



HAL
open science

Methods for the reconstruction of the socio-semantic dynamics of political activist Twitter networks

Noe Gaumont, Maziyar Panahi, David Chavalarias

► To cite this version:

Noe Gaumont, Maziyar Panahi, David Chavalarias. Methods for the reconstruction of the socio-semantic dynamics of political activist Twitter networks: Application to the 2017 French Presidential elections. PLoS ONE, 2018, 13 (9), 10.1371/journal.pone.0201879 . hal-01575456v3

HAL Id: hal-01575456

<https://hal.science/hal-01575456v3>

Submitted on 21 Sep 2018

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Reconstruction des dynamiques socio-sémantiques des réseaux de militants politiques sur Twitter - Méthode et application à l'élection présidentielle française de 2017

Noé Gaumont¹, Mazyar Panahi², David Chavalarias^{1,2,*}

1 Centre d'Analyse et de Mathématique Sociales (CAMS), Centre National de la Recherche Scientifique (CNRS) / École des Hautes Études (EHESS), Paris, France

2 Institut des Systèmes Complexes Paris Île-de-France (ISC-PIF), CNRS, Paris, France

* david.chavalarias@ehess.fr

Pour citer cet article : Gaumont G., Panahi M. et Chavalarias D. (2018) Reconstruction of the socio-semantic dynamics of political activist Twitter networks - Method and application to the 2017 French presidential election, *PLOS ONE*, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201879>.

Résumé

Contexte Les espaces numériques, et en particulier les sites de réseaux sociaux, sont de plus en plus présents et influents dans le fonctionnement de nos démocraties. Dans cet article, nous proposons une méthodologie intégrée pour la collecte de données, la reconstruction, l'analyse et la visualisation du développement du paysage politique d'un pays à partir des données de Twitter.

Méthode La méthode proposée repose uniquement sur les interactions entre les comptes Twitter et est indépendante des caractéristiques des contenus partagés tels que la langue des tweets. Nous validons notre méthodologie sur une étude de cas portant sur l'élection présidentielle française de 2017 (60 millions d'échanges Twitter entre plus de 2,4 millions d'utilisateurs) via deux méthodes indépendantes : la comparaison entre notre catégorisation politique automatisée et une catégorisation humaine basée sur l'évaluation d'un échantillon de 5000 descriptions de profils ; et la correspondance entre les reconfigurations détectées dans le paysage politique reconstruit et les événements politiques clés rapportés dans les médias. Cette dernière validation démontre la capacité de notre approche à refléter avec précision les reconfigurations de la scène politique d'un pays à partir des données des réseaux sociaux.

Résultats Nous nous sommes appuyés sur cette reconstruction pour donner un aperçu des dynamiques d'opinion et de la reconfiguration des communautés politiques qui sont en jeu lors d'une élection présidentielle. Tout d'abord, nous proposons une description et une analyse quantitative de l'engagement politique des membres des communautés politiques. Ensuite, nous analysons l'impact des communautés politiques sur la diffusion de l'information et en particulier sur leur rôle dans le phénomène des fausses nouvelles. Nous mesurons un effet différentiel de chambre d'écho (*echo chamber*) sur les différents types de nouvelles politiques (fausses nouvelles, démentis, nouvelles standards) causées par la structure communautaire. Nous soulignons l'importance de prendre en compte les méso-structures des réseaux politiques pour comprendre les phénomènes de type "fausses nouvelles" (*fake news*).

Conclusions En donnant accès à un niveau intermédiaire, entre les enquêtes sociologiques de terrain et les grandes études statistiques (telles que celles menées par des organisations nationales ou internationales), nous démontrons que les données des réseaux sociaux permettent de qualifier et de quantifier l'activité des communautés politiques dans un environnement politique multipolaire, ainsi que leur évolution et reconfiguration temporelle, leur structure, leurs stratégies d'alliance et leurs particularités sémantiques au cours d'une campagne présidentielle à travers l'analyse de leurs traces numériques. Nous concluons ce document par un commentaire sur les implications politiques et éthiques de l'utilisation des données des réseaux sociaux en politique. Nous insistons sur l'importance de développer des macroscopes sociaux qui permettront aux citoyens de mieux comprendre la manière dont collectivement ils font société. Nous proposons comme exemple le *Politoscope*, un microscope qui restitue certains de nos résultats d'une manière interactive.

Mots-clés : opinion dynamics, political activism, elections, socio-semantic networks, networks dynamics, Twitter.

Introduction

Les espaces numériques sont de plus en plus présents et influents dans le fonctionnement de nos démocraties, au point que ce qui s'y passe ne peut plus être ignoré lorsqu'il s'agit d'analyser l'évolution du paysage politique. Après chacun des grands séismes politiques ayant secoué les démocraties occidentales ces derniers mois (ex. le Brexit, l'élection de Trump), les commentateurs se sont tournés vers les réseaux sociaux pour y trouver des éléments de compréhension.

Les réseaux sociaux ont également été, en quelques années, totalement intégrés aux stratégies des partis politiques. Barack Obama a été qualifié de "*game changer*" en 2012 alors qu'il était le premier candidat à la présidentielle à utiliser massivement les réseaux sociaux pour sa campagne. La campagne présidentielle USA 2016 a confirmé cette tendance avec un budget cumulé dédié aux réseaux sociaux estimé à près de 500M Euros [1].

La France a suivi cette même tendance. Les partis politiques les plus importants ont aujourd'hui recours aux services de sociétés spécialisées pour mieux comprendre leurs électeurs potentiels via l'analyse des masses de données issues des réseaux sociaux. Parmi ceux-ci, Twitter détient une bonne place dans les sources disponibles. C'est par ailleurs un réseau social utilisé par la plupart des personnalités politiques pour réagir à l'actualité ou donner de la visibilité à leurs opinions et leurs prises de paroles. Ce réseau social est ainsi devenu un média incontournable pour le monde politique lorsqu'il s'agit de faire passer des messages ou d'asseoir une légitimité.

Dans le but d'évaluer le potentiel de plateformes comme Twitter pour la recherche sur les dynamiques sociales, et particulièrement en sciences politiques, nous proposons une méthode s'appuyant sur l'analyse des réseaux complexes permettant de reconstruire et d'analyser les communautés politiques et leur évolution dans un environnement politique multipolaire.

Comme cas d'étude, nous évaluons cette méthode sur les données que nous avons recueillies lors des élections présidentielles françaises de 2017, représentant près de 60 millions d'échanges Twitter entre plus de 2,4 millions d'utilisateurs (projet *Politoscope* - <https://politoscope.org>). La nature exceptionnellement mouvementée de ces élections (voir S1 Texte section 1 pour un résumé) représente un scénario des plus difficiles pour ce type de reconstruction en raison de la grande variabilité des dynamique et garantit la robustesse de notre approche. Nous présentons ensuite quelques perspectives que ce type de reconstruction apporte aux sciences sociales et discutons de sa pertinence pour les sciences politiques.

Related work

Twitter fait l'objet d'une littérature académique assez importante, car cette plateforme cumule deux avantages. D'une part, elle est accessible au grand public, même si elle n'est pas représentative de l'ensemble de la population mondiale ou même Française [3], et ses contenus sont massivement repris par la presse traditionnelle ou la télévision. D'autre part, la collecte de données y est relativement aisée grâce aux différentes API proposées par Twitter¹.

Il existe de nombreuses méthodologies pour aborder les productions de Twitter dans le cadre de l'étude de l'action politique. Certaines méthodes [4-6] s'attachent à tirer profits du contenu des tweets pour déduire l'orientation politique d'un tweet ou d'un utilisateur. Cependant, les contraintes sur le format des tweets semblent limiter leur efficacité. Par exemple, d'après Cohen et Ruths [7], les hashtags utilisés pour parler de politique varient en fonction du militantisme des personnes et ne semblent que peu différents pour les personnes les moins militantes.

Afin de contourner les problèmes liés au traitement automatique du langage, les recherches se sont tournées vers des méthodes analysant la structure des échanges ayant lieu sur Twitter au lieu de leur contenu. Elles s'inspirent ainsi d'une étude pionnière sur les blogs politiques aux États Unis [8], qui a démontré qu'il y a plus de liens entre deux blogs s'ils sont du même bord politique.

Dans le cadre de Twitter, il existe plusieurs définitions possibles d'un lien entre deux comptes. Un compte peut suivre ou mentionner un autre compte. Il peut répondre, citer, commenter ou simplement reprendre à son compte (retweet) le tweet d'un autre compte. Toutes ces notions de liens sont pertinentes et dépendent de l'objectif de l'étude [9]. Cependant pour classer les comptes en fonction de leur orientation politique, il semble que l'utilisation du retweet comme définition d'un lien entre deux comptes soit la plus utilisée. Garimella *et al.* [10] et Connover *et al.* [6] ont notamment observé que pour étudier les débats politiques sur Twitter l'utilisation du retweet est plus pertinente que la mention.

Connover *et al.* [6] ont été parmi les premiers à exploiter les retweets en étudiant la structure de l'espace politique aux USA. Ils ont retrouvé sa nature bipartite et ils ont démontré qu'il existe significativement moins de retweets entre deux comptes lorsqu'ils sont de partis différents. De nombreuses autres études ont par la suite utilisé ce type de relation. Boutet *et al.* [11] ont appliqué la même méthode pour étudier la structure de l'espace politique en Angleterre et analyser les différences de pratiques dans la manière dont les groupes politiques utilisent les mentions, hashtags et les liens URL.

Garimella *et al.* [10] et Morales *et al.* [13] ont utilisé la notion de retweet pour mesurer l'existence de polarisation et de controverses sur Twitter. Amor *et al.* [14] s'en servent pour caractériser le débat sur *care.data* du *National Health System* anglais et Guerrero *et al.* [15] pour étudier le mouvement *Catalan Process Toward Independence*.

Ces approches ont permis de retrouver des structures politiques binaires mais la notion de retweet permet également de retrouver des structures multipolaires. C'est le cas de Cherepnalkoski *et al.* [16,17] qui ont retrouvé sur Twitter l'affiliation politique des députés européens via les retweets. Les auteurs notent cependant que l'approche se généralise mal à l'ensemble des utilisateurs car en plus des accointances politiques il existe des préférences nationales.

Le retweet n'est pas la seule relation exploitable pour caractériser la structure d'un espace politique. Les travaux de Barberá [18] ont utilisé les relations suiveur/suivi (follower/followee) entre comptes. L'originalité de leur travaux est d'assigner à chaque compte une position entre -1 et 1 sur un axe idéologique gauche-droite. Ainsi, un

¹<https://dev.twitter.com/overview/api>

compte n'est pas assigné au groupe démocrate mais à une position proche d'autres démocrates. Cette approche permet de comparer un ensemble de comptes sur un continuum idéologique. Ils ont été répliqués avec succès sur des données Twitter en France par Briatte et Gallic [19].

En utilisant leur méthode d'estimation idéologique, Barberá *et al.* [20] ont également montré qu'il y a une forte polarisation des retweet lorsque le sujet évoqué est politique, c'est-à-dire qu'il existe de nombreux retweets entre personnes proches idéologiquement et peu lorsqu'elles sont éloignées.

Enfin il est important de noter que malgré l'existence de ces nombreux travaux évaluant l'alignement idéologique, l'une des applications des sciences politiques consistant à effectuer des prévisions de vote reste problématique [21], notamment à cause de la représentativité de Twitter par rapport à la population réelle.

En quelques années, Twitter est ainsi devenu un domaine de recherche productif pour les sciences sociales. Bien qu'il présente certaines limites en termes de types de population qui peuvent être correctement étudiés, il convient de noter qu'il y a eu une forte augmentation de l'utilisation de Twitter au cours des dernières années avec plus de 15 millions d'utilisateurs actifs en France en 2017 [22] (la population française en 2017 était de 66,9 millions d'habitants) et un nombre estimé d'utilisateurs actifs qui a triplé dans le monde entier depuis les premières études en 2011 [23]. Par conséquent, on peut s'attendre à ce que certaines limites qui ont été relevées dans les articles précédents soient maintenant atténuées et que de nouveaux résultats puissent être obtenus.

Dans cet article, nous réévaluons les techniques d'analyse de retweet susmentionnées et nous les améliorons pour prendre en compte les dynamiques et l'aspect multipolaire des paysages politiques, en les appliquant à un corpus généré par des activistes politiques sur une période de près d'un an. Nous combinons également deux techniques d'analyse du flux Twitter (suivi de comptes et suivi de mots-clés) qui ont généralement été utilisés séparément. Ceci nous a permis de constituer un corpus Twitter de très haute qualité.

La taille de l'échantillon de tweets collecté ainsi que l'étendue de la période d'observation combinées avec une méthodologie améliorée et de nouvelles analyses ont, pour l'une des premières fois, permis d'étudier la *dynamique socio-sémantique* d'un paysage politique multipolaire.

Méthodologie

Données

Nous avons utilisé la plateforme Multivac du CNRS (<http://multivac.iscpif.fr>), une plateforme dédiée à la capture et à l'analyse des données de flux, pour capturer en temps réel l'activité politique de la twittosphère en France. Pendant une période de 11 mois précédant l'élection présidentielle, nous avons collecté en continu près de 60 millions de tweets concernant plus de trois milliers de personnalités politiques françaises et quelques dizaines de mots-clés politiques, qui ont été utilisés par plus de 2,4 millions de comptes individuels (voir Texte B dans S1 File). Nous appelons cet ensemble de données \mathcal{D} .

Toutes les données ont été collectées conformément aux conditions de service et de confidentialité de Twitter et sont mises à disposition conformément aux conditions d'utilisation de Twitter [24]. Toutes les données anonymes dérivées de l'analyse de cette base de données dans le contexte de cet article sont disponibles pour téléchargement [?].

Nous avons estimé la proportion de tweets de \mathcal{D} générés par des robots à moins de 5% (voir Texte C dans S1 File), ce qui nous assure que nos résultats restent représentatifs de l'activisme politique en ligne.

Rétrospectivement, étant donné la nature inhabituelle et mouvementée de ces élections (voir la section A du Texte S1 pour un résumé) qui se sont déroulées dans un

environnement politique multipolaire, cet ensemble de données s’est avéré à la fois d’un grand intérêt pour comprendre le paysage politique français et d’une grande valeur pour tester les méthodes de reconstruction de communautés.

Identification des communautés politiques

Definition des groupes sociaux

Les groupes sociaux, quelle que soit leur définition, ont des frontières floues et changeantes. Notre approche n’est pas centrée sur la localisation exacte des frontières des groupes sociaux, mais plutôt sur la façon dont elles se déplacent avec le temps, les scissions qu’ils peuvent produire, les sous-corpus de tweets qui peuvent être constitués via l’estimation des affiliations à des groupes sociaux et les processus de diffusion qu’ils permettent d’étudier.

La méthodologie proposée pour l’identification des groupes sociaux politiques ne suppose pas que ces groupes peuvent être déterminés de manière précise ou exhaustive. La méthode prend en compte les inexactitudes de qualification au niveau de l’individu (par exemple, quelques erreurs dans la détermination de certaines affiliations politiques), et le caractère arbitraire de certains choix liés au positionnement des frontières entre ces groupes (par exemple le choix des valeurs de paramètres pour les algorithmes de reconstruction).

Une définition du groupe social d’un point de vue sociologique est particulièrement adaptée aux études Twitter. En 1890, Gabriel Tarde [?] a défini un *groupe social* comme “*Une collection d’êtres en tant qu’ils sont en train de s’imiter entre eux ou en tant que, sans s’imiter actuellement, ils se ressemblent et que leurs traits communs sont des copies anciennes d’un même modèle*”. La notion d’imitation invoquée par Tarde était relativement générale et couvre tout des notions plus larges liées aux concepts d’apprentissage social et d’influence sociale (voir [?, 27] pour la conceptualisation de ces idées). Il couvre en particulier tous les cas où une attitude, une manière de faire ou de s’exprimer, ou une opinion est transmise avec peu de modifications entre deux personnes.

Twitter offre un terrain idéal pour l’adaptation de cette définition aux espaces numériques, en analysant la façon dont les gens s’expriment sur cette plate-forme. En effet, Twitter combine trois éléments clés, ce qui permet d’analyser la propagation des traits culturels :

- Tout le monde peut publier, sans filtrage, du contenu visible pour tous les autres utilisateurs,
- Le système follower/followee permet à quiconque de s’abonner aux publications de tout autre utilisateur,
- Le système de retweet permet de copier sans effort le contenu produit par une autre personne en transmettant une copie identique de ce contenu à tous ses followers.

Ainsi, en se concentrant sur les événements de type retweet (relais d’information sans modification), nous disposons de marqueurs permettant de définir, pour une période donnée, un ensemble de groupes d’utilisateurs au sein desquels les contenus Twitter ont tendance à circuler en subissant des modifications moindres (reprise à l’identique) que lorsqu’ils circulent entre ces groupes (transitions pendant lesquelles ils peuvent être détournés, commentés de manière ironique ou humoristique, contredits ou tout simplement non repris). Cela étant dit, nous avons considéré la définition suivante:

[Définition] Groupes sociaux dans la twittosphère politique : réseaux denses de comptes qui relaient entre eux, de façon récurrente et sans modification, des contenus (retweet).

Les communautés de retweet

Afin de définir des groupes d'utilisateurs sur la base des retweets, nous avons extrait les retweets de notre ensemble de données \mathcal{D} et défini un graphe $G_T = (V, E)$ dans lequel chaque nœud, $u \in V$, est un compte Twitter et chaque tuple $(u, v, w) \in E$ est un lien pondéré et non dirigé entre les nœuds u et les nœuds v, w étant proportionnel au nombre de retweets échangés entre les comptes u et v pendant la période de temps T .

Pour qualifier les groupes politiques dans G_T , nous adoptons une approche similaire à celle proposée par [29–31] en appliquant un algorithme de détection de communauté sans chevauchement à G_T . Parmi les algorithmes de détection de communauté état-de-l'art [32–34], nous avons choisi l'algorithme de Louvain développé par Blondel *et al.* qui, étant donnée la définition adoptée pour la notion de lien, calcule des communautés de nœuds qui correspondent exactement à la définition de groupe social que nous recherchons : ce sont des sous-ensembles de comptes qui échangent beaucoup plus d'informations sans modification au sein du groupe qu'avec le reste du réseau.

Dans ce qui suit, nous adoptons les définitions suivantes :

[Définition] Lien w-fort: deux comptes A et B sont liés par un lien w-strong sur la période T-fort s'ils se sont retweeté l'un l'autre au moins w fois, quelque soit la direction.

[Définition] w-communautés sur la période T: communautés détectées avec l'algorithme de Louvain sur les graphes de retweet, pour lesquels seuls les liens au moins w-forts ont été retenus.

Les communautés politiques trouvées sur une période T selon cette définition sont notées $\mathcal{C}_w^T = \{C_i^T\}_{1 \leq i \leq N}$. Dans les analyses et les visualisations, elles sont labellisées avec les noms des principales personnalités politiques présentes dans ces groupes, ceci afin de donner un aperçu de l'évolution des communautés politiques par rapport aux personnalités clés de la campagne présidentielle française (cf. figure 1 par exemple).

Dans ce qui suit, les labels ont été pris dans la liste des candidats à la présidence à laquelle a été ajoutée un petit nombre de personnalités politiques qui, bien que n'ayant pas été candidates, ont joué un rôle important dans cette campagne (par exemple, le président sortant, l'ancien président, les personnalités préférées dans les sondages, etc.). Le paramètre de résolution utilisé pour l'algorithme de Louvain était sa valeur par défaut.

Une partie importante des comptes n'appartiennent à aucune communauté avec cette approche. Ils correspondent à des comptes qui, bien qu'abordant des questions politiques, ne s'inscrivent pas dans la sphère informationnelle d'un groupe d'activistes politiques particulier. Nous appellerons cet ensemble de comptes "la mer".

Granularité sociale et temporelle

Il est possible d'étudier l'environnement politique à différentes résolutions, en sélectionnant une liste plus ou moins longue de comptes représentatifs, et en ajustant la résolution de l'algorithme de Louvain afin d'identifier les communautés distinctes dans le voisinage de ces comptes. Nous avons observé que les différents courants politiques au sein d'un parti pouvaient être identifiés de cette façon. Une telle analyse dépasse toutefois le cadre de la présente étude et nous avons choisi de conserver les paramètres

par défaut de l’algorithme de Louvain qui ont permis d’identifier avec précision les principaux courants et partis politiques.

Les réseaux socio-sémantiques co-évoluent : les acteurs sont influencés par les contenus auxquels ils sont exposés, ce qui les conduit à développer leur réseau local, et donc à modifier la nature des contenus auxquels ils seront exposés à l’avenir. En conséquence, il est courant pour les individus de changer de sphère d’influence politique pendant une campagne présidentielle, ou d’être influencés par de nouvelles personnalités politiques, ce qui conduit à une modification des structures d’échange sur Twitter.

Afin de tenir compte de ce phénomène, nous considérons des ensembles de graphes $\mathcal{G} = \{G_T\}_{T \in \mathcal{T}}$ où \mathcal{T} est un ensemble d’intervalles de temps, $\mathcal{T} = \{T_i\}_{1 \leq i \leq K}$, de telle sorte que $T_i = [t_{min}^i, t_{max}^i]$, t_{min}^i et t_{max}^i étant des entiers positifs représentant les dates dans le système de référence Unix (auquel on se réfère en utilisant la date conventionnelle). Les liens dans un graphe G_{T_i} ne sont définis que par des retweets qui se produisent dans l’intervalle de temps T_i .

