.

Jiang R, Chiappa S, Lattimore T, et al. (2019) Degenerate Feedback Loops in Recommender Systems. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society - AIES '19*: 383–390. DOI: 10.1145/3306618.3314288.

Ledwich M and Zaitsev A (2019) Algorithmic Extremism: Examining YouTube’s Rabbit Hole of Radicalization. *arXiv:1912.11211 [cs]*. Available at: <http://arxiv.org/abs/1912.11211> (accessed 30 December 2019).

O’Callaghan D, Greene D, Conway M, et al. (2015) Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems. *Social Science Computer Review* 33(4): 459–478. DOI: 10.1177/0894439314555329.

Our ongoing work to tackle hate (2019). Available at: <https://youtube.googleblog.com/2019/06/our-ongoing-work-to-tackle-hate.html> (accessed 2 February 2020).

Ribeiro MH, Ottoni R, West R, et al. (2019) Auditing Radicalization Pathways on YouTube. *arXiv:1908.08313 [cs]*. Available at: <http://arxiv.org/abs/1908.08313> (accessed 30 November 2019).

Zhou R, Khemmarat S, Gao L, et al. (2016) How YouTube videos are discovered and its impact on video views. *Multimedia Tools and Applications* 75(10): 6035–6058. DOI: 10.1007/s11042-015-3206-0.