Le choix de \mathcal{T} est important : si les intervalles sont trop longs (plusieurs mois), il ne sera pas possible de suivre les reconfigurations des groupes sociaux suite à une agrégation temporelle excessive. Par contre, si les intervalles sont trop courts (par exemple 1 jour), les graphiques correspondants seront presque vides et n’auront aucune signification statistique. Il est donc important de trouver un compromis approprié entre l’adaptabilité temporelle de la méthode de reconstruction des groupes politiques et sa stabilité.

En raison des variations d’opinion politique, résultant de l’enchevêtrement de plusieurs processus dynamiques aux constantes de temps très différentes (par exemple, reconfiguration des courants politiques à l’intérieur d’un parti, reconfiguration des partis eux-mêmes, développement de la scène politique européenne, etc.), il peut être pertinent d’analyser les données à plusieurs échelles de temps différentes.

Par exemple, l’analyse des 5–communautés sur une période de 5 mois du 1er août au 31 décembre 2016 (cf. Fig 1) nous donne une image assez précise du paysage politique préélectoral. Sa pertinence a été confirmée par les alliances ultérieures qui ont eu lieu en 2017 au cours de la campagne. Celles-ci ont respecté la proximité des personnalités au sein de cette carte : Hamon avec Jadot, Macron avec Bayrou, Sarkozy avec Fillon, Le Pen avec Dupont-Aignan.

Pour la présente étude, nous avons choisi une échelle de temps correspondant au temps caractéristique de Twitter et des médias dans le contexte d’une campagne politique. En analysant la distribution temporelle des retweets dans notre corpus, nous avons observé que les événements politiques actuels et l’activité Twitter ont une périodicité hebdomadaire, et qu’en moyenne la période de retweet d’un tweet original est de l’ordre d’une semaine, avec 98.13\$ de retweets émis dans un délai de sept jours (voir Texte D dans S1 File).

Nous avons donc considéré des intervalles de sept jours comme période caractéristique de l’activité politique sur Twitter, et nous avons utilisé des intervalles T_i de deux semaines afin d’obtenir des périodes qui se chevauchent dans la reconstruction temporelle (voir ci-dessous). Dans les analyses suivantes, nous considérons les communautés calculées chaque jour à partir des 14 jours précédents pour assigner une communauté à chaque tweet de cette journée. Pour l’analyse de la dynamique des communautés, nous avons pris le sous-ensemble des communautés calculé tous les lundis, avec $\mathcal{G} = \{G_T\}_{T \in \mathcal{T}}$ où \mathcal{T} est un ensemble d’intervalles temporels, tels que $\forall k \in [2, K - 1], |T_k| = 14 \text{ days}, |T_{k-1} \cap T_k| = |T_k \cap T_{k+1}| = 7 \text{ days}$, lorsque les intervalles sont ordonnés selon leur date de début.

Chaque tweet de notre ensemble de données est donc pris en compte dans exactement deux graphes de retweet pour l’analyse des reconfigurations communautaires. Comme nous le montrons ci-dessous, ce choix permet de suivre de

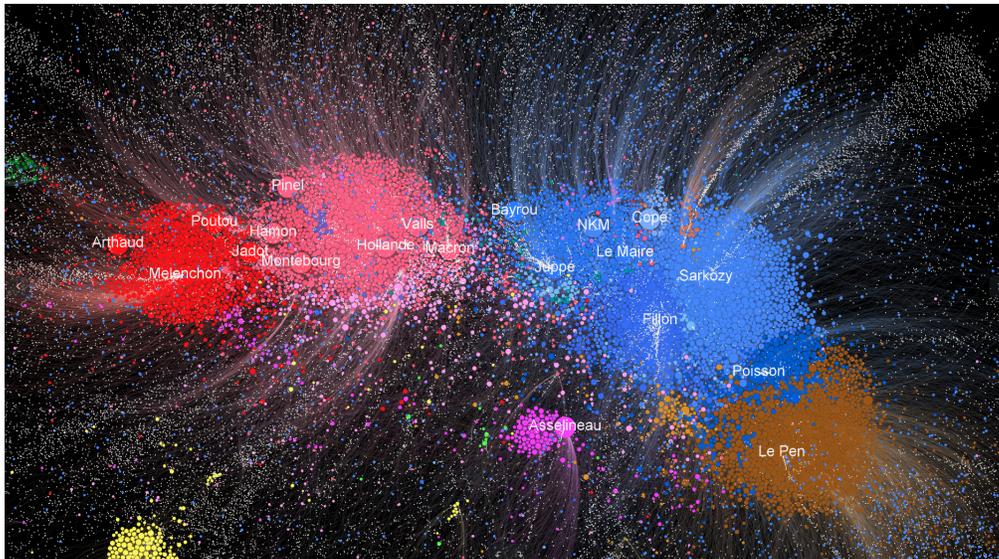


Fig 1. Graphe des 5-communautés de la twittosphère politique pré-électorale française calculée sur la période du 1er août au 31 décembre 2016. Le paysage politique multipolaire est clairement décrit avec une disposition sur un axe gauche-droite de la plupart des partis politiques. Il faut noter que les alliances observées ultérieurement lors de la campagne politique se sont déroulées entre des personnalités politiques proches sur cette carte : Hamon-Jadot, Macron-Bayrou, Le Pen-Dupont-Aignan. La droite (*Les Républicains*, bleu) est divisée en plusieurs courants qui sont reflétés par différentes communautés. La communauté Juppé est à gauche de la droite tandis que les communautés Fillon et Sarkozy sont proches l’une de l’autre et beaucoup plus à droite. Ces deux dernières ont fusionné après les primaires de la droite tandis que la communauté Juppé a pour partie progressivement migré vers la communauté Macron. Le *Parti socialiste* était également composé d’une ”gauche” (communauté Hamon, Montebourg et Pinel) et d’une ”droite” (Hollande, Valls). Ce graphique a été spatialisé avec le logiciel Gephi [36] en utilisant l’algorithme ForceAtlas2 [37]. Voir File A1 sur DOI:10.7910/DVN/AOGUIA pour la version anonymisée de ce graphique.

manière raisonnablement continue les transformations des communautés politiques et de leurs espaces sémantiques. 280

Pour fixer la valeur de w dans la définition de communauté, nous avons choisi de conserver en moyenne les 10% de relations de retweet les plus fortes au cours de T pour l’analyse, ce qui nous a amené à effectuer notre analyse sur les 3-communautés. 281
282
283
284
L’évolution dans le temps du nombre de liens et de comptes de ces à 3-communautés est illustrée dans la Fig 2. 285
286

Analyses détaillées 287

Tous les candidats n’ont pas la même stratégie de communication et certains utilisent relativement peu Twitter. Nous respectons ce choix et nous avons décidé de ne présenter des statistiques détaillées que pour les communautés dont les leaders ont une présence significative sur Twitter. 288
289
290
291

En calculant le nombre de tweets de chaque candidat au 31 mars 2017, nous avons observé un écart entre les six candidats les plus actifs ayant tweeté plus de 500 fois et les autres. Par conséquent, nous conservons ces six communautés pour nos analyses détaillées : Asselineau, Fillon, Hamon, Le Pen, Macron et Mélenchon. Parmi celles-ci, la communauté Asselineau a été classée dans la catégorie des ” petits ” candidats (0,92% des voix au premier tour). Les cinq autres représentent les principales tendances politiques en France : la gauche de la gauche (Mélenchon), la gauche (Hamon), le centre libéral (Macron), la droite (Fillon), et l’extrême-droite (Le Pen). 292
293
294
295
296
297
298
299

Les cinq autres candidats ont reçu moins de 1,5% des votes au premier tour, à 300

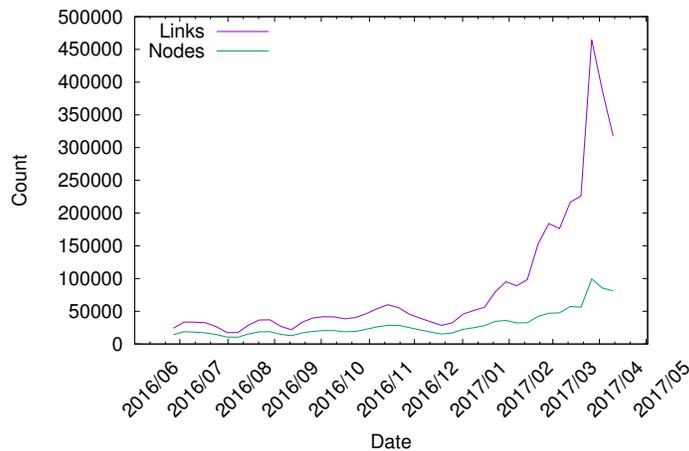


Fig 2. Variation du nombre de nœuds et de liens pour les graphes de G_{T_i} .

l'exception de Dupont-Aignan qui a reçu 4,7% des votes. Le total cumulé de leurs voix au premier tour a représenté 7,82% des votes.

Remarques sur l'interprétation des retweets

Sur Twitter, les marqueurs d'influence sociale liés aux formes d'expression ou d'opinion s'étendent bien au-delà du cadre des retweets. Les individus peuvent par exemple s'influencer mutuellement dans l'utilisation d'expressions ou de hashtags, sans nécessairement se retweeter. De plus, le retweet peut être interprété de différentes manières, avec par exemple une interprétation négative commune lorsque le retweet est tardif (voir Texte J dans S1 File). Cependant, dans la grande majorité des cas, lorsque le retweet est répété dans un contexte politique, il contribue à la propagation des idées et des opinions à travers le réseau [38–40, 40, 41], et peut être considéré comme une forme d'adhésion aux idées d'un candidat ou d'un parti politique [42].

Le suivi de compte est une autre fonctionnalité bien étudiée sur Twitter, qui pourrait aussi servir de proxy pour identifier une certaine forme d'influence sociale. Il est cependant assez rare qu'un utilisateur cesse de suivre un compte, de telle sorte que, de ce point de vue, l'action "suivre" est moins informative que le retweet, qui est intrinsèquement temporel et indique l'affinité d'un utilisateur avec les commentaires faits par un autre compte, à un moment donné. De plus, la fonction "follow" peut être utilisée dans le simple but de se tenir informé des déclarations et commentaires faits par un autre compte, tout en maintenant une certaine distance par rapport à cet utilisateur.

Le retweet a enfin l'avantage de quantifier l'influence sociale entre deux comptes puisque sur une période donnée, il peut exister aucun, peu ou beaucoup de retweets entre deux comptes. Ces relations de retweets peuvent de plus être uni ou bi-directionnelles, information que nous n'exploiterons pas ici.

Enfin, le retweet a l'avantage de quantifier l'influence sociale présente entre deux comptes puisque, pour une période donnée, il peut n'y avoir pas, peu ou beaucoup de retweets entre deux comptes. De plus, ces relations retweet peuvent être mono- ou bidirectionnelles, mais cette information n'est pas prise en compte dans la présente étude.

Caractérisation des thèmes de la campagne présidentielle de 2017

Pour mieux caractériser les communautés politiques, reconstruites uniquement à partir de l'analyse des interactions sociales, nous avons analysé leur profil sémantique. Notre objectif n'est pas de les caractériser de manière exhaustive, mais de donner un aperçu de la façon dont les communautés diffèrent dans leur agenda politique et dans leurs manière de contextualiser l'information.

Pour ce faire, nous avons effectué une analyse du discours politique afin d'extraire les principaux débats de cette campagne. Nous avons utilisé pour cela deux types de sources : les mesures politiques rapportées dans les programmes de campagne, écrites dans un langage contrôlé, et les tweets des candidats, qui empruntent souvent au langage parlé.

Pour l'analyse des programmes des candidats, nous avons d'abord contacté, le 1er mars 2017, toutes les équipes de campagne qui étaient sur le point de recueillir les 500 signatures nécessaires pour pouvoir candidater à l'élection présidentielle (le Conseil constitutionnel n'a publié la liste officielle des candidats que le 20 mars 2017). Nous les avons invités à nous fournir les éléments et les expressions clés qu'ils jugeraient caractéristiques de leur programme. Comme on pouvait s'y attendre, nous n'avons eu presque aucune réponse, seule l'équipe d'un "petit" candidat nous a indiqué qu'étant donné la taille de l'équipe de campagne, elle n'était pas certaine de pouvoir participer. Nous avons ensuite utilisé les programmes des onze candidats téléchargés à partir de leurs sites Web (entre le 14 février et le 24 mars 2017), documents à partir desquels nous avons extrait manuellement 980 mesures politiques (voir Fichier A2 sur le dataverse DOI:10.7910/DVN/AOGUIA pour une version éditable de ces mesures politiques).

Pour compléter ce corpus, nous avons extrait de la base de données *Politoscope* 25.883 tweets publiés par les onze candidats et quelques autres personnalités politiques clés entre août 2016 et janvier 2017 (voir Texte B dans S1 File). Ce deuxième corpus a l'avantage de mettre en évidence les thèmes qui se sont dégagés lors des débats politiques, indépendamment des orientations programmatiques des candidats.

Il existe deux grandes catégories de méthodes traditionnellement utilisées pour l'extraction de sujets à partir de textes non structurés : l'analyse de par co-termes [43] et le *topic modelling* avec des méthodes de type LDA [44]. Dans ces approches, les sujets sont définis comme des "sacs de mots", déduits des statistiques d'apparition au sein des documents de termes issus d'une liste de mots clés définie à l'avance. Cette liste est elle-même obtenue par des méthodes de text-mining plus ou moins avancées dans les domaines du traitement du langage naturel (NLP) et de l'apprentissage machine.

Il a été démontré que le LDA comporte certaines limites pour l'analyse de documents courts [45] ou de corpus de petite taille [46], qui sont deux contraintes présentes dans nos corpus Twitter (messages courts) et de mesures politiques (moins de 1000 documents).

Par conséquent, nous avons analysé ces deux corpus à l'aide du logiciel de fouille de texte du CNRS *Gargantext* (<https://gargantext.org> ; open source à <https://github.com/ISCFIP/gargantext>), qui met en œuvre des méthodes avancées de traitement automatique du langage et de détection de thématiques par l'analyse des co-termes ; ainsi que des méthodes de visualisation pour la représentation des résultats et la navigation interactive.

Dans les premières étapes, *Gargantext* utilise une combinaison de lemmatisation, de post-tagging, d'analyse statistique comme tf-idf [47] et d'analyse des relations de généralité entre termes [48] pour identifier dans le texte quelques milliers de mots clés qui sont spécifiques au discours politique. Ces mots-clés ont ensuite été triés par les auteurs afin de sélectionner les plus significatifs (*i.e.* les mots vides ou les expressions mal formées qui n'auraient pas été filtrée lors des étapes de text-mining ont été supprimés, des hashtags ou néologismes importants de Twitter comme *frexit* ont été ajoutés). Enfin, toutes les mesures politiques ont été relues attentivement avec les

mots-clés sélectionnés mis en évidence dans le texte afin de vérifier qu'il ne manquait aucun mot-clé important. Cela a conduit à un vocabulaire de près de 1600 groupes de mots-clés qualifiant les thèmes de la campagne présidentielle (voir Texte I dans S1 File pour la liste des mots clés).

Nous avons utilisé la mesure de proximité *confidence* pour évaluer la proximité thématique entre les termes sélectionnés. Cette mesure est le maximum entre deux probabilités conditionnelles. Si $P(x|y)$ est la probabilité qu'un document mentionne le terme x sachant qu'il mentionne déjà le terme y , la mesure *confidence* est définie par $\max(P(x|y), P(y|x))$. Il a été démontré qu'il s'agit de l'un des meilleurs choix pour induire automatiquement des relations de noms spécifiques/génériques à partir des de l'analyse co-apparition de termes dans les corpus [49].

Nous avons appliqué l'algorithme de Louvain [35] pour identifier des groupes de termes délimitant des sujets. Enfin, nous avons généré la carte thématique pour chacun de ces deux corpus (cf. Fig 3 pour la carte des programmes présidentiels 2017). Toutes ces étapes de traitement font partie du workflow *Gargantext*.

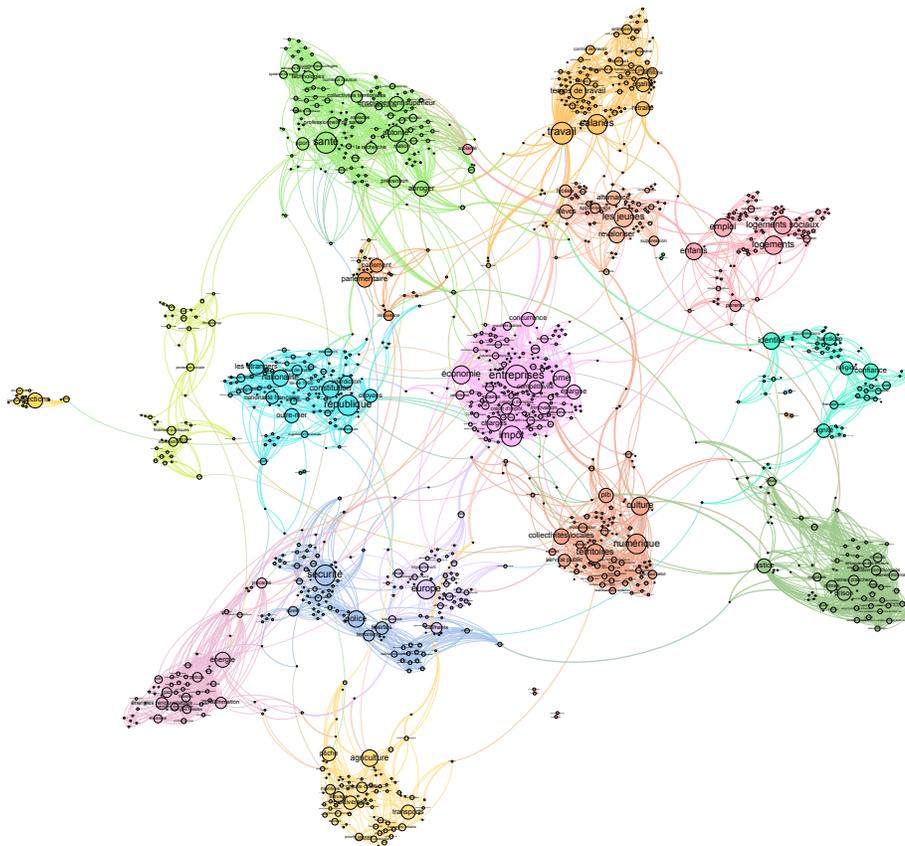


Fig 3. Carte des thèmes couverts par les programmes présidentiels de 2017, établie à partir de *Gargantext*. La carte a été construite à partir des mesures politiques extraites des programmes des candidats. Les nœuds de la carte sont des labels pour des groupes de termes jugés équivalents dans le discours politique. Le lien entre un label A et un label B indique que la probabilité que A et B soient mobilisés conjointement dans la même mesure politique est élevée. *Gargantext* applique l'algorithme de Louvain pour identifier les groupes de labels qui relèvent de la même thématique et leur assigne la même couleur. Pour améliorer la lisibilité, la carte a été éditée dans le logiciel Gephi (<https://gephi.org>) pour définir la taille des nœuds et des labels selon une fonction monotone de leur PageRank [50]. Le Fichier A3 disponible sur DOI:10.7910/DVN/AOGUIA fournit une version éditable de cette carte (gexf).

À partir de ces cartes, nous avons sélectionné onze thèmes que nous avons identifiés
comme particulièrement importants et représentatifs des débats de cette présidentielle.

Validation de la méthode

Afin de valider notre méthode de reconstruction, nous avons vérifié manuellement la
catégorisation politique des comptes pour la journée du lundi 6 février 2017
(communautés calculées sur la période d'activité du lundi 23 janvier au dimanche 5
février 2017). Cette période correspond à la fin de la primaire de la gauche, avant que le
paysage politique ne change suite à des alliances entre candidats (écologistes/Jadot avec
les socialistes/Hamon) ; centre/Bayrou avec En Marche/Macron, DLF/Dupont-Aignan
avec FN/Le Pen). Nous avons considéré pour cela deux échantillons distincts : tous les
comptes suivis actifs ce jour-là (2 440) et un échantillon aléatoire de 2 500 comptes
actifs ce jour-là (dont la grande majorité ne faisaient pas partie des comptes suivis).

Pour ces deux échantillons, nous avons extrait les métadonnées Twitter des comptes
et nous avons codé manuellement l'affiliation politique des comptes chaque fois qu'il y
avait un soutien clair ou une déclaration d'appartenance à un parti dans la description
du compte.

Nous avons associé tous les principaux dirigeants politiques à un parti de référence à
l'aide de la table de correspondance Table. 1.

Table 1. Correspondance entre les leaders politiques et leur parti au 6 février 2017.

Leader	Party
Nathalie Arthaud	LO
François Asselineau	UPR
François Bayrou	MODEM
Jacques Cheminade	S&P
Nicolas Dupont-Aignan	DLF
François Fillon	LR
Benoît Hamon	PS
François Hollande	PS
Yves Jadot	EELV
Alain Juppé	LR
Marine Le Pen	FN
Emmanuel Macron	EM
Jean-Luc Mélenchon	FI
Philippe Poutou	NPA
Nicolas Sarkozy	LR
Manuel Valls	PS

Dans la description de leur compte, certains utilisateurs de Twitter faisaient
référence à des partis qui n'avaient aucun candidat en lice pour cette élection. Nous
avons écarté les mentions qui correspondent à des partis qui n'avaient clairement rallié
aucun candidat en lice (le PCF et le Parti progressiste parmi les plus cités). Nous avons
considéré comme équivalent à *La France Insoumise* (FI) tous les partis qui faisaient
partie de l'ancienne coalition de *Front de Gauche* menée par Mélenchon (avant qu'il
décide de se présenter à la présidence et crée la *France Insoumise*). Compte tenu des
déclarations officielles du parti UDI [51] nous avons également considéré l'UDI (*Union
des Démocrates Indépendants*) comme soutenant *Les Républicains* (LR) au 06 février
2017, bien que cette position n'ait manifestement pas été unanime puisque quelques
semaines plus tard, une douzaine d'élus de l'UDI ont rallié Macron [52].

Cette analyse manuelle a abouti à deux listes distinctes de 885 comptes suivis et de
414 comptes aléatoires respectivement, qui soutenaient publiquement l'un des 11
candidats en lice au moment de la reconstruction des communautés politiques du 6
février 2017 (les autres comptes dans les échantillons ne donnaient pas suffisamment
d'information dans leur profil pour identifier leur affiliation politique avec certitude).

Nous avons ensuite comparé la catégorisation automatique faite sur cette période avec notre codage manuel pour évaluer la précision et le rappel de notre algorithme. Nous avons limité cette comparaison aux comptes ayant fait au moins 10 tweets au cours de la période d'observation de 14 jours.

Bien que nous puissions nous attendre à un biais de sélection dû au fait que les personnes les plus engagées politiquement sont à la fois les plus susceptibles de déclarer une affiliation politique dans leur profil Twitter et sont plus facilement catégorisées par des méthodes automatisées, nous considérons ceci comme une limitation raisonnable à la précision de nos estimations de notre la qualité de notre reconstruction. Il convient également de souligner qu'il n'y a pas de référence absolue en ce qui concerne les affiliations politiques d'un groupe de personne puisque en permanence une proportion importante de la population est indécise ou en train de changer d'opinion. Il y aura donc toujours une incertitude intrinsèque dans la catégorisation des comptes, quelle que soit la méthodologie utilisée.

Pour compléter cette validation, nous avons comparé les reconfigurations clés mises en évidence dans la reconstruction du paysage politique avec les événements majeurs de l'actualité (voir la section résultats).

Résultats

Précision et rappel de la reconstruction du paysage politique

La comparaison entre la catégorisation à révélé une très bonne correspondance entre la catégorisation automatique et le codage manuel des affiliations politiques : une précision supérieure à 93 % pour les deux échantillons et un rappel supérieur à 63 % pour les deux échantillons.

Les tables 2, 3, 4 and 5 montrent le rappel, la précision et les intervalles de confiance pour chaque parti politique et chaque échantillon.

La précision et le rappel sont semblables d'un parti politique à l'autre et entre les échantillons, à deux exceptions près.

Le FN (extrême droite) a un rappel plus élevé pour les comptes suivis (échantillon 1). Cela indique que les militants du FN expriment leurs opinions politiques en ligne plus ostensiblement que tous les autres militants.

En Marche ! (le nouveau parti de Macron) a une très faible précision pour l'échantillon des comptes suivis (7.14% contre une moyenne de 93.54%). Cela signifie que ces comptes n'ont pas mentionné leur soutien à la *En Marche !* dans leur profil, mais que leur comportement en ligne pouvait être analysé comme relayant des informations pro-Macron. En effet, la liste des comptes suivis était basée sur les personnalités politiques élues avant la création du mouvement *En Marche !*. Au moment de la capture de l'échantillon (23/01/06-02/2018), le mouvement de Macron qui en était à ses début prenait de l'ampleur. Il commençait à recruter des sympathisants d'autres partis, mais apparemment pas au point que ceux-ci aillent jusqu'à exprimer ouvertement leur soutien dans leur profil Twitter.

Évolution de communautés politiques

Une autre validation externe de notre approche vient de sa capacité à suivre avec une grande précision temporelle les changements du paysage politique français rapportés par les médias.

Les campagnes politiques sont faites de retournements, coups de théâtre et reconfigurations des soutiens politiques. Les commentateurs s'efforcent de les anticiper

Table 2. Rappel pour les comptes suivis qui étaient actifs entre le 23 janvier et le 6 février 2017 (au moins 10 tweets émis).

	Recoupement auto vs. manuel	Étiqueté manuellement	rappel	Interval de confiance (95%)
LR	162	265	61.13%	[-5.99%, +5.67%]
PS	132	215	61.40%	[-6.65%, +6.25%]
FN	47	50	94.00%	[-10.22%, +3.94%]
FI	18	36	50.00%	[-15.53%, +15.52%]
MODEM	2	3	66.67%	[-45.90%, +27.18%]
EM	1	2	50.00%	[-40.55%, +40.55%]
Total	362	571	63.40%	[-4.03%, +3.85%]

Table 3. Précision pour les comptes suivis qui étaient actifs entre le 23 janvier et le 6 février 2017 (au moins 10 tweets émis).

	Étiquetés de manière consistante	Étiquetés auto.	Précision	Interval de confiance
LR	164	165	99.40%	[-2.75%, +0.50%]
PS	132	138	95.65%	[-4.81%, +2.34%]
FN	47	47	100.00%	[-7.56%, +0.00%]
FI	18	18	100.00%	[-17.59%, +0.00%]
EM	1.0	14	7.14%	5.87%, +24.32%]
DLF	0	1	0.00%	[0.00%, +79.35%]
EELV	0	4	0.00%	[0.00%, +48.98%]
Total	362	387	93.54%	[-2.90%, +2.04%]

Table 4. Rappel pour les comptes de l'échantillon aléatoire qui étaient actifs entre le 23 janvier et le 6 février 2017 (au moins 10 tweets émis).

	Recoupement auto vs. manuel	Étiqueté manuellement	Rappel	Intervalle de confiance (95%)
LR	69	122	56.56%	[-8.86%, +8.46%]
PS	33	60	55.00%	[-12.51%, +11.91%]
FN	66	86	76.74%	[-9.95%, +7.67%]
FI	37	52	71.15%	[-13.43%, +10.52%]
EM	31	51	60.78%	[-13.70%, +12.19%]
UPR	7	8	87.50%	[-34.59%, +10.26%]
EELV	5	6	83.33%	[-39.68%, +13.66%]
S&P	1	1	100.00%	[-79.35%, +0.00%]
DLF	3	3	100.00%	[-56.15%, +0.00%]
MODEM	0	3	0.00%	[0.00%, +56.15%]
PP	0	1	0.00%	[0.00%, +79.35%]
Total	252	393	64.12%	[-4.86%, +4.58%]

Table 5. Précision pour les comptes de l'échantillon aléatoire qui étaient actifs entre le 23 janvier et le 6 février 2017 (au moins 10 tweets émis).

	Coherently tagged	Auto. tagged	Precision	Confidence Interval
LR	69	71	97.18%	[-6.88%, +2.04%]
PS	33	39	84.62%	[-14.34%, +8.14%]
FN	66	66	100.00%	[-5.50%, +0.00%]
FI	37	39	94.87%	[-11.76%, +3.71%]
EM	31	32	96.88%	[-12.62%, +2.57%]
UPR	7	7	100.00%	[-35.43%, +0.00%]
EELV	5	5	100.00%	[-43.45%, +0.00%]
S&P	1	1	100.00%	[-79.35%, +0.00%]
DLF	3	4	75.00%	[-44.94%, +20.44%]
Total	252	264	94.55%	[-3.23%, +1.93%]

et de les analyser tandis que les partis politiques s'évertuent à masquer ce qui se trame en coulisse pour prendre leurs adversaires de court.

Le principal résultat de notre étude est qu'une analyse automatique des motifs de retweets est en mesure de révéler, parfois avant que la presse n'en parle, l'évolution des communautés politiques face à ces événements et, dans certains cas, face à leur anticipation.

Ces changements peuvent être d'une grande ampleur sur un très court laps de temps. Par exemple, la comparaison entre le paysage politique des 3-communautés la veille du premier tour (Fig 4) et celui de la veille du second tour (Fig 5) montre une réorganisation complète des communautés, certaines ayant connu des scissions en cours de route.

Dans le but de suivre les reconfigurations des communautés politiques au fil du temps, nous avons comparé les communautés deux à deux, sur des intervalles de temps consécutifs. Cette approche permet de suivre de manière continue les transformations des communautés politiques.

Entre chaque période, pour chaque communauté, nous avons calculé le nombre de comptes ayant changé de communauté et généré la structure globale synthétisant l'évolution du paysage politique. Cette structure peut être visualisée à l'aide d'un diagramme alluvial (cf. Fig 6), une visualisation qui résume les reconfigurations de l'espace politique sur une période de 11 mois, avec une résolution suffisante pour proposer une narration précise de cette élection présidentielle (Fig 7). On voit que si la plupart des grands événements politiques peuvent être interprétés via cette représentation, elle met également en avant des relations plus subtiles entre communautés qui n'ont pas l'objet de déclarations officielles, contrairement aux ralliements ou aux victoires aux primaires. Par exemple, peu avant l'affaire Penelopegate (voir analyse ci-dessous), on observe une prise de distance de la communauté Sarkozy par rapport à celle de Fillon suivie du ralliement à Dupont-Aignan, quelque semaines plus tard d'un nombre important de militants Sarkozy. Ce dernier ralliement a été relativement éphémère, puisque la plupart de ces comptes ont cessé de participer à la sphère informationnelle de Dupont-Aignan lorsque, entre les deux tours, il s'est lui-même rallié à Le Pen .

Les nombreux épisodes que l'on peut décrire par cette visualisation alluviale sont autant de preuves qu'une analyse des tweets politiques de Twitter permet de reconstruire avec succès la dynamique du paysage politique : les événements clés de la campagne présidentielle correspondent à des éléments clés de la reconstruction, et donc de la visualisation (comme la fusion ou la scission des communautés, le changement brutal de la taille des communautés, etc). Entre autres, les événements suivants peuvent facilement être identifiés : fusion de communautés $[M]$ scission $[S]$, émergence $[E]$ ou changement brutal de taille de certaines communautés. Chacun a été associé à une une de média (cf. Fig 7 où chacun des événements suivants a été identifié) :

1. **2016/11/27-2016/12/01 [53]**: reconfiguration de la droite après les primaires. Ralliement après négociations de Sarkozy à Fillon et disparition de la communauté de Juppé,
2. **2016/12/01 [54]**: effet de levier du renoncement de Hollande sur la campagne de Macron,
3. **2017/01/17 [55]**: $[S]$ Frictions entre *les Républicains* à la suite de contestations de certaines investitures aux législatives, dont celle de Nathalie Kosciusko-Morizet (NKM), à qui Fillon offre une circonscription facile à gagner,
4. **2017/01/25 [56]**: $[M + S + E]$ Début du Pénélope Gate, quelques fillionnistes rejoignent la communauté Sarkozy, la communauté Juppé réapparaît,

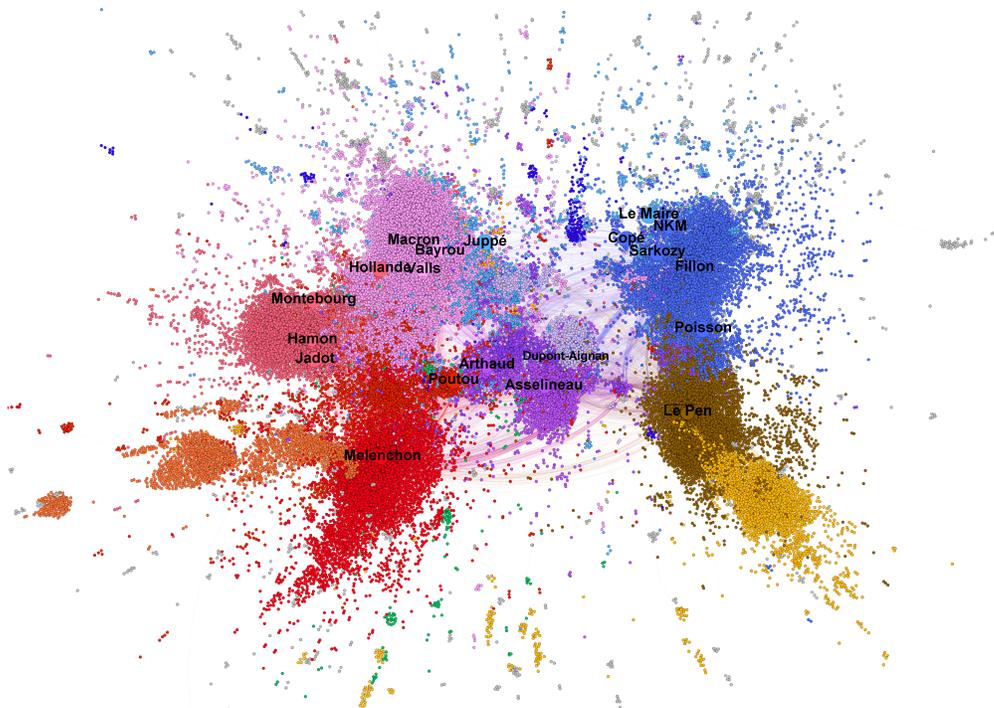


Fig 4. 3-communautés du paysage politique français le jour du 1er tour (T=[9 avril 2017 - 23 avril 2017]). Le paysage est très morcelé et multipolaire, les principales forces politiques étant représentées (les pourcentages indiqués correspondent aux scores obtenus au premier tour) : Mélenchon et la *France Insoumise* (rouge vif - 19.58%), Hamon et le *Parti Socialiste* (rouge clair - 6.36%), Macron et *En Marche !* (rose - 24.01%), Fillon et *Les Républicains* (bleu - 20.01%), Le Pen et le *Front National* (marron - 21.30%), Dupont-Aignan et *Debout la France* (lilas - 4.7%). Pour les “petits” candidats, Asselineau (0.92%) et Arthaud (0.64%) se retrouvent dans la même communauté (violet), Poutou (1.09%) se retrouve avec Mélenchon mais n’a pas vraiment de communauté Twitter. On remarquera que non seulement Bayrou est dans la communauté Macron, ce qui est normal après sont ralliement le 23 février 2017, mais que Juppé, ex-candidat à la primaire *Les Républicains*, est beaucoup plus proche de la communauté Macron que de celle de Fillon et Sarkozy. Son “lieutenant” Edouard Philippe, deviendra premier ministre de Macron. Deux communautés non labellisées par un leader politique sont néanmoins importantes. En orange à gauche, proche de la communauté Mélenchon on retrouve une communauté identifiée comme proche de la *Discorde Insoumise*, communauté du jeu vidéo qui a joué un rôle important dans la campagne de Mélenchon. En jaune en bas à droite, proche de la communauté Le Pen, nous retrouvons une communauté de comptes populistes et nationalistes anglophones qui a fortement soutenu Le Pen. Elle est elle-même composée de sous-communautés, en particulier des supporters anglophones de Le Pen, des supporters de Trump et des nationalistes UK pro-Brexit. On relèvera également des différences importantes avec la carte pré-campagne présidentielle de 2016 (figure 1) : en plus de l’axe “gauche-droite” qui était alors dominant, un nouvel axe semble important (“haut-bas” sur la carte) qui distingue les partis à tendance nationaliste ou protectionniste des partis adhérant à la mondialisation et à la globalisation des échanges économiques (le fait que les petits candidats soient au “centre” de la carte ne doit pas être interprété comme une forme de centralité. Seules les contiguïtés topologiques sont significatives sur ce type de graphe. Les communautés des “petits” candidats étant relativement peu connectées aux autres se retrouvent au centre du fait de l’algorithme de spatialisation du graphe) Le Fichier A4 disponible sur DOI:10.7910/DVN/AOGUIA donne une version éditable de ce graph.

5. **2017/01/29 [57]:** $[C + M + S]$ Deuxième tour des primaires de gauche remporté par Hamon, 527
528
6. **2017/02/23 [58]:** $[M]$ Bayrou se rallie à Macron, 529
7. **2017/02/26 [59]:** $[M]$ Jadot se rallie à Hamon, 530
8. **2017/03/05-06 [60]:** $[S+M+C]$ François Fillon organise un meeting au Trocadéro (Paris). Il annonce 200 000 participants (y compris les principaux 531
532

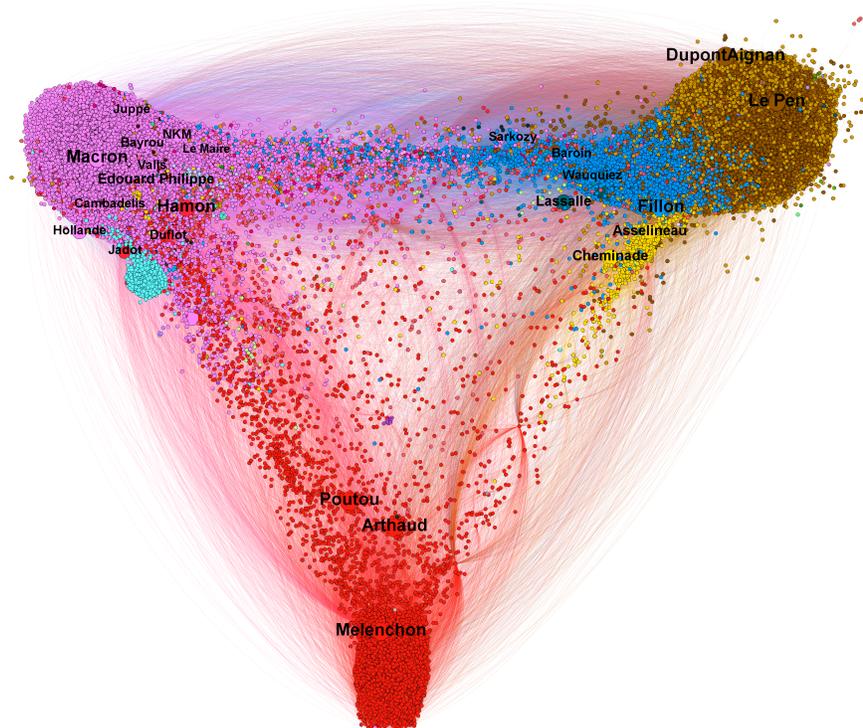


Fig 5. 3-communautés du paysage politique français entre les deux tours de scrutin (T=[27 avril 2017 - 7 mai 2017]). Avec seulement deux candidats restants (Macron et Le Pen), les motifs du militantisme politique sur Twitter ont radicalement changé avec deux pôles correspondant aux partisans des deux candidats restants (issus de la recombinaison des anciennes communautés) et un troisième pôle constitué par ceux qui refusaient de prendre parti. Ces derniers se positionnaient pour être la prochaine force d’opposition du futur gouvernement. Il est à noter que le parti *Les Républicains* s’est scindé dans l’entre-deux tours, les ténors de ce parti étant dispersés entre les sphères informationnelles de Macron et de Le Pen. Le Fichier A5 disponible sur DOI:10.7910/DVN/AOGUIA fournit une version éditable de ce graphique.

soutiens de Sarkozy) et présente ceci comme un soutien au maintien de sa candidature, les sarkozistes soutiennent le retour de Juppé mais ce dernier renonce finalement publiquement à toute perspective de candidature présidentielle. Confirmation de la légitimité de Fillon par le bureau politique de *Les Républicains*, qui avait été initialement convoqué pour demander sa démission [61],

9. **2017/03/23 [62]:** [M] Dupont-Aignan prétend que “de nombreux sarkozistes se joignent à [son] mouvement”,
10. **2017/04/23 [63]:** Premier tour de l’élection présidentielle mené par Macron et Le Pen. Mélençon refuse de donner des instructions de vote à ses partisans, contrairement à ce qu’il avait fait en 2002 (il avait appelé à voter contre le *Front national*),
11. **2017/04/28 [64]:** [M] Dupont-Aignan se rallie à Le Pen.

Nous analysons ces événements clés plus en détail dans les Figs 8, 9, 10, 11 et 12.

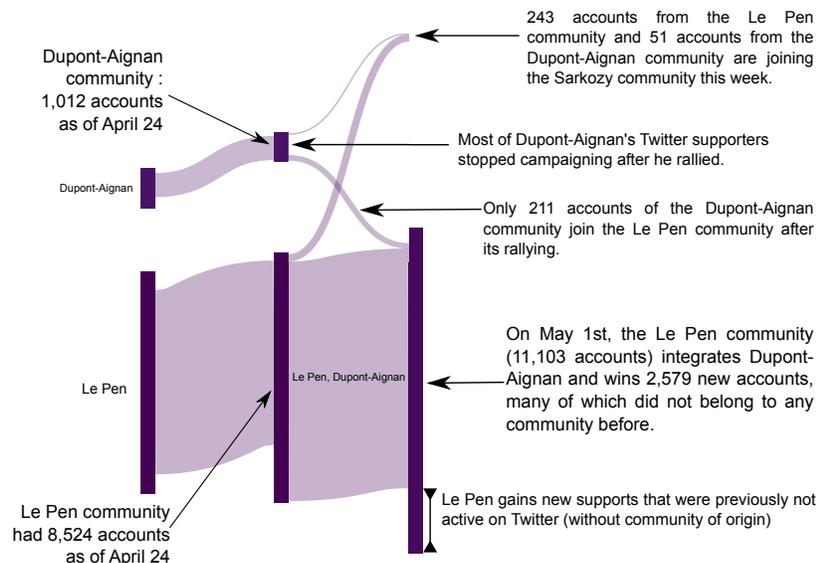


Fig 6. Représentation de l'évolution des communautés par un diagramme alluvial

La méthode de reconstruction que nous proposons met en évidence les profondes reconfigurations que peut subir un paysage politique au cours d'une campagne présidentielle dans un environnement politique multipolaire. En décrivant le comportement d'un ensemble de militants, correspondant à un large éventail d'affiliations politiques, elle se situe à un niveau intermédiaire entre les dirigeants politiques et leur base électorale.

Cette approche donne accès à un niveau d'observation rarement accessible dans le cadre d'études sociologiques ou en sciences politiques. Au total, entre le 1er août 2016 et le 8 mai 2017, près de 190 000 comptes individuels ont été associés à une communauté politique, la plupart des communautés impliquant plusieurs dizaines de milliers de militants (voir tableau C dans S1 File). Ces communautés sont composées d'un grand nombre de leaders d'opinion et de partis politiques. À partir de ces résultats, on peut raisonnablement émettre l'hypothèse que les tribulations politiques des sous-populations étudiées par le Politoscope sont des indicateurs, et dans certains cas les précurseurs, du changement et de la reconfiguration du paysage politique.

Dans les sections suivantes, nous avançons quelques idées et posons des questions ouvertes sur ce qu'il est possible de tirer de ces reconstructions, en plus d'un récit précis et synthétique d'une campagne politique. Nous explorons deux caractéristiques des communautés : leur structure sociale et leurs profils sémantiques.

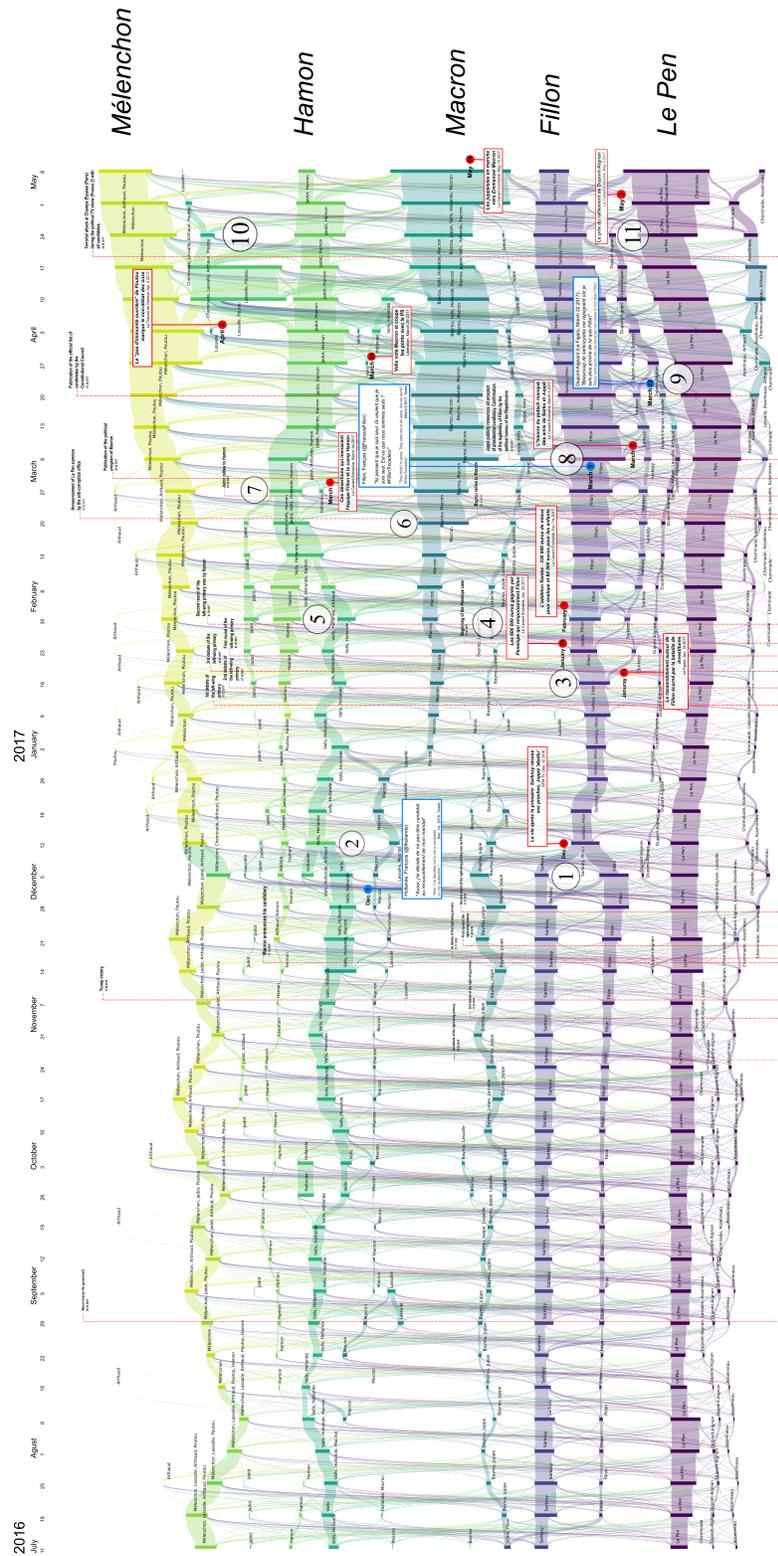


Fig 7. Reconfigurations des communautés politiques entre juin 2016 et avril 2017. Certains grands titres de journaux ont été ajoutés pour apporter un éclairage sur les bifurcations observées. Les labels des communautés encore actives lors du second tour ont été ajoutés en fontes larges pour une meilleure visibilité. Chaque barre verticale correspond à une communauté politique sur Twitter labellisée par les comptes de candidats qui y participent. La hauteur est proportionnelle au nombre de comptes Twitter faisant partie de cette communauté. Celles-ci sont calculées tous les lundis.

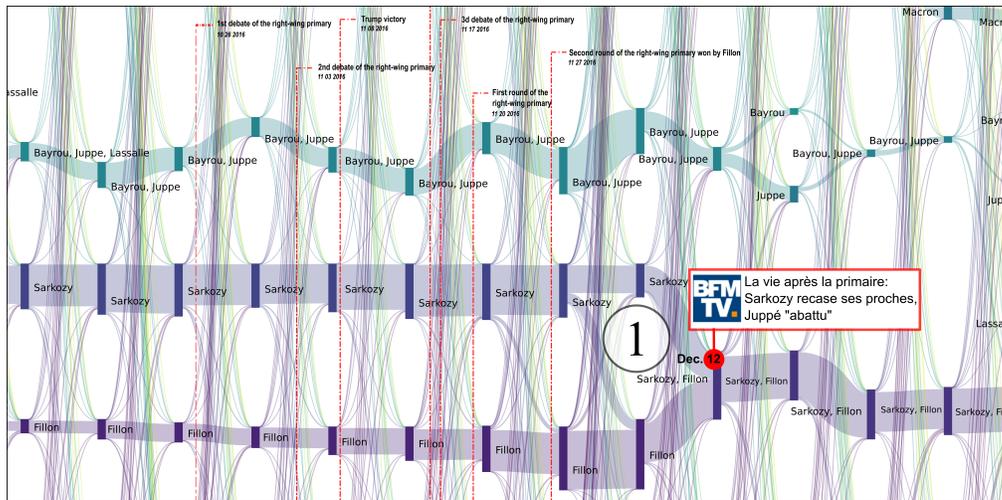


Fig 8. Reconfiguration des communautés politiques au moment des primaires de la droite. Les primaires de la droite ont été rythmées par trois débats et deux votes. On peut observer les renforcements progressifs et réguliers de la communauté Fillon semaines après semaines, alors que personne ne le pronostiquait au second tour de la primaire, Juppé étant le favori des sondages. A l'issue du premier tour, de la victoire surprise de Fillon et de l'élimination également surprenante de Sarkozy (président du parti *Les Républicains* et ex-Président de la République), seule une infime partie de la communauté Sarkozy a rejoint la communauté Fillon. La communauté Sarkozy trainera des pieds pour se rallier à la communauté Fillon et Sarkozy négociera pendant plusieurs jours la place de ses membres dans le dispositif Fillon. La communauté Juppé, en revanche, prendra immédiatement ses distances par rapport aux nouvelles orientations politiques prises par le parti. Cette ligne de fracture de la droite était apparente dans le paysage politique pré-électoral (cf. figure 1). Après l'échec du retour de Juppé en mars 2017 (cf. figure 11), une partie de la communauté Juppé rejoindra Macron. Les bifurcations observées ici sont bien résumées par le titre d'un article de BFMTV du 12 décembre 2017 : *La vie après la primaire : Sarkozy recase ses proches, Juppé "abattu"*. Chaque barre verticale correspond à une communauté politique sur Twitter labellisée par les comptes des candidats qui y participent. La hauteur est proportionnelle au nombre de comptes Twitter faisant partie de cette communauté. Celles-ci sont calculées tous les lundis.

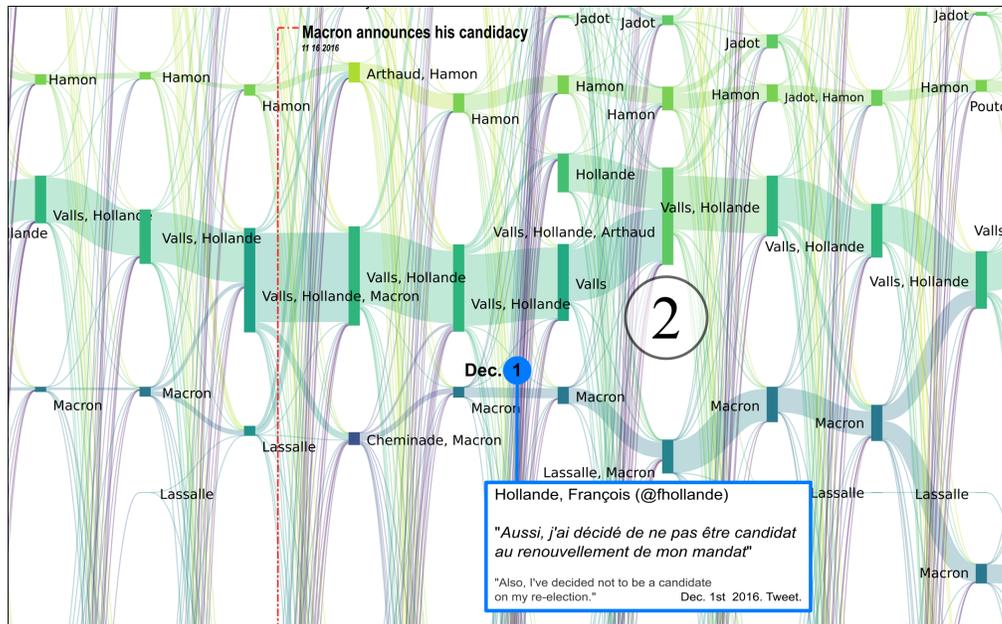


Fig 9. Reconfiguration des communautés politiques au moment où le Président de la République François Hollande annonce qu'il ne se représentera pas. Cette déclaration a eu lieu peu après que Macron ait annoncé sa candidature. Il apparaît clairement sur le graphe alluvial que la non candidature de Hollande a dynamisé la communauté de Macron au détriment de celle de Valls. Alors que la taille de la communauté Macron était restée stable au moment de l'annonce de sa candidature (427 comptes la semaine précédente, 561 la semaine de l'annonce, 472 la semaine suivante), elle a bondit à 749 comptes la semaine du renoncement de Hollande, puis 1,479 comptes la semaine suivante. Elle a dès lors été en croissance régulière jusqu'aux élections. Cette évolution est concomitante avec une perte de vitesse de la communauté Valls-Hollande, aile droite du *Parti Socialiste* la plus proche de Macron d'un point de vue idéologique. La communauté Hamon, plus à gauche, est en revanche restée relativement stable sur cette période. Chaque barre verticale correspond à une communauté politique sur Twitter labellisée par les comptes des candidats qui y participent. La hauteur est proportionnelle au nombre de comptes Twitter faisant partie de cette communauté. Celles-ci sont calculées tous les lundis.

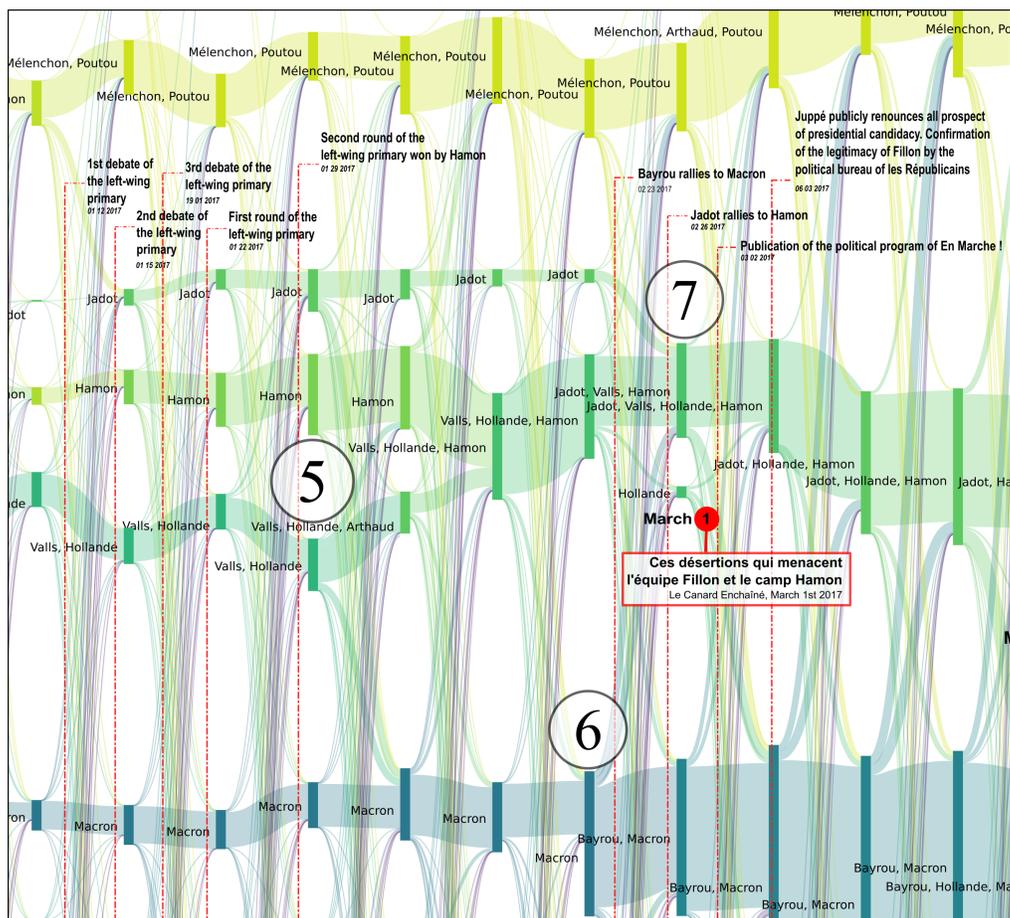


Fig 10. Reconfiguration des communautés politiques au moment des primaires de la gauche. Benoît Hamon (36,03% au 1er tour et 58,69% au second tour) et Manuel Valls (31,48% au 1er tour et 41,31% au second tour) étaient les deux favoris de ces primaires et représentaient deux courants du parti socialiste qualifiés par Valls de “gauches irréconciliables” lors de sa campagne. Bien que François Hollande ne se soit pas exprimé officiellement en faveur d’un candidat particulier, Valls affirmait que celui-ci le soutenait (cf. par exemple *Le Monde* du 4 janvier 2017 “Selon Valls, Hollande le soutient”). En effet, sur notre reconstruction, la communauté Hamon est, pendant toute notre période d’observation, bien distincte de la communauté Valls. Cette dernière est quant à elle, la plupart du temps également la communauté de Hollande, notamment au moment des primaires, indiquant une proximité idéologique entre les deux personnalités politiques. Après la victoire de Hamon, une polémique éclata sur l’engagement pris par les candidats à cette primaire de se ranger derrière le vainqueur alors que certains affichaient clairement leur préférence pour Macron. Ce flottement est clairement visible sur le graphe alluvial avec des mouvements importants de militants de la communauté Valls-Hollande vers la communauté Macron dès le lendemain du deuxième tour de la primaire. Sur les 2075 comptes que comportait la communauté Valls-Hollande juste avant sa fusion avec la communauté Hamon le 13 février (Valls rejoignant également la communauté Hamon), seuls 959 comptes resteront. Des mouvements importants continueront d’avoir lieu entre ces deux communautés jusqu’au premier tour de la présidentielle. Si Valls finira par soutenir officiellement Macron le 29 mars, il n’apparaissait déjà plus dans la communauté Hamon dès le 13 mars. Il apparaît dans la communauté Macron à partir du 17 avril et jusqu’au deuxième tour, indiquant un soutien actif de sa part. On remarquera également sur ce zoom le ralliement de l’écologiste Jadot à Hamon le 23 février 2017, qui se manifeste immédiatement par la fusion de sa communauté avec celle de Hamon. Chaque barre verticale correspond à une communauté politique sur Twitter labellisée par les comptes des candidats qui y participent. La hauteur est proportionnelle au nombre de comptes Twitter faisant partie de cette communauté. Celles-ci sont calculées tous les lundis.

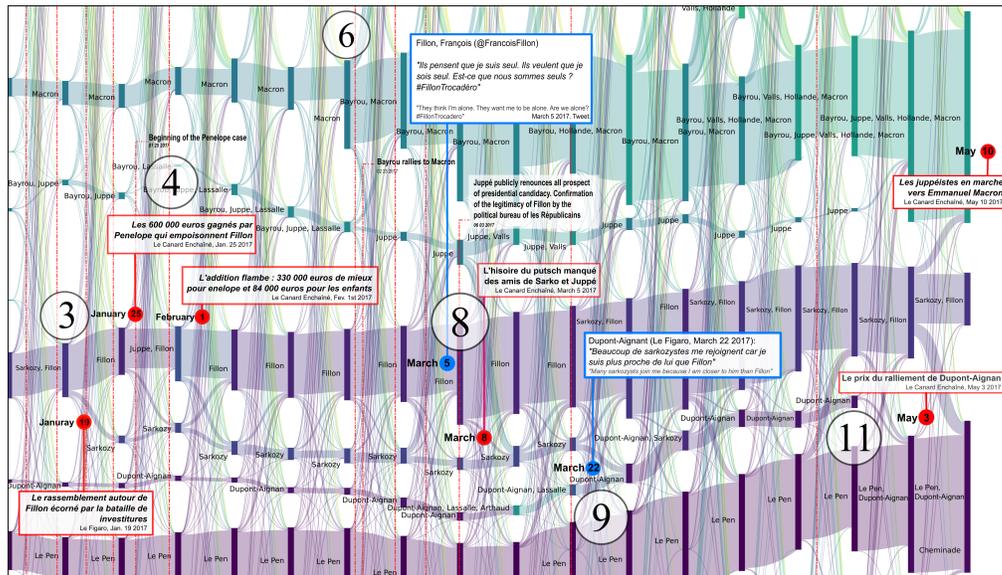


Fig 11. Reconfiguration des communautés politiques au moment du Pénelope Gate. Le Pénelope Gate arrive à un moment où le rassemblement autour de Fillon est fragilisé en raison d’une polémique sur les investitures aux législatives, mal vécues par les sarkozystes. Une partie de ceux-ci avait déjà commencé à se détacher de la communauté *Les Républicains* dès le 23 janvier 2017 et formeront une communauté à part pendant toute la polémique. Les différentes révélations du *Canard Enchaîné* provoquent de larges mouvements de militants entre différentes communautés, les communautés autour de Juppé (potentielle alternative à la candidature Fillon ou “plan B”) et Sarkozy se renforçant progressivement. Au 1er mars, jour de la convocation de Fillon en vue d’une mise en examen, une grande partie des Sarkozystes soutiennent le retour de Juppé dans un ultime effort pour pousser Fillon à la démission, ce qui se manifeste par un flux inhabituel de comptes de la communauté Sarkozy vers la communauté Juppé. Le rassemblement du Trocadéro le 5 mars met fin aux perspectives de plan B. La communauté Fillon gagne beaucoup de nouveaux militants suite à ce coup de force mais une partie des sarkozistes et des juppéistes restent à l’écart et rejoindront respectivement, Dupont-Aignan et Macron. Sarkozy quant à lui, restera plusieurs semaines en retrait, ce qui se traduit par son absence dans la plupart des communautés Fillon entre le 16 janvier et le 10 avril. Ce n’est qu’à partir du 7 avril, qu’il soutiendra de nouveau activement la candidature Fillon, comme le mentionne par exemple le titre du *Parisien* de ce jour “Présidentielle : Sarkozy soutient (enfin) Fillon” et ce qui se manifeste par sa présence permanente dans la communauté Fillon à partir du 10 avril 2017. Comme on peut le constater, les bifurcations observées dans le Politoscope sont concomitantes et même parfois anticipent les commentaires de la presse sur l’activité des communautés politiques. Chaque barre verticale correspond à une communauté politique sur Twitter labellisée par les comptes des candidats qui y participent. La hauteur est proportionnelle au nombre de comptes Twitter faisant partie de cette communauté. Celles-ci sont calculées tous les lundis.

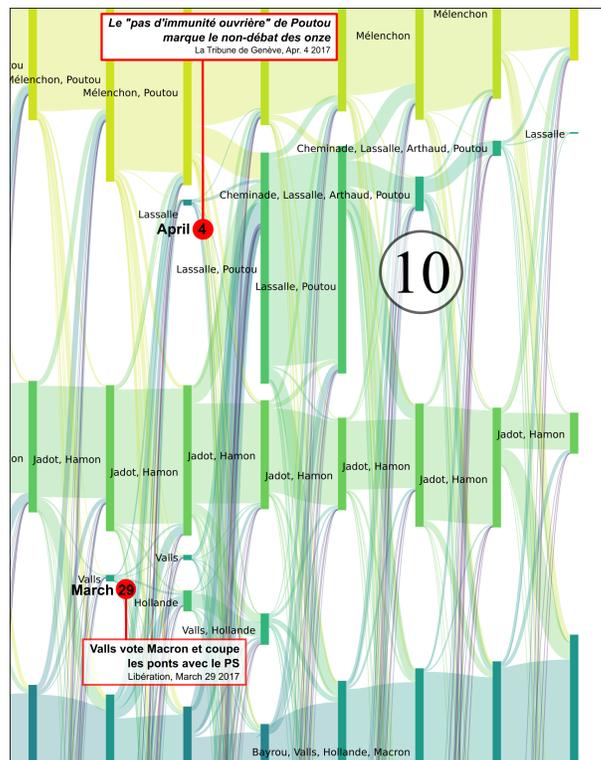


Fig 12. Reconfiguration des communautés politiques au moment du “Grand Débat” télévisé du 4 avril 2017. Philippe Poutou a créé le buzz lors de cette confrontation en dénonçant ouvertement la triche et la corruption des candidats Fillon et Le Pen et en critiquant les positions anti-système : “Quand nous on est convoqué par la police, on n’a pas d’immunité ouvrière, on y va.” Son intervention a été relayée par la plupart des médias et a recueilli plus d’un million de vues en ligne. Sa communauté Twitter, quasi inexistante jusque là et souvent fusionnée avec celle de Mélenchon, s’est développée considérablement (> 13,000 comptes) sur une courte période suite à cette intervention.

Caractérisation de la structure des communautés et de l'engagement politique

Stabilité et composition des communautés

Les communautés politiques peuvent évoluer en gagnant ou perdant des membres qui ne comptaient pas auparavant parmi les comptes politisés sur Twitter ou en gagnant l'adhésion de membres appartenant à d'autres communautés. Notre reconstruction du paysage politique décrit la co-évolution entre la structure d'une communauté et les opinions de ses membres. Les communautés régissent la circulation de l'information et les membres d'une communauté donnée sont exposés à un contenu spécifique. Cependant, cette exposition peut contribuer à modifier leur opinion et éventuellement les amener à se tourner vers une autre communauté, modifiant ainsi la structure globale des communautés.

Un défi important pour comprendre l'évolution d'un paysage politique est de comprendre quelles sont les particularités structurelles de chaque communauté et comment celles-ci peuvent influencer leur évolution.

Nous avons analysé la stabilité et la composition des communautés des 6 candidats qui ont le plus utilisé Twitter dans leur campagne. Un candidat donné c peut apparaître dans des communautés à différentes périodes de temps sur un ensemble de périodes \mathcal{T} . Puisque ces 6 candidats ont toujours eu des communautés distinctes, par souci de clarté, nous désignerons par "la communauté de c " l'ensemble des communautés auxquelles c a appartenu au cours des périodes de temps \mathcal{T} (noté $C_c^{\mathcal{T}}$).

De plus, afin de s'assurer que les indicateurs calculés sur les communautés sont comparables, nous les avons calculés sur l'ensemble \mathcal{T} des périodes du 1er février au 22 avril 2017. Le 1er février est le jour où tous les noms des candidats à la présidence ont été confirmés. Le 22 avril 2017 est la veille du premier tour.

L'analyse de la distribution cumulative du temps passé dans une communauté donnée (voir Texte E dans S1 File) montre que pour la plupart des communautés (sauf Asselineau qui était un "petit" candidat avec 0,92% des votes), environ 15% des comptes ont présents pendant seulement 1 jour, et plus de 10% des comptes étaient membre d'une communauté plus de 40 jours.

Il existe donc une structure communautaire politique générique sur Twitter, composée d'un noyau de militants qui restent plusieurs semaines au sein de la communauté, d'une périphérie beaucoup plus volatile et d'un ensemble intermédiaire de comptes couvrant un large spectre d'implication militante.

Ceci n'est cependant qu'une description très partielle de ce qui caractérise la structure d'une communauté politique et l'engagement politique de ses membres. Pour affiner l'analyse des communautés, nous proposons plusieurs indicateurs qui contribuent à caractériser quantitativement les types et le degré d'engagement politique de leurs membres.

Indicateurs d'engagement politique

Soit $C_c^{\mathcal{T}}$ la communauté d'un candidat c sur l'ensemble des périodes \mathcal{T} . Nous pouvons calculer le nombre de tweets catégorisés dans cette communauté et le nombre de comptes ayant été impliqués au moins un jour dans cette communauté.

La *contribution journalière* (*day contribution*) d'un compte a à chaque communauté $C_c^{\mathcal{T}}$ est le nombre de jours pendant lesquels a a été impliqué dans $C_c^{\mathcal{T}}$ sur la période \mathcal{T} . En moyenne sur une communauté, nous obtenons le nombre moyen de jours pendant lesquels les comptes contribuant à $C_c^{\mathcal{T}}$ ont été impliqués dans $C_c^{\mathcal{T}}$ (noté $\langle days \rangle$). La distribution cumulative des jours passés au sein d'une communauté est donnée par la figure F dans S1 File.

Dans l'ensemble, sur la période \mathcal{T} , dans les communautés Twitter des 6 leaders qui ont utilisé le plus intensément ce media pour leur campagne (resp. toutes les communautés Twitter), il y a eu 18.018.417 de tweets émis (resp. 19.483.583) par 92.762 comptes distincts (resp. 105.418) avec un nombre moyen de tweets par compte de 184,82 (resp. 194,24).

Définissons **l'engagement (commitment)** d'un compte a auprès un candidat c comme le pourcentage des tweets communautaires de a qui appartiennent à $C_c^{\mathcal{T}}$: $co(a, c, \mathcal{T}) = \frac{N_{tw}^{\mathcal{T}}(a, c)}{N_{tw}^{\mathcal{T}}(a)}$, where $N_{tw}^{\mathcal{T}}(a, c)$ est le nombre de tweets émis par a dans $C_c^{\mathcal{T}}$ et $N_{tw}^{\mathcal{T}}(a)$ est le nombre total de tweets émis par a quelque soit la communauté sur la même période. Cet indice indique dans quelle mesure a contribue à la circulation des idées dans la communauté de c lorsque a est impliqué dans des interactions avec des groupes politiques. L'engagement moyen (average commitment) des membres d'une communauté $C_c^{\mathcal{T}}$ est alors défini comme la moyenne des engagements des comptes participant à $C_c^{\mathcal{T}}$: $\langle co(C_c^{\mathcal{T}}) \rangle = \frac{\sum_{a \in C_c^{\mathcal{T}}} co(a, c, \mathcal{T})}{|C_c^{\mathcal{T}}|}$.

Plusieurs autres indicateurs peuvent être calculés au niveau du compte Twitter. Puisqu'un compte peut être impliqué dans plusieurs communautés politiques sur une série de périodes (ce qui pourrait refléter le fait que le propriétaire du compte est indécis), nous avons généralisé ces indicateurs au niveau communautaire en les pondérant par l'engagement des comptes (voir Tables 6 et 7) :

- **l'activité (activity)** d'un compte a sur \mathcal{T} est le nombre de tweets politiques émis pour cet ensemble de périodes (*i.e.* nombre de tweets capturés par le *Politoscope*). L'activité d'une communauté $C_c^{\mathcal{T}}$ est alors définie comme la moyenne pondéré des activités des comptes participant à $C_c^{\mathcal{T}}$.
- la **politisation (politicization)** d'un compte a sur un ensemble de périodes \mathcal{T} est le pourcentage de tweets politiques émis par un compte (*i.e.* nombre de tweets capturés par le *Politoscope* pendant \mathcal{T} sur le nombre total estimé de tweets pendant la même période). Puisque nous ne suivons pas la plupart des comptes qui interagissent avec le *Politoscope*, nous estimons pour chaque compte le nombre total de tweets émis pendant \mathcal{T} comme la différence entre le nombre total de tweets pour ce compte indiqué dans les méta-données du dernier tweet observé et la même quantité mentionnée en dans le premier tweet observé. Cette estimation peut être biaisée de deux façons. D'abord, nous ne capturerons peut-être pas tous les tweets politiques. Deuxièmement, les utilisateurs peuvent supprimer des tweets. Cela diminue le nombre de tweets mentionnés dans les méta-données mais pas le nombre de tweets vus dans le *Politoscope*, ce qui fait que la *politisation* est supérieur à 1 pour certains utilisateurs. Par conséquent, nous fixons le maximum de *politisation* d'un compte à 1. Comme nous pouvons raisonnablement faire l'hypothèse que ces deux biais s'appliquent uniformément à toutes les communautés, nous considérons cette estimation comme un bon indicateur pour comparer le degré de politisation des communautés politiques. La *politisation* de $C_c^{\mathcal{T}}$ est alors définie comme la moyenne pondérée des *politisation* des comptes participant à $C_c^{\mathcal{T}}$.
- **l'intégration politique (political integration)** d'un compte a au cours de \mathcal{T} est le pourcentage de tweets émis dans une communauté parmi le nombre total de tweets politiques de aa capturés. Un compte qui twitte principalement en interaction avec une communauté aura un indice élevé d'intégration politique. Au contraire, un compte dont le contenu politique n'est pas en relation avec un groupe structuré sur Twitter (*i.e.* ce compte est souvent classé comme appartenant à "la mer") aura un index d'*intégration politique* bas. L'*intégration*

Table 6. Statistiques détaillées pour les communautés des 6 candidats ayant l'utilisation la plus intensive de Twitter dans leur campagne ($T_k \in [\text{Feb. 1er 2017} - \text{22 avril 2017}]$) : temps moyen passé par compte dans la communauté ($\langle \text{jours} \rangle$), nombre total de tweets de cette communauté ($|\text{Tweets}|$), nombre de comptes présents au moins une fois dans la communauté ($|\text{comptes}|$) et nombre de comptes présents au moins une fois dans la communauté pondéré par l'engagement ($|\text{comptes ponderes}|$). Les valeurs extrêmes sont en gras pour une meilleure lisibilité.

Name	$\langle \text{jours} \rangle$	$ \text{Tweets} $	$ \text{comptes} $	$ \text{comptes ponderes} $
Jean-Luc Mélenchon	12.11	3,110,517	22,118	17,375.96
Benoît Hamon	11.69	1,828,418	18,950	14,340.52
Emmanuel Macron	13.92	3,643,822	24,803	18,717.05
François Fillon	18.49	5,032,109	18,264	14,428.71
Marine Le Pen	17.18	3,693,650	18,438	14,622.39
François Asslineau	8.95	709,901	6,597	3,713.01

Table 7. Statistiques détaillées pour les communautés des 6 candidats ayant l'utilisation la plus intensive de Twitter dans leur campagne ($T_k \in [\text{1er Fév. 2017} - \text{22 Avr. 2017}]$) : engagement moyen des comptes dans la communauté ($\langle \text{engagement} \rangle$) ; quantités pondérées par l'engagement : nombre moyen de tweets par compte ($\langle \text{activite} \rangle$), intégration politique ($\langle \text{integration} \rangle$) et politisation ($\langle \text{politisation} \rangle$) Les valeurs extrêmes sont en gras pour une meilleure lisibilité.

Name	$\langle \text{engagement} \rangle$	$\langle \text{activite} \rangle$	$\langle \text{integration} \rangle$	$\langle \text{politisation} \rangle$
Jean-Luc Mélenchon	93.88%	140.63	54.19%	2.62%
Benoît Hamon	91.96%	96.49	57.16%	6.16%
Emmanuel Macron	91.93%	146.91	54.51%	2.91%
François Fillon	93.24%	275.52	63.40%	5.42%
Marine Le Pen	95.74%	200.33	60.72%	3.63%
François Asslineau	85.43%	107.61	56.05%	2.80%

moyenne de C_c^T est alors définie comme la moyenne pondérée de l'intégration politique des comptes participant à C_c^T .

663
664

Ces indicateurs permettent de caractériser différents types d'engagement politique qui pourraient être spécifiques à certaines communautés. En retour, l'identification de certaines particularités au niveau des communautés pourrait indiquer des spécificités dans les modes de recrutement des nouveaux membres.

665
666
667
668

La communauté autour d'Asslineau est clairement un cas particulier dans cet ensemble de communautés politiques. Étant de loin la plus petite en nombre de comptes, le niveau d'engagement y est le plus faible et le niveau de politisation et d'intégration figurent parmi les plus faibles.

669
670
671
672

En excluant cette communauté atypique et en classant les cinq autres selon un axe gauche-droite, nous observons une courbe d'engagement en forme de U qui indique que les personnes ayant les opinions politiques les plus extrêmes ont tendance à être les plus inflexibles dans leur idéologie. Ceci est à rapprocher de l'observation de Preotçiuç-Pietro *et al.* [65] aux États-Unis selon laquelle les personnes dont les idées politiques sont les plus extrêmes ont le plus haut niveau d'engagement politique.

673
674
675
676
677
678

Les membres de la communauté Le Pen et Fillon sont les plus actifs sur Twitter avec une activité de 200, 33 (resp. 275, 52). Ce sont aussi ceux qui ont la plus longue

679
680

participation au sein d'une communauté (respectivement 17, 18 et 18, 49 jours) et ceux qui sont les plus intégrées politiquement (respectivement 63, 40 et 60, 72).

Alors que le nombre moyen de jours passés au sein d'une communauté est directement lié au niveau d'activité, une forte intégration signifie que, plus que pour les autres communautés, lorsque les membres des communautés de Fillon et Le Pen expriment leurs opinions politiques sur Twitter, c'est pour soutenir un groupe politique particulier. Ces deux communautés se distinguent cependant par l'engagement de leurs membres, les membres de la communauté de Le Pen ayant le score d'engagement le plus élevé de toutes les communautés (95, 74%). Le haut niveau d'engagement et d'intégration des membres de la communauté Le Pen font qu'ils sont les plus inflexibles de tous les militants politiques dans leurs opinions et les plus engagés envers un leader. Étant donné que la communauté Le Pen est aussi l'une des plus actives sur Twitter, c'est celle qui possède le noyau communautaire le plus stable. C'est probablement la raison pour laquelle les leaders de la droite ont rarement réussi à attirer les électeurs d'extrême droite malgré leur volonté affichée de le faire. Ceci explique également pourquoi, comme relevé dans [42], au lieu de combattre Le Pen qui était en bonne position pour passer le premier tour des élections, la plupart des dirigeants ont préféré se battre entre-eux pour avoir l'opportunité de lui faire face lors du second tour des élections.

Les communautés de Macron et de Mélenchon sont de loin celles qui comptent le plus grand nombre de comptes (24.000 et 22.100 respectivement), mais aussi celles qui ont les plus faibles intégration et de politisation ainsi que des moyennes assez faibles en contribution journalière aux communautés. Cela reflète le caractère très dynamique de ces deux candidatures dont les résultats finaux étaient imprévisibles en début de campagne et qui ont pris de l'ampleur vers la fin de la campagne. Cela laisse également penser ces deux leaders ont réussi à recruter un public relativement peu politisé jusque là.

La communauté Hamon est celle dont la politisation est la plus forte (6, 16) avec des niveaux modérés d'engagement et d'intégration, ce qui tend à indiquer qu'il s'agissait d'une communauté avec une forte culture politique mais qui était plus axée sur les idées que sur le soutien à un leader particulier. Ceci fait écho avec l'analyse de [42] qui a démontré à partir des données du *Politoscope* que la communauté Hamon était celle qui exprimait ses idées politiques de la manière la plus détachée de toute référence à un candidat ou un parti. La faible contribution journalière moyenne de la communauté reflète le phénomène du *vote utile* : à un moment donné (vers mars), de nombreuses personnes de la communauté Hamon ont rejoint les communautés Mélenchon ou Macron lorsque la croyance collective s'est fixée sur l'idée que Hamon ne serait pas qualifié pour le second tour. Et en effet, il a fini loin derrière tous les autres candidats.

La décomposition de l'engagement des membres des communautés politiques en ligne suivant les facteurs *activité*, *politisation*, *intégration politique* et *engagement* permet donc de distinguer différents types de groupes politiques. Nous conjecturons que l'analyse de leur évolution dans le temps pourrait contribuer à prédire la destinée des communautés politiques.

Propagation de l'information politique

Les systèmes sociaux comptent parmi les systèmes complexes les plus sophistiqués en termes d'enchevêtrement de niveaux et de dynamiques multi-échelles. Il faut donc s'attendre à ce que leurs structures multi-niveaux soit un facteur clé dans les phénomènes de diffusion, les réseaux sociaux étant un lieu privilégié pour observer leurs rôles.

Dans cette section, nous analysons la manière dont les communautés influencent la diffusion et la circulation de l'information politique, en défendant certains thèmes ou en

mettant en avant certains sujets avec un cadrage spécifique. Une attention particulière est accordée au phénomène des fausses nouvelles (“fake news”) qui peut être étudié ici dans toute sa dimension structurelle et communautaire.

Profils sémantiques et effet de “chambre d’écho”

Étant donné notre méthode de reconstruction, un premier résultat intuitif est que les communautés politiques génèrent des effets important de chambre d’écho (*echo chamber*) [66] : l’analyse de la propagation des retweets (voir Texte F dans S1 File) montre que la grande majorité des tweets sont retweetés dans leur communauté d’origine.

Pour aller au-delà de cette observation, nous avons sélectionné 11 des principaux sujets de débat de cette élection (voir Texte G dans S1 File) et analysé l’évolution des profils sémantiques des tweets ayant circulé dans chaque communauté en relation avec ces sujets. On peut observer une grande hétérogénéité inter-communautaire, tant en termes de sujets, de manière de contextualiser l’information et d’agenda.

Plus précisément, nous avons analysé la spécificité du discours politique, en fonction du thème, à plusieurs niveaux :

- Pour chaque thème, diversité des partis politiques traitant ce thème,
- Diversité des thèmes abordés par un parti politique donné,
- Diversité des expressions utilisées par chaque communauté politique.

Pour chaque thème, la valeur calculée de l’entropie, basée sur le nombre de communautés abordant ce thème, montre que tous les thèmes sont couverts par une moyenne d’environ six communautés, avec un degré moindre de diversité pour les thèmes de *sécurité intérieure* et *immigration et multiculturalisme*, qui sont fortement prônés par l’extrême droite (voir tableau D dans S1 File).

Pour chaque communauté politique, le calcul de la diversité des thèmes abordés révèle clairement un inter-communautés (voir Tableau 8) et un positionnement différencié des thèmes dans leur agenda politique en fonction des moments de la campagne (voir l’analyse de la dynamique des thèmes par communauté dans le texte G.2 de S1 File). Cette hétérogénéité se retrouve également au niveau du vocabulaire utilisé par les membres de chaque communauté politique (voir tableau ??). Toutefois, comme cette analyse ne recherche pas à identifier les thèmes de manière exhaustive, la présente étude ne peut être utilisée pour qualifier les candidats en fonction de la diversité des thèmes qu’ils abordent.

Le *Politoscope*, la plateforme associée à cet article (<https://presidentielle2017.politoscope.org/dashboard>), permet d’explorer, pour chacun des principaux candidats et chacun des onze thèmes, les changements dans l’utilisation par ces communautés des 15 mots clés qui leur sont les plus caractéristiques. On peut également observer, outre la spécificité sémantique des communautés par rapport à ces thèmes, une variation en fonction du temps pour les différents candidats, l’utilisation de certains mots-clés apparaissant, se renforçant ou disparaissant avec le temps (voir Texte G.3 dans S1 File pour des détails sur la méthode et des exemples).

Circulation des fausses nouvelles au sein des communautés

Les fausses nouvelles sont un concept mal défini qui recouvre la propagande, la désinformation et les fausses allégations dans les médias. Néanmoins, l’étude de l’influence de ces phénomènes est d’une importance capitale. Les fausses nouvelles font actuellement l’objet de discussions aux Etats-Unis d’Amérique, notamment dans le cadre de l’élection présidentielle de 2016 [67, 68] et font l’objet en 2018 d’une adaptation

Table 8. Entropie des thèmes abordés par chaque candidat et sa

3—communauté. Les deux thèmes dominants ainsi que les pourcentages qui leur sont associés sont également présentés. L'entropie maximale est de 3,46 lorsque les 11 thèmes sont discutés de manière uniforme. En ce qui concerne les thèmes retenus, Benoît Hamon est le candidat qui a abordé le plus grand nombre de thèmes, tandis que François Asselineau est le candidat qui s'est le plus concentré sur un sous-ensemble de thèmes (dont la politique étrangère et la démocratie). Les variations de l'attention accordée par les candidats à ces thèmes sont représentées sur les Figures I à N dans S1 File.

Name	Entropy	# themes	Dominant themes	%
François Asselineau	2.63	6.2	foreign policy, democracy	57.7
François Fillon	3.06	8.3	immigration and multiculturalism, democracy	33.5
Benoît Hamon	3.19	9.1	employment, foreign policy	39.3
Marine Le Pen	2.78	6.9	immigration and multiculturalism, foreign policy	49.0
Emmanuel Macron	3.03	8.2	foreign policy, democracy	44.2
Jean-Luc Mélenchon	3.10	8.6	democracy, foreign policy	37.3
All candidates	3.08	8.4	foreign policy, democracy	36.9

de la législation en France depuis les tentatives de manipulation de l'opinion publique lors des élections présidentielles françaises de 2017 [69].

La question de savoir comment des informations vraies et fausses se propagent sur un réseau social comme Twitter et quels contextes influencent cette propagation est une question importante pour la "science des fausses nouvelles" qui est sensée aider à résoudre la crise post-vérité [70, 71]. Pour évaluer l'importance de la reconstruction des communautés politiques dans la compréhension de ce phénomène, nous avons utilisé la base de données *Decodex*, une initiative du journal *Le Monde*, l'un des principaux journaux français (orienté centre-gauche). Celui-ci a recueilli, au cours de l'année 2017, les URLs des pages web et des messages Facebook qui contenaient des fausses nouvelles [72]. Les URLs se référant à la même fausse nouvelle (e.g. "Jean-Luc Mélenchon porte une Rolex" ou "plus de 30% de la campagne d'Emmanuel Macron a été financée par l'Arabie Saoudite") ont été regroupées en rumeur et associés à une réfutation (*debunk*), un article en ligne dans lequel les journalistes du *Le Monde* expliquent en quoi cette nouvelle est fausse. 179 histoires ont ainsi été identifiées au total et 179 réfutations ont été écrites (voir Files A6 et A7 sur DOI:10.7910/DVN/AOGUIA pour une version json).

La base de données *Decodex* est un bon point d'entrée pour identifier, au prisme de la reconstruction de notre paysage politique, les enjeux du phénomène des fausses nouvelles.

Effet de chambre d'écho différentiel pour les fake news.

Après avoir nettoyé et normalisé les URLs du *Decodex* ; et après avoir vérifié que les tweets mentionnant un lien vers une fausse nouvelle ou une réfutation ont bien pour but de les diffuser (voir Texte H dans S1 File), nous avons étudié leur diffusion. Nous n'avons observé que 4 888 partages de tweets contenant un faux liens dans notre ensemble de données de 60 millions de tweets (0,0081%). Cela indique soit que les fausses nouvelles ne sont pas très partagées en France par les personnes intéressées par la politique sur Twitter, soit que le *Decodex* n'a pas réussi à identifier une proportion significative de fausses nouvelles circulant sur Twitter. De même, les réfutations n'ont été partagées que 1 275 fois.

Le nombre moyen de partage d'une fausse nouvelle est de 27,3 alors que le nombre moyen de partage d'un démenti est de 7,12. Par conséquent, pour cet ensemble de

Nom	Entropie	Nb keyword	Keyword dominant	%
Jean-Luc Mélenchon	7.53	184.8	loi travail, peuple, travail, république, nuitdebout, sondage, élus, médias, Pénélope Fillon, chômeurs	24.3
François Asselineau	5.98	63.2	frexit, europe, sondage, médias, brexit, peuple, élus, démocratie, traités, l'euro	48.7
Marine Le Pen	7.82	226.2	peuple, islamistes, patriotisme, police, médias, sondage, chômeurs, migrants, élus, terrorisme	20.7
François Fillon	7.75	215.8	médias, sondage, république, redressement, élus, chômeurs, police, peuple, travail, sécurité	19.6
Emmanuel Macron	7.71	209.1	Pénélope Fillon, république, élus, travail, sondage, justice, chômeurs, europe, police, mise en examen	19.8
Benoît Hamon	7.77	218.6	république, travail, les jeunes, revenu universel, démocratie, europe, élus, écologie, peuple, enfants	21.1
Philippe Poutou	7.55	187.5	nuitdebout, loi travail, république, travail, Pénélope Fillon, élus, peuple, sondage, chômeurs, médias	26.1
Nathalie Arthaud	7.32	159.2	frexit, sondage, europe, élus, chômeurs, médias, travail, peuple, nuitdebout, république	29.8
All	8.23	301.23	peuple, république, sondage, élus, médias, chômeurs, travail, frexit, police, europe	17.5

données de fausses nouvelles/démenti, les fausses nouvelles ont été partagées presque quatre fois plus que leurs démentis. 810

Cette différence peut être de nature structurelle ou conjoncturelle. Elle pourrait aussi être dû au fait qu'il est possible que nous n'ayons pas capté une proportion importante de réfutations qui auraient été publiées par d'autres journaux que *Le Monde*. Elle est néanmoins comparable aux résultats de Vosoughi *et al.*. 2018 [73] sur les fausses nouvelles politiques américaines sur Twitter : "Les fausses nouvelles diffusent beaucoup plus loin, plus vite, plus profondément et plus largement que la vérité dans toutes les catégories d'information, et les effets [sont] plus prononcés pour les fausses nouvelles politiques. ("falsehood diffuse significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects [is] more pronounced for false political news.") 811-821

Néanmoins, notre reconstruction des communautés politiques nous invite à reconsidérer les résultats de Vosoughi *et al.* étant données les nouvelles perspectives qu'elle offre sur la diffusion de fausses nouvelles. En calculant les entropies des fausses nouvelles et des diffusions de réfutations au niveau communautaire, nous avons constaté que le phénomène décrit Vosoughi *et al.* s'inverse (voir tableau 9) : en moyenne, les fausses nouvelles se propagent dans beaucoup moins de communautés (3,78) que les réfutations associées (7,67), qui à leur tour se propagent dans beaucoup moins de communautés que la moyenne des messages politiques (13,45). 822-829

Les fausses nouvelles peuvent toucher en moyenne plus de personnes que les autres nouvelles politiques, mais elles se propagent plus localement. Cet effet d'immersion de la structure communautaire sur la microdynamique, qui implique un renversement de perspective pendant la transition du niveau micro (les fausses nouvelles atteignent plus de comptes que les autres nouvelles) au niveau macro (les fausses nouvelles atteignent moins de communautés que les autres nouvelles) est très important pour aborder cette 830-835

Table 9. Entropie et nombre moyen de communautés [entre crochets] atteintes en moyenne par les fausses nouvelles, les réfutations et les tweets avec ou sans mention d’URL sur la période. Nous présentons également les résultats correspondant à l’agrégation de toutes les communautés en une seule à l’exception des 6 principales communautés politiques en ligne identifiées lors de l’élection présidentielle française. Les tweets sans communauté (43,44%) ont été considérés comme une seule communauté pour le traitement de l’entropie. Pour comparaison avec les données sur les réfutations et les fausses nouvelles, identifiées par les URL qu’elles diffusent, nous fournissons également les statistiques globales pour le sous-ensemble de tweets qui mentionnent des URLs, quelle qu’elles soient.

entropie [# communautés]	Toutes les communautés	6 principales communautés
Fake news retweet (n=4,888)	1.92 [3.78]	1.53 [2.88]
Retweet de réfutations (n=1,275)	2.94 [7.67]	2.23 [4.69]
Tous retweets with url (n=7,318,878)	3.75 [13.45]	2.28 [4.85]
Tous retweets (n=43,101,891)	3.72 [13.17]	2.54 [5.82]

question du point de vue sociétal.

Sur Twitter, les fausses nouvelles proviennent principalement de certaines communautés politiques, pas de la mer. La proportion de retweets

n’est pas la seule variable pertinente. L’analyse des fausses nouvelles et des réfutations circulant au sein des cinq grandes communautés politiques (François Fillon, Benoît Hamon, Marine Le Pen, Emmanuel Macron et Jean-Luc Mélenchon) révèle des tendances claires en termes de localisation et de distribution temporelle.

18.9% (922) des fausses nouvelles correspondent à des tweets sans affiliation communautaire, *i.e.* de ce que nous avons appelé “la mer” (nous rappelons que les communautés sont calculées chaque jour et que la communauté d’un tweet est la communauté attribuée au compte qui émet le tweet au moment de l’émission). Puisque globalement 43,44% (25 695 888) des tweets n’ont pas de communauté, cela signifie que les comptes qui sont les moins politisés ou non impliqués dans un groupe de militants partagent en moyenne beaucoup moins de fausses nouvelles que les comptes les plus politiquement intégrés.

Il existe également une grande hétérogénéité dans la pratique du partage de fausses nouvelles au sein des communautés. Les communautés de Fillon (droite) et Le Pen (extrême droite) sont les plus actives et ont émis respectivement 50.75% et 22.21% du nombre total de fausses nouvelles partagées (89.9% de fausses nouvelles partagées au sein des communautés) ; alors que la communauté de Hamon est la moins active avec seulement 0.43% du total (voir tableau 10).

La dynamique temporelle de la diffusion des fake news présente également des caractères remarquables. Pour chaque rumeur, nous avons classé les tweets selon leur date d’émission avec une résolution temporelle d’une seconde. En supprimant les rumeurs qui ont été partagées moins de 30 fois, ceci afin d’atténuer les effets de petite taille, nous avons calculé le pourcentage de *premiers relais* par communauté définis comme les utilisateurs qui se classent dans le percentile 1/30 (cf. tableau 10). La communauté de Fillon est celle qui a le pourcentage le plus élevé de premiers relais (6% des comptes de cette communauté sont des premiers relais), devant la communauté de Le Pen (5%) ; alors qu’aucun des comptes des communautés de Hamon et de Macron ne comptent parmi les premiers relais.

En résumé, sur la base de la liste de fausses nouvelles du *Decodex*, la communauté de Fillon est à la fois celle qui a été la plus prompte à diffuser des fausses nouvelles et la principale source de partage de fausses nouvelles.

Table 10. Nombre d'utilisateurs, nombre de tweets mentionnant un lien référant à une fake news et proportion de premiers relais pour les tweets contenant des liens à des "fake news". Les premiers relais sont définis comme des utilisateurs qui ont sont dans le premiers 1/30 percentile d'utilisateurs à partager une fausse nouvelle.

Candidats	#Comptes	#Tweets	% premiers relais
François Fillon	1236	2481	0.06
Benoît Hamon	19	21	0
Marine Le Pen	784	1086	0.05
Emmanuel Macron	61	65	0
Jean-Luc Mélenchon	11	143	0.03
Autres communautés	123	170	0.03
Mer	614	922	0.03

Nous ne devons pas tirer de conclusions trop hâtives sur l'origine de ces fausses nouvelles. Plusieurs hypothèses sont compatibles avec ces résultats préliminaires. Au sein des communautés, le statut sociologique des comptes diffusant de fausses nouvelles est de la plus haute importance. Il peut s'agir de l'œuvre de leaders, de seconds couteaux, d'activistes, de robots ou de comptes tenu par des personnes externes à ces communautés qui tentent d'influencer une communauté particulière par une pratique d'astroturfing (ces comptes peuvent même être tenus par des personnes situées à l'étranger). L'estimation de ce statut social n'entre pas dans le cadre de cet article, mais on peut dire que ces comptes ont tissés des interactions suffisamment étroites avec les communautés politiques pour être associés à celles-ci via notre méthode de détection communautaire.

L'analyse des tweets contenant des réfutations offre un point de vue radicalement différent (cf. tableau 11). Tandis que 38.11% des réfutations circulent dans la mer, la communauté la plus active est celle de Macron avec 39.18% du total des réfutations à elle seule. De plus, 12% des utilisateurs qui diffusent des réfutations au sein de la communauté Macron font partie du premier 1/10 percentile des utilisateurs qui partagent une réfutation lorsqu'ils sont triés par date du tweet (premiers relais). La communauté de Mélenchon est également très réactive avec 14% de premiers relais, même s'ils sont beaucoup moins actifs.

Table 11. Nombre d'utilisateurs, nombre de tweets et proportion de premiers relais pour les tweets contenant une réfutation. Les premiers relais sont des utilisateurs qui ont été parmi les premiers 1/10 percentile d'utilisateurs à partager une réfutation.

Candidats	#Comptes	#Tweets	% premiers relais
François Fillon	15	15	0.07
Benoît Hamon	47	69	0.11
Marine Le Pen	7	8	0
Emmanuel Macron	361	499	0.12
Jean-Luc Mélenchon	83	98	0.14
Autres communautés	82	100	0.12
Mer	345	486	0.05

Pour comparer les propensions au partage de fausses nouvelles et de réfutations, la Fig 13 montre la propension normalisée de chaque communauté à partager des fausses nouvelles et des réfutation (une propension de 1 correspond à la moyenne). La communauté Macron a partagé presque quatre fois plus de réfutations que la moyenne,

tandis que la communauté Fillon a partagé plus de quatre fois plus de fausses nouvelles que la moyenne.

893
894

<debunk>= 2.156e-05 – <fake>= 8.264e-05

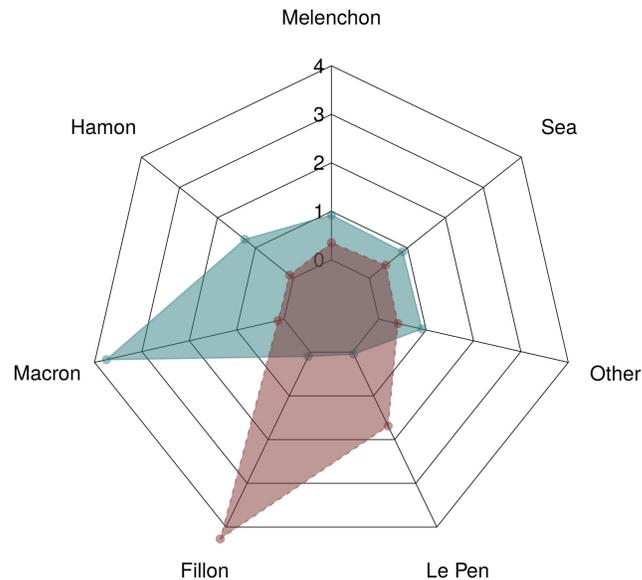


Fig 13. Diagramme radar de la propension normalisée à propager des fausses nouvelles (en pointillé rouge) et des réfutations (en bleu) par communauté.

Cette étude préliminaire dépeint un paysage politique schizophrénique “split brain” du point de vue des fausses nouvelles : un hémisphère produit ou diffuse des fausses nouvelles, l’autre produit ou diffuse des réfutations. Cette dissociation entre le public qui lit les fausses nouvelles et celui qui lit leurs réfutations doit être prise en compte dans les stratégies d’atténuation du phénomène des fake news.

895
896
897
898
899

Combattre les fausses nouvelles est une tâche très délicate avec au moins deux effets contre-intuitifs : le *backfire effect* [74] - la répétition d’un mythe augmente la familiarité et le renforce - et l’effet d’influence continue (*continued influence effect*) [75] - malgré une réfutation, les gens continuent de se fier à la désinformation. De plus, en l’absence d’exposition médiatique, les fausses nouvelles se propagent localement, de proche en proche. Mais en exposant un large public à la réfutation d’une fausse nouvelle, il n’y a aucune garantie qu’une proportion importante de ce public ne choisira pas de faire confiance à la source de la fausse nouvelle plutôt qu’à celle de la réfutation (*i.e.* les partisans des théories du complot). Ainsi, exposer certaines communautés qui n’étaient pas au courant de l’existence d’une fausse nouvelle pourrait déclencher un nouveau foyer de fausses nouvelles dans des secteurs des médias sociaux qui étaient jusque-là épargnés.

900
901
902
903
904
905
906
907
908
909
910

Contrairement à la vaccination pour laquelle les gens doivent être traités préventivement, dans le domaine des fausses nouvelles, les gens doivent être démystifiés après exposition. En ce sens, le fait que les fausses nouvelles aient tendance à se propager dans moins de communautés que les autres types d’information est une bonne nouvelle : les contre-mesures peuvent se restreindre à ces communautés particulières ou aux médias qu’elles consultent sans interférer avec un public plus large. La deuxième bonne nouvelle, quand on regarde de près les chiffres, c’est qu’au moins sur Twitter, les fausses nouvelles politiques ne semblent pas être très populaires. Notre étude a identifié qu’en France, seulement 0,08% des tweets politiques contenaient de fausses nouvelles.

911
912
913
914
915
916
917
918
919

Bien qu'il soit probable que de *Decodex* n'ait pas réussi à identifier une proportion significative de fausses nouvelles, cette proportion est néanmoins assez faible. Cette estimation est également comparable avec les observations de Vosoughi *et al.* (2018) qui ont identifié "seulement" 4,5 millions de partages de fausses nouvelles sur une période de 10 ans de Twitter, à comparer au nombre de tweets envoyés par jour en 2017, estimé 500 millions.

Conclusions

Nous avons démontré que les méthodes d'analyse des échanges de contenus politiques sur des sites de réseaux sociaux tels que Twitter permettent de reconstruire automatiquement des communautés politiques qui, par leur profil sémantique et leurs évolutions, permettent de mieux comprendre les dynamiques en jeu dans l'espace politique militant.

En appliquant ces méthodes aux données de l'élection présidentielle française de 2017, nous avons reconstruit avec une grande précision l'espace politique multipolaire français, récupéré les principaux événements survenus pendant la campagne et révélé les multiples reconfigurations des communautés politiques, dont certaines ont eu lieu dans des circonstances relativement confidentielles. En n'utilisant que des données d'interaction, notre approche a l'avantage d'être indépendante du langage utilisé lors de ces échanges, à condition que l'activité sociale étudiée implique un certain niveau d'engagement de la part des participants, ce qui est le cas en politique.

Après avoir vérifié l'exactitude de cette reconstruction, nous avons démontré comment cette approche peut aider à mieux caractériser les communautés politiques en ligne, tant du point de vue de leur structure que du point de vue du type d'informations qu'elles propagent, y compris les fausses nouvelles :

1. Les communautés politiques peuvent être reconstruites à partir des données de Twitter :
 - L'utilisation de Twitter par les communautés politiques permet de reconstituer l'évolution du paysage politique d'un pays pendant une campagne présidentielle à travers la correspondance entre les différents courants politiques et les communautés reconstruites algorithmiquement à partir de leurs traces numériques,
 - Au événements politiques clés (officiels et non officiels) correspondent des structures mathématiques identifiables dans le paysage politique reconstruit, comme la division, la fusion, la disparition, l'apparition ou le changement abrupt de la taille des communautés,
2. Cette reconstruction donne un aperçu des structures des communautés politiques et de leur dynamique (cas de la France) :
 - Les communautés en ligne peuvent se reconfigurer en très peu de temps à la suite d'événements clés tels que des victoires aux primaires ou des alliances entre dirigeants,
 - Il existe une structure communautaire politique générique sur Twitter, composée d'un noyau de militants qui restent plusieurs semaines au sein de la communauté, d'un ensemble de membres périphériques beaucoup plus volatile et d'un ensemble intermédiaire de comptes couvrant un large spectre de participation,

- Les communautés politiques sont hétérogènes en ce qui concerne l'activité en ligne, l'engagement, l'intégration politique et la politisation de leurs membres, ce qui suggère à la fois des profils différents d'engagement politique d'une communauté à l'autre et une dynamique distincte pour recruter de nouveaux membres. La caractérisation, à partir de traces numériques, de la variété des processus qui soutiennent la cohésion d'une communauté reste un défi ouvert, 965-970
 - Les personnes ayant des opinions politiques extrêmes ont tendance à être les plus attachées à leur idéologie, 971-972
3. Cette reconstruction donne un aperçu de l'impact des communautés sur la diffusion de l'information et en particulier sur leur rôle dans le phénomène des fausses nouvelles (le cas de la France) : 973-975
- La structure de la communauté en ligne limite la diffusion de l'information parmi les activistes de Twitters, la plupart des cascades de retweets se produisant au sein d'une ou deux communautés, 976-978
 - Le paysage politique français de Twitter ressemble à un "split brain" du point de vue des fausses nouvelles : un "hémisphère" produit ou diffuse des fausses nouvelles, l'autre produit ou diffuse des réfutations, 979-981
 - Les fausses nouvelles semblent être plus populaires que leurs réfutations en termes de nombre de comptes atteints, mais le phénomène s'inverse au niveau communautaire : la diffusion des fausses nouvelles tend à se limiter à moins de communautés que les autres nouvelles politiques. Par conséquent, dans l'évaluation de l'impact des fausses nouvelles, il faut tenir compte non seulement du nombre de personnes touchées, mais aussi de la diversité des communautés concernées. 982-988
 - Dans l'ensemble, sur Twitter, pour la politique française, le phénomène des fausses nouvelles semble assez limité tant en termes de volume (ordre de grandeur de 0,1%) qu'en termes de diversité des communautés concernées. 72,9% de toutes les fausses nouvelles analysées ont été produites et diffusées par seulement deux communautés politiques. Leur diffusion est donc avant tout l'œuvre des communautés politiques et profanes en politiques. 989-994

En raison de sa durée de 11 mois, de sa résolution temporelle (saisie des données en temps réel avec une précision à la seconde) et de son ampleur (2,4 millions de comptes Twitter, dont 180.000 ont eu suffisamment d'interactions avec le *Politoscope* pour être associés à une communauté politique), cette étude établit les bases d'une série d'analyses permettant de mieux comprendre la dynamique des opinions et la reconfiguration des communautés politiques lors d'une élection présidentielle. Elle permet d'accéder à un niveau intermédiaire entre les enquêtes sociologiques de terrain et les études statistiques globales (telles que celles menées par les agences nationales de recensement), dont on peut espérer qu'à l'avenir, il représentera une source majeure de connaissances politiques et sociologiques complémentaires aux approches plus traditionnelles. 995-1004

Nous concluons cet article par une remarque sur les implications politiques et éthiques du développement de ces méthodes. Depuis quelques années, la frontière entre vie privée et vie publique n'a cessé de s'estomper à mesure que les traces numériques laissées par les citoyens ont été analysées de manière systématique, en croisant un nombre croissant de bases de données (achats, historiques de navigation web, votes, cartes de fidélité, transports, GPS, etc.), en considérant un historique de plus en plus conséquent portant sur des tailles d'échantillon de la population qui tendent à être de l'ordre de grandeur de la population elle-même. Le marketing et la sphère économique 1005-1012

ont été les premiers domaines d'activité à s'emparer de ces nouvelles opportunités à des 1013
fin de valorisation. Plus récemment, la sphère politique a pris la mesure de ces nouvelles 1014
opportunités et a cherché à les mettre à profit afin de mieux connaître les préférences et 1015
orientations politiques des électeurs. En témoigne la fuite survenue en juillet 2017 de 1016
données personnelles sensibles de près de 200 millions d'américains compilées pour le 1017
compte du parti Républicain ($\approx 99\%$ du corps électoral) [76]. Nous avons pris soin sur 1018
l'ensemble du projet *Politoscope* d'anonymiser les données traitées et de ne livrer que 1019
des données agrégées, rendant impossible l'identification d'individus particuliers mis à 1020
part quelques personnalités politiques connues. Les travaux présentés dans cet article 1021
démontrent cependant qu'il est possible, à partir de traces laissées ostensiblement dans 1022
l'espace publique et avec des moyens computationnels relativement accessibles, de 1023
catégoriser une proportion significative de l'électorat d'un pays en fonction de ses 1024
orientations politiques, même si, comme nous l'avons souligné, une certaine marge 1025
d'erreur persiste. Dès lors, il est important de prendre conscience que toute trace laissée 1026
sur le web engage son auteur-e de ce point de vue. La plateforme ouverte *politoscope.org* 1027
a notamment pour but d'accompagner cette prise de conscience. Nous ne pouvons 1028
compter que sur l'honnêteté et la bonne foi des dirigeants politiques actuels et futurs 1029
pour ne pas utiliser, comme cela a pu se produire maintes fois par le passé et encore 1030
actuellement dans certains pays [77–80], des données de ce type à des fins de 1031
manipulation, de discrimination, voire de répression ou de persécution. 1032

Remerciements 1033

Cette recherche a été soutenue par l'Institut des Systèmes Complexes de Paris 1034
Ile-de-France (<https://iscpif.fr>), le projet *Algodiv* (ANR-15-CE38-0001), le projet 1035
Science en Poche (programme *Emergence(s)*, Ville de Paris) et le DIM *Problématiques* 1036
transversales aux systèmes complexes (Domaine d'intérêt majeur) de la Région 1037
Île-de-France. 1038

Nous remercions chaleureusement nos évaluateurs ainsi que tous nos collègues pour 1039
leurs commentaires et suggestions fructueux sur notre travail. 1040

References

1. Green K. The Game Changer: Social Media and the 2016 Presidential Election. The Huffington Post. 2015;.
2. Brunel A. Les technologies BigData au service de la campagne électorale; 2017. Available from: <https://www.franceinter.fr/politique/les-technologies-bigdata-au-service-de-la-campagne-electorale>.
3. Barberá P, Rivero G. Understanding the Political Representativeness of Twitter Users. *Social Science Computer Review*. 2015;33(6):712–729. doi:10.1177/0894439314558836.
4. Rao D, Yarowsky D. Detecting Latent User Properties in Social Media. Proc of the NIPS MLSN Workshop. 2010; p. 1–7.
5. Pennacchiotti M, Popescu AM. A Machine Learning Approach to Twitter User Classification. *Icwsn*. 2011;11(1):281–288.
6. Conover MD, Gonçalves B, Ratkiewicz J, Flammini A, Menczer F. Predicting the political alignment of twitter users. In: Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011; 2011. p. 192–199.

7. Cohen R, Ruths D. Classifying Political Orientation on Twitter: It's Not Easy! In: International AAAI Conference on Web and Social Media; 2013. p. 91–99. Available from: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6128>.
8. Adamic LA, Glance N. The Political Blogosphere and the 2004 U.S. Election. Proceedings of the 3rd international workshop on Link discovery - LinkKDD '05. 2005; p. 36–43. doi:10.1145/1134271.1134277.
9. Darmon D, Omodei E, Garland J. Followers are not enough: A multifaceted approach to community detection in online social networks. PLoS ONE. 2015;10(8). doi:10.1371/journal.pone.0134860.
10. Garimella K, De Francisci Morales G, Gionis A, Mathioudakis M. Quantifying Controversy in Social Media. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining - WSDM '16. New York, New York, USA: ACM Press; 2016. p. 33–42. Available from: <http://dl.acm.org/citation.cfm?doid=2835776.2835792>.
11. Boutet A, Hyoungshick Kim, Yoneki E. What's in Twitter: I Know What Parties are Popular and Who You are Supporting Now! In: 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE; 2012. p. 132–139. Available from: <http://ieeexplore.ieee.org/document/6425772/>.
12. Larsson AO, Moe H. Studying political microblogging: Twitter users in the 2010 Swedish election campaign. New Media & Society. 2012;14(5):729–747. doi:10.1177/1461444811422894.
13. Morales AJ, Borondo J, Losada JC, Benito RM. Measuring political polarization: Twitter shows the two sides of Venezuela. Chaos. 2015;25(3). doi:10.1063/1.4913758.
14. Amor BRC, Vuik SI, Callahan R, Darzi A, Yaliraki SN, Barahona M. Community detection and role identification in directed networks: understanding the Twitter network of the care. data debate. Dynamic Networks and Cyber-Security. 2016;1:111.
15. Guerrero-Sole F. Community Detection in Political Discussions on Twitter: An Application of the Retweet Overlap Network Method to the Catalan Process Toward Independence. Social Science Computer Review. 2015;35(2):244 – 261. doi:10.1177/0894439315617254.
16. Cherepnalkoski D, Mozetic I. A retweet network analysis of the European Parliament. In: 2015 11TH International Conference On Signal-Image Technology & Internet-Based Systems (SITIS); 2015. p. 350–357.
17. Cherepnalkoski D, Mozetič I. Retweet networks of the European Parliament: evaluation of the community structure. Applied Network Science. 2016;1(1):2. doi:10.1007/s41109-016-0001-4.
18. Barberá P. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. Political Analysis. 2015;23(1):76–91. doi:10.1093/pan/mpu011.
19. Briatte F, Gallic E. Recovering the French Party Space from Twitter Data. In: Science Po Quanti. Paris, France; 2015. Available from: <https://halshs.archives-ouvertes.fr/halshs-01511384>.

20. Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R. Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological science*. 2015;26(10):1531–42. doi:10.1177/0956797615594620.
21. Huberty M. Can we vote with our tweet? On the perennial difficulty of election forecasting with social media. *International Journal of Forecasting*. 2015;31(3):992–1007. doi:https://doi.org/10.1016/j.ijforecast.2014.08.005.
22. Mediametrie. L'Audience Internet Mobile en France en septembre 2017; 2017. Available from: www.mediametrie.fr/internet/communiqués/telecharger.php?f=db29450c3f5e97f97846693611f98c15.
23. Statista. Twitter MAU worldwide 2017 | Statistic; 2018. Available from: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>.
24. Chavalarias D, Panahi M, Gaumont N. Politoscope data : tweets Ids for the 2017 French Presidential Elections; 2018. Available from: doi:10.7910/DVN/6739SP.
25. Gaumont N, Panahi M, Chavalarias D. Data for : Reconstruction of the socio-semantic dynamics of political activist Twitter networks; 2018. Available from: 10.7910/DVN/AOGUIA.
26. Tarde G. *Les lois de l'imitation*. Les classiques des Sciences Sociales ed. Jean-Marie Tremblay; 1890.
27. Martin A, Guéguen N, Fischer-Lokou J. L'imitation humaine: Une synthèse de 50 années de recherche en psychologie sociale. *Canadian Psychology/Psychologie canadienne*. 2016;57(2):101.
28. Tomasello M. *The cultural origins of human cognition*. Harvard university press; 2009.
29. Hopcroft J, Khan O, Kulis B, Selman B. Tracking evolving communities in large linked networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101 Suppl:5249–5253. doi:10.1073/pnas.0307750100.
30. Falkowski T, Spiliopoulou M, Bartelheimer J. Community dynamics mining. In: *Proceedings of 14th European Conference on Information Systems (ECIS 2006)*. Citeseer; 2006. p. 318–329. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.5334{&}rep=rep1{&}type=pdf>.
31. Palla G, Barabási AL, Vicsek T. Quantifying social group evolution. *Nature*. 2007;446(7136):664–667. doi:10.1038/nature05670.
32. Xie J, Kelley S, Szymanski BK. Overlapping community detection in networks. *ACM Computing Surveys*. 2013;45(4):1–35. doi:10.1145/2501654.2501657.
33. Kanawati R. Seed-centric approaches for community detection in complex networks. In: Meiselwitz G, editor. *6th international conference on Social Computing and Social Media*. vol. 8531. Springer International Publishing; 2014. p. 197–208. Available from: http://link.springer.com/chapter/10.1007/978-3-319-07632-4_{19}{%}5Cnhttp://link.springer.com/content/pdf/10.1007{%-}2F978-3-319-07632-4_{19}.pdf.
34. Fortunato S, Hric D. Community detection in networks: A user guide. *Physics Reports*. 2016;doi:10.1016/j.physrep.2016.09.002.

35. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008;2008(10):P10008. doi:10.1088/1742-5468/2008/10/P10008.
36. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: *International AAAI Conference on Weblogs and Social Media*; 2009.
37. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLOS ONE*. 2014;9(6):e98679. doi:10.1371/journal.pone.0098679.
38. Boyd D, Golder S, Lotan G. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: *Proceedings of the Annual Hawaii International Conference on System Sciences*; 2010. p. 1–10.
39. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? In: *Proceedings of the 19th international conference on World wide web - WWW '10*; 2010. p. 591. Available from: <http://portal.acm.org/citation.cfm?doid=1772690.1772751https://www.cs.bgu.ac.il/~snean151/wiki.files/22-WhatIsTwitterASocialNetworkOrANewsMedia.pdf>.
40. Cha M, Haddai H, Benevenuto F, Gummadi KP. Measuring User Influence in Twitter : The Million Follower Fallacy. *International AAAI Conference on Weblogs and Social Media*. 2010; p. 10–17. doi:10.1.1.167.192.
41. Bakshy E, Hofman JM, Mason WA, Watts DJ. Everyone’s an influencer: quantifying influence on twitter. In: *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM; 2011. p. 65–74.
42. Chavalarias D, Gaumont N, Panahi M. Sur Twitter, lepénistes prosélytes et hamonistes idéalistes. *Libérationfr*. 2017;.
43. Callon M, Courtial J, Turner W, Bauin S. From Translations to Problematic Networks - an Introduction to Co-Word Analysis. *Social Science Information Sur Les Sciences Sociales*. 1983;22(2):191–235. doi:10.1177/053901883022002003.
44. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of machine Learning research*. 2003;3(Jan):993–1022.
45. Hong L, Davison BD. Empirical study of topic modeling in Twitter. In: *Proceedings of the first workshop on social media analytics*. ACM Press; 2010. p. 80–88. Available from: <http://portal.acm.org/citation.cfm?doid=1964858.1964870>.
46. Leydesdorff L, Nerghes A. Co-word Maps and Topic Modeling: A Comparison from a User’s Perspective. *arXiv:151103020 [cs]*. 2015;.
47. Sparck Jones K. A Statistical Interpretation Of Term Specificity And Its Application In Retrieval. *Journal of Documentation*. 1972;28(1):11–21. doi:10.1108/eb026526.
48. Cointet JP, Chavalarias D. Multi-level Science mapping with asymmetric co-occurrence analysis: Methodology and case study. *Networks and Heterogeneous Media*. 2008; p. 267–276.

49. Dias G, Mukelov R, Cleuziou G. Mapping General-Specific Noun Relationships to WordNet Hypernym/Hyponym Relations. In: Knowledge Engineering: Practice and Patterns. Springer; 2008. p. 198–212.
50. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the web. Stanford InfoLab; 1999.
51. Challenges. L’UDI reste rallié à LR mais pose ses conditions à Fillon; 2017. Available from: https://www.challenges.fr/election-presidentielle-2017/presidentielle-2017-l-udi-reste-rallie-a-lr-mais-pose-ses-conditions-a-fillon_458990.
52. Forcari C. L’UDI est divisée entre un sommet tourné vers Les Républicains et une base favorable à Macron; 2017. Available from: http://www.liberation.fr/politiques/2017/03/25/l-udi-est-divisee-entre-un-sommet-tourne-vers-les-republicains-et-une-base-favorable-a-ma_1558188.
53. Hussonnois-Alaya C. La vie après la primaire: Sarkozy recase ses proches, Juppé “abattu”; 2016-12-12. Available from: <http://www.bfmtv.com/politique/nicolas-sarkozy-et-alain-juppe-la-vie-apres-l-echec-a-la-primaire-1069609.html>.
54. Lepelletier P. Jean-Pierre Mignard, intime de Hollande, soutient désormais Macron; 2016-12-02. Available from: <http://www.lefigaro.fr/politique/le-scan/citations/2016/12/02/25002-20161202ARTFIG00106-jean-pierre-mignard-intime-de-hollande-soutient-desormais-macron.php>.
55. Point L, AFP. Législatives : NKM sera candidate dans la circonscription de Fillon; 2017-01-17. Available from: http://www.lepoint.fr/politique/legislatives-nkm-sera-candidate-dans-la-circonscription-de-fillon-17-01-2017-2097902_20.php.
56. Enchaîné LC. Les 600 000 Euros gagnés par Penelope qui empoisonnent Fillon; 2017-01-25. Available from: <https://www.lecanardenchaine.fr/la-une-du-25-janvier-2017/>.
57. Bonnefous B, Pietralunga C, de Royer S. Primaire à gauche : après la victoire, le plus dur commence pour Benoît Hamon; 2017-01-30. Available from: http://www.lemonde.fr/election-presidentielle-2017/article/2017/01/30/primaire-de-la-gauche-apres-la-victoire-le-plus-dur-commence-pour-benoit-hamon_5071263_4854003.html.
58. Monde L, AFP. Présidentielle : Bayrou et Macron, une alliance sous conditions; 2017-02-22. Available from: http://www.lemonde.fr/election-presidentielle-2017/article/2017/02/22/presidentielle-fin-du-suspense-autour-de-la-candidature-de-francois-bayrou_5083592_4854003.html.
59. Siraud M. Jadot annonce qu’il retire sa candidature au profit de Hamon; 2017-02-24. Available from: <http://www.lefigaro.fr/elections/presidentielles/2017/02/23/35003-20170223ARTFIG00367-hamon-et-jadot-s-entendent-sur-un-accord.php>.

60. décodeurs L. Manifestation au Trocadéro : les chiffres exagérés du camp Fillon; 2017-03-06. Available from: http://www.lemonde.fr/les-decodeurs/article/2017/03/06/trocadero-chiffres-exageres-et-bataille-d-images_5089869_4355770.html.
61. Lemarié A, Goar M. Juppé renonce, la droite au bord de la rupture; 2017-03-06. Available from: http://www.lemonde.fr/election-presidentielle-2017/article/2017/03/06/juppe-renonce-la-droite-au-bord-de-la-rupture_5089836_4854003.html.
62. Poussart A. Nicolas Dupont-Aignan se voit comme une alternative à François Fillon; 2017-03-27. Available from: <https://www.publicsenat.fr/article/politique/nicolas-dupont-aignan-se-voit-comme-une-alternative-a-francois-fillon-57945>.
63. Besse Desmoulières R. Présidentielle : Mélenchon ne donne pas de consigne de vote; 2017-04-23. Available from: http://www.lemonde.fr/election-presidentielle-2017/article/2017/04/23/presidentielle-melenchon-ne-donne-pas-de-consigne-de-vote-dans-l-immediat_5116116_4854003.html.
64. Cavelier J. Le spectaculaire revirement de Dupont-Aignan sur le FN; 2017-04-29. Available from: http://www.lemonde.fr/election-presidentielle-2017/article/2017/04/29/le-spectaculaire-revirement-de-dupont-aignan-sur-le-fn_5120038_4854003.html.
65. Preot,ciuc Pietro D, Liu L, Hopkins D, Ungar L. Beyond binary labels: political ideology prediction of twitter users. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. vol. 1; 2017. p. 729–740.
66. Colleoni E, Rozza A, Arvidsson A. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*. 2014;64(2):317–332.
67. Clifton D. Fake news on Twitter—including from Russia—flooded swing states that helped Trump win; 2017-10-01. Available from: <http://www.motherjones.com/politics/2017/09/fake-news-including-from-russian-sources-saturated-battleground-states-trump-barely-won/>.
68. Howard P, Kollanyi B, Bradshaw S, Neudert LM. Social Media, News and Political Information during the US Election: Was Polarizing Content Concentrated in Swing States? | The Computational Propaganda Project; 2017. Available from: <http://comprop.oii.ox.ac.uk/2017/09/28/social-media-news-and-political-information-during-the-us-election-was-polarizing-content->
69. Chrisafis A. Emmanuel Macron promises ban on fake news during elections; 2018. Available from: <http://www.theguardian.com/world/2018/jan/03/emmanuel-macron-ban-fake-news-french-president>.
70. Lewandowsky S, Ecker UKH, Cook J. Beyond Misinformation: Understanding and Coping with the “Post-Truth” Era. *Journal of Applied Research in Memory and Cognition*. 2017;6(4):353–369. doi:10.1016/j.jarmac.2017.07.008.

71. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. *Science*. 2018;359(6380):1094–1096. doi:10.1126/science.aao2998.
72. Décodeurs L. Fausses informations : les données du Décodex en 2017. *Le Monde.fr*. 2017;.
73. Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science*. 2018;359(6380):1146–1151. doi:10.1126/science.aap9559.
74. Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J. Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*. 2012;13(3):106–131. doi:10.1177/1529100612451018.
75. Johnson H, Seifert C. Sources of the Continued Influence Effect: When Misinformation in Memory Affects Later Inferences. *Journal of Experimental Psychology*. 1994;20(1):1420–1436.
76. O’Sullivan D. The RNC Files: Inside the Largest US Voter Data Leak; 2018. Available from: <https://www.upguard.com/breaches/the-rnc-files>.
77. Greenwald G. No place to hide: Edward Snowden, the NSA, and the US surveillance state. Picador ed.; 2015. Available from: <http://us.macmillan.com/noplacetohide/glenngreenwald>.
78. Stoycheff E. Under Surveillance Examining Facebook’s Spiral of Silence Effects in the Wake of NSA Internet Monitoring. *Journalism & Mass Communication Quarterly*. 2016; p. 1077699016630255. doi:10.1177/1077699016630255.
79. Alonso P. L’allié français du paranoïaque Kadhafi; 2016. Available from: http://www.liberation.fr/france/2016/03/15/l-allie-francais-du-paranoiaque-kadhafi_1439846.
80. Botsman R. Big data meets Big Brother as China moves to rate its citizens;. Available from: <http://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion>.

Supporting information

S1 File : Exhaustive data description and complementary analyses. Table of content:

- Text A: Main political events of the 2017 French presidential campaign,
- Text B: Data collection platform
 - B.1 Collection focused on main political accounts (follow API),
 - B.2 Collection focused on political terms (track API) ,
 - B.3 List of the followed political figures (track and follow API),
 - B.4 Platform architecture,
 - B.5 Availability of the data,
- Text C: Bots detection,
- Text D: Lifetime of tweet and Twitter activity patterns,

- Text E: Community stability,
- Text F: Correlation between community structure and retweets patterns: confirmation of the echo chamber effect,
- Text G: Analysis of the communities' semantic background
 - G.1 Definition of the themes through queries,
 - G.2 Distribution of topics by political communities,
 - G.3 Evolution of community vocabulary,
 - Text H: Fake news data cleaning,
 - Text I: Vocabulary of the political discourse of the presidential election 2017,
 - Text J: Analysis of late retweets.

S1 Figure : High resolution of figure 7 : Reconfigurations of political communities between June 2016 and April 2017.

Supporting Information :

RECONSTRUCTION DES DYNAMIQUES SOCIO-SÉMANTIQUES DES RÉSEAUX DE MILITANTS POLITIQUES SUR TWITTER - Méthode et application à l'élection présidentielle française de 2017

Noé Gaumont¹, Mazyar Panahi², David Chavalarias^{1,2,*}

1 Centre d'Analyse et de Mathématiques Sociales (CAMS), Centre National de la Recherche Scientifique (CNRS) / École des Hautes Études (EHESS), Paris, France

2 Institut des Systèmes Complexes Paris Île-de-France (ISC-PIF), CNRS, Paris, France

* david.chavalarias@ehess.fr

Résumé

Contexte Les espaces numériques, et en particulier les sites de réseaux sociaux, sont de plus en plus présents et influents dans le fonctionnement de nos démocraties. Dans cet article, nous proposons une méthodologie intégrée pour la collecte de données, la reconstruction, l'analyse et la visualisation du développement du paysage politique d'un pays à partir des données de Twitter.

Méthode La méthode proposée repose uniquement sur les interactions entre les comptes Twitter et est indépendante des caractéristiques des contenus partagés tels que la langue des tweets. Nous validons notre méthodologie sur une étude de cas portant sur l'élection présidentielle française de 2017 (60 millions d'échanges Twitter entre plus de 2,4 millions d'utilisateurs) via deux méthodes indépendantes : la comparaison entre notre catégorisation politique automatisée et une catégorisation humaine basée sur l'évaluation d'un échantillon de 5000 descriptions de profils ; et la correspondance entre les reconfigurations détectées dans le paysage politique reconstruit et les événements politiques clés rapportés dans les médias. Cette dernière validation démontre la capacité de notre approche à refléter avec précision les reconfigurations de la scène politique d'un pays à partir des données des réseaux sociaux.

Résultats Nous nous sommes appuyés sur cette reconstruction pour donner un aperçu des dynamiques d'opinion et de la reconfiguration des communautés politiques qui sont en jeu lors d'une élection présidentielle. Tout d'abord, nous proposons une description et une analyse quantitative de l'engagement politique des membres des communautés politiques. Ensuite, nous analysons l'impact des communautés politiques sur la diffusion de l'information et en particulier sur leur rôle dans le phénomène des fausses nouvelles. Nous mesurons un effet différentiel de chambre d'écho (*echo chamber*) sur les différents types de nouvelles politiques (fausses nouvelles, démentis, nouvelles standards) causées par la structure communautaire. Nous soulignons l'importance de prendre en compte les méso-structures des réseaux politiques pour comprendre les phénomènes de type "fausses nouvelles" (*fake news*).

Conclusions En donnant accès à un niveau intermédiaire, entre les enquêtes sociologiques de terrain et les grandes études statistiques (telles que celles menées par des organisations nationales ou internationales), nous démontrons que les données des réseaux sociaux permettent de qualifier et de quantifier l'activité des communautés politiques dans un environnement politique multipolaire, ainsi que leur évolution et reconfiguration temporelle, leur structure, leurs stratégies d'alliance et leurs particularités sémantiques au cours d'une campagne présidentielle à travers l'analyse de leurs traces numériques. Nous concluons ce document par un commentaire sur les implications politiques et éthiques de l'utilisation des données des réseaux sociaux en politique. Nous insistons sur l'importance de développer des macroscopes sociaux qui permettront aux citoyens de mieux comprendre la manière dont collectivement ils font société. Nous proposons comme exemple le *Politoscope*, un microscope qui restitue certains de nos résultats d'une manière interactive.

Mots-clés : opinion dynamics, political activism, elections, socio-semantic networks, networks dynamics, Twitter.

Contents

1 Principaux événements politiques de la campagne présidentielle française de 2017	48
2 Dispositif de collecte des données	50
2.1 Collecte centrée sur les comptes des leaders politiques (API follow)	50
2.2 Collecte centrée sur les termes (API track)	50
2.3 Liste des principales personnalités politiques suivies (API track et follow)	52
2.4 Architecture de la plateforme	55
2.5 Disponibilité des données	56
3 Détection des bots	58
4 Durée de vie des tweet et motif d'activité sur Twitter	60
5 Stabilité des communautés	60
6 Corrélation entre structure communautaire et motifs de retweets: une confirmation des chambres d'écho	62
7 Analyse du contexte sémantique des communautés	64
7.1 Définition des thèmes par requêtes	64
7.2 Distribution des thèmes par communautés politiques	66
7.3 Évolution sémantique des communautés	67
8 Nettoyage des données sur les fausses nouvelles	72
9 Vocabulaire du discours politique lors des élections présidentielles de 2017	73
10 Analyse des retweets tardifs	86
11 Listes des tables et des figures	93

Listes des fichiers additionnels disponible sur Harvard Dataverse

Le lien persistant de ces fichiers est:

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2FAOGUIA>

- **A1 File - Graph_August_December.gexf:** Graphe anonymisé des 5-communautés enter août et décembre 2016. Les id des comptes sont anonymisé mais le nom des comptes des principales figures politiques est affiché.
- **A2 File - 2017 French Presidential Elections Programs.csv:** Liste des mesures politiques extraite des programmes de campagne des candidats à l'élection présidentielle de 2017.
- **A3 File - French_Presidential_Election_2017_Topics.gexf:** Carte thématique (analyse de co-occurrences des termes et proximité calculé par la *confidence*) provenant de l'analyse des programmes des candidats à l'élection présidentielle française de 2017.
- **A4 File - Politoscope_RetweetGraph_2017-04-09_2017-04-23_min_3_anon.tab:** Graphe anonymisé des 3-communautés sur la période [9 avril 2017 – 23 avril 2017] (le jour du premier tour des élections).
- **A5 File - Politoscope_RetweetGraph_2017-04-24_2017-05-08_min_3_anon.tab :** Graphe anonymisé des 3-communautés sur la période [24 avril 2017 – 8 mai 2017] (le jour du second tour des élections).
- **A6 File - hoax_lines.json** Liste des fausses nouvelles crée par *Les Decodeurs* (<https://github.com/multivacplatform/presidentielle2017-info>). Tous les liens ont été vérifiés manuellement pour supprimer les liens jugés ouvertement parodiques.
- **A7 File - debunk_lines.json:** Liste des debunks crée par *Les Decodeurs* (<https://github.com/multivacplatform/presidentielle2017-info>).
- **A8 File - topic_queries.txt:** Requêtes utilisées pour définir les principaux thèmes de l'élections présidentielle de 2017 en France. (<https://github.com/multivacplatform/presidentielle2017-info>).
- **A9 File - keywords-query.txt:** Termes extrait des tweets et des programmes de campagne des candidats à l'élection présidentielle de 2017 en France. (<https://github.com/multivacplatform/presidentielle2017-info>).

La liste des id des tweets collectés par la plateforme Politoscope.org entre le 1er août 2016 et le 8 mai 2017 (**A10 File - tweet_id.tar.xz**) est disponible à l'adresse

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6739SP>

1 Principaux événements politiques de la campagne présidentielle française de 2017

Au printemps 2016, beaucoup anticipaient une alternance gauche-droite, Alain Juppé étant le grand favori de tous les sondages. Les élections présidentielles françaises de 2017 ont été exceptionnellement mouvementée avec de nombreux retournements de situation en des laps de temps très courts et l'effondrement des deux principaux candidats (Fillon et Hamon) et partis ayant dominé la scène politique depuis plus de quarante ans. Nous en présentons ici les principaux jalons. Deux événements internationaux majeurs se sont également déroulés de manière concomitante qui ont surpris tous les analystes et ouvert le champ des possibles pour les futures élections : le Brexit et l'élection de Donald Trump.

- **2017/06/12** : Interview de Juppé sur TF1 déplorant son surnom "Ali Juppé" (action de dénigrement lancée par l'extrême droite puis reprise par certains membres des communautés Sarkozy et Fillon pendant les primaires de la droite),
- **2017/06/23** : Le Royaume-Uni vote le Brexit,
- **2016/07/14**: Attentat de Nice,
- **2016/08/28**: François Fillon, candidat à la primaire de la droite, a attaqué son rival Nicolas Sarkozy sur le volet judiciaire en posant la question "Qui imagine un seul instant le général de Gaulle mis en examen?". Cette phrase sera ressortie à plusieurs reprises pendant le Pénelope Gate,
- **2016/08/30**: Macron quitte le gouvernement
- **2016/10/13**: 1er débat primaires de la droite, Sarkozy et Juppé sont les favoris des sondages,
- **2016/11/08**: Élection de Donald Trump ,
- **2016/11/03**: 2ème débat primaires de la droite,
- **2016/11/16**: Annonce de la candidature Macron,
- **2016/11/17**: 3ème débat primaires de la droite,
- **2016/11/20**: Premier tour des primaires de la droite,
- **2016/11/27**: Deuxième tour des primaires de la droite,
- **2016/12/01**: Hollande renonce à se représenter,
- **2017/01/12**: 1er débat primaires de la gauche,
- **2017/01/15**: 2ème débat primaires de la gauche,
- **2017/01/17**: Frictions chez les *Républicains* après des investitures aux législatives contestées, dont celle de Nathalie Kosciusko-Morizet (NKM), parachutée chez dans la circonscription de Fillon réputée imperdable,
- **2017/01/19**: 3ème débat primaires de la gauche. La présidente du Front national est créditée de 25% à 26% d'intentions de vote au premier tour, selon un sondage Ipsos publié par Le Monde. De son côté, Emmanuel Macron se rapproche d'un François Fillon en perte de vitesse,
- **2017/01/20**: Prise de fonction de Trump. Charge de Rachida Dati (ex. Garde de Sceaux sous Sarkozy) contre NKM,
- **2017/01/21**: Maurice Leroy (UDI et soutien de Sarkozy pendant la primaire) met en garde François Fillon,
- **2017/01/22**: 1er tours primaires de la gauche,
- **2017/01/25**: Le *Canard Enchaîné* dévoile les premières informations sur ce qui sera appelé le *Penelope Gate*,
- **2017/01/29**: 2ème tours primaires de la gauche,
- **2017/02/21**: Fillon annonce son plan santé, revirement par rapport à ses déclarations précédentes,
- **2017/02/22**: Annonce de la convocation de Marine Le Pen par les policiers de l'Office anti-corruption de la police judiciaire (OCLCJFF), elle ne s'y rendra pas,
- **2017/02/23**: Ralliement de Bayrou à Macron,
- **2017/02/26**: Ralliement de Jadot à Hamon,
- **2017/03/01**: Convocation de Fillon pour une mise en examen, plusieurs ténors des *Républicains* appellent à un plan B. Juppé revient sur le devant de la scène comme possible remplaçant de Fillon,

- **2017/03/02:** Publication du programme d'*En Marche !*,
- **2017/03/03:** Convocation de Marine Le Pen en vue d'une possible mise en examen (affaire des assistants parlementaires),
- **2017/03/05:** François Fillon organise un rassemblement au Trocadéro. Il annonce qu'il y avait 200.000 personnes (dont des témoins soutiens de Sarkozy) et présente cela comme un soutien au maintien de sa candidature,
- **2017/03/06:** Juppé annonce qu'il renonce à toute perspective d'être candidat à la présidentielle. Confirmation de la légitimité de la candidature de Fillon par le bureau politique des *Républicains*,
- **2017/03/20:** Publication de la liste officielle des candidats par le Conseil Constitutionnel,
- **2017/03/20:** Débat TF1 entre les cinq "principaux" candidats à la présidentielle (Mélenchon, Hamon, Macron, Fillon, Le Pen),
- **2017/03/29:** Valls annonce qu'il votera Macron,
- **2017/04/04:** Grand débat sur BFMTV à onze candidats. Poutou se fait remarquer avec sa charge contre Le Pen et Fillon à propos de leurs "affaires",
- **2017/04/16:** Une commission du Parlement européen pourrait convoquer Marine Le Pen avant le second tour de la présidentielle,
- **2017/04/20:** Attentat des Champs Élysées pendant l'émission politique sur la chaîne France 2 alors que les onze candidats étaient invités à s'exprimer. Le Pen radicalise son discours avec une polémique : "la colonisation a beaucoup apporté, notamment à l'Algérie",
- **2017/04/23:** Premier tour de la présidentielle et victoire de Macron et Le Pen. Mélenchon refuse de donner des consignes de vote à ses électeurs, contrairement à ce qu'il avait fait en 2002 où il avait appelé à faire barrage contre le Front National,
- **2017/04/28:** Dupont-Aignan annonce son ralliement à Le Pen,
- **2017/05/03:** Contre-performance de Le Pen lors du débat télévisé Macron-Le Pen de l'entre deux tours.

2 Dispositif de collecte des données

Notre objectif étant d'étudier l'organisation de l'espace politique à l'occasion de la présidentielle française de 2017, nous avons choisi de collecter les données Twitter via deux méthodes complémentaires.

2.1 Collecte centrée sur les comptes des leaders politiques (API follow)

Nous avons établi une liste de plus de 3700 comptes de personnalités politiques françaises (candidats aux primaires, députés, sénateurs, personnes ayant un mandat important au sein d'un parti). Nous avons suivi chacun de ces comptes grâce à l'API 'follow' de Twitter (voir encadré).

La documentation de l'API donne la description exacte des contenus collectés pour le suivi d'un compte Twitter d'un utilisateur particulier (*user*):

*From Twitter **follow** API documentation^a.* For each user specified, the stream will contain:

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

The stream will not contain:

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

^a<https://dev.twitter.com/streaming/overview/request-parameters>

Ainsi, si l'on note $T(X, A)$ le fait que le compte X tweete le contenu A , nous avons récolté, dans la limite des volumes autorisés par les API de Twitter, pour chaque compte X suivi, et tout compte Y les contenus $T(X, A)$, $T(Y, T(X, A))$ et, dans le cas où A contient une réponse à X ("@ X ...") tous les $T(Y, A)$ (cf. fig. A1).

2.2 Collecte centrée sur les termes (API track)

La méthode de collecte centrée sur comptes via l'API 'follow' ne permet pas de collecter les tweets faisant référence à des personnes en dehors des cas mentionnés par Twitter. Pour pallier à cette limitation et obtenir les tweets qui citent un candidat important dans le corps du texte, nous avons constitué une liste de termes liés aux candidats (screenname, nom de parti, etc.) et utilisé l'API 'track' qui permet de capturer tous les tweets mentionnant une expression donnée. Plusieurs hashtags ont été créés spécifiquement à l'occasion de la présidentielle pour catégoriser les tweets pertinents pour cet événement. Nous en avons recensé la plupart et les avons ajoutés à la liste

d'expressions suivies afin de collecter les tweets les mentionnant (ex. #legranddebat [1,360,502 mentions], #2017ledebat [761,400 mentions]).

*From Twitter **track** API documentation^a.*

A comma-separated list of phrases which will be used to determine what Tweets will be delivered on the stream. A phrase may be one or more terms separated by spaces, and a phrase will match if all of the terms in the phrase are present in the Tweet, regardless of order and ignoring case. By this model, you can think of commas as logical ORs, while spaces are equivalent to logical ANDs (e.g. 'the twitter' is the AND twitter, and 'the,twitter' is the OR twitter).

The text of the Tweet and some entity fields are considered for matches. Specifically, the text attribute of the Tweet, expanded_url and display_url for links and media, text for hashtags, and screen_name for user mentions are checked for matches.

^a<https://dev.twitter.com/streaming/overview/request-parameters>

La liste des expressions suivies par l'API track est la suivante :

- présidentielle2017
- présidentielle2017
- #presidentielle2017
- primaires2016
- loitravail
- frexit
- Nuit Debout
- nuitdebout
- #nuitdebout
- #Fillon
- #Macron
- #PenelopeGate
- #Hamon2017
- #Valls2017
- #Montebourg2017
- #Marine2017
- #Fillon2017
- #Peillon2017
- #Asselineau2017
- #Macron2017
- #Marinelepen2017
- #Mlp2017 MLP_officiel
- FrancoisFillon
- fhollande
- manuelvalls
- alainjuppe
- nk_m
- JLMelenchon
- NicolasSarkozy
- benoithamon
- EmmanuelMacron
- yjadot
- n_arthaud
- montebourg
- UPR_Asselineau
- jfpoisson78
- JLM2017
- #NDA2017
- dupontaignan
- jeanlassalle
- PhilippePoutou
- n_arthaud
- manuelvalls
- #legranddebat
- presidentielles2017
- présidentielles2017
- présidentielle
- presidentielle
- jevote
- avote
- avoté
- sansmoiLe7mai
- Debat2017
- 2017LeDebat
- jamaismacron
- jamaislepen
- jamaismarine
- jamaislefn
- JeVoteMarine
- JeVoteMacron
- ChoisirLaFrance
- Dupont-Aignan
- enmarche
- frontnational
- #mfga

L'API de Twitter ne donne accès qu'à un échantillon de la totalité des tweets. Mais chaque retweet contient dans ses méta-données le nombre total de retweets jusqu'à la date du retweet considérée. Nous avons ainsi pu estimer la proportion de retweets capturés par notre dispositif en comparant le nombre total de retweets capturés au nombre maximum de retweets mentionné dans les méta-données.

Plus précisément, pour chaque jour t et pour chaque tweet original Tw émis ce jour là, si l'on note $n_t^c(Tw)$ le nombre de retweets de Tw captés et $n_t^n(Tw)$ le nombre

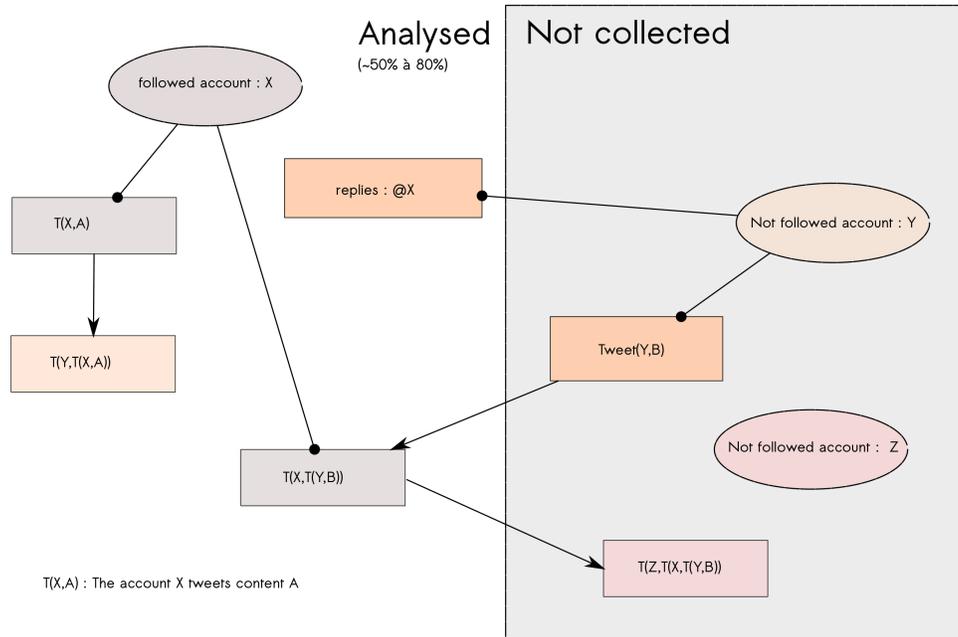


Fig A1. Description du dispositif de capture centré sur les leaders politiques. Celui-ci est complété par un dispositif de capture centré sur les mots-clé (Twitter track API).

maximum de retweets mentionnés dans les méta-données des retweets captés de Tw , nous estimons la proportion de retweets captés par :

$$\nu(t) = \frac{\sum_{\{Tw\}} n_t^c(Tw)}{\sum_{\{Tw\}} n_t^m(Tw)}$$

La figure A2 présente l'évolution de la proportion journalière de retweets capturée ainsi que le nombre total de tweets. Cette proportion est très variable en fonction du jour de collecte et dépend de l'ensemble du trafic Twitter d'une manière qui n'est pas spécifiée par l'API. En moyenne, sur l'ensemble de l'échantillon testé, cette proportion est égale à 34%.

Twitter ne publiant pas le code source de ses APIs, nous ne savons pas comment sont produits les échantillons fournis par les APIs. En particulier, nous ne savons pas si le nombre de retweets mentionnés dans les méta-données d'un retweet entre dans les paramètres de sélection des tweets de l'échantillon. Néanmoins, nous pouvons raisonnablement penser que la proportion de tweets captés (mentions, citations, etc) par notre dispositif est de l'ordre de 34% du total des tweets correspondant à nos critères de capture.

Nos résultats montrent cependant que cette incertitude sur la précision de l'échantillon intrinsèque aux limites imposées par les API Twitter n'est pas un frein à l'analyse de la dynamique des communautés politiques.

2.3 Liste des principales personnalités politiques suivies (API track et follow)

La campagne présidentielle française a été rythmée par deux primaires organisées par les deux partis ayant occupés le pouvoir ces dernières années, le *Parti Socialiste* (gauche) et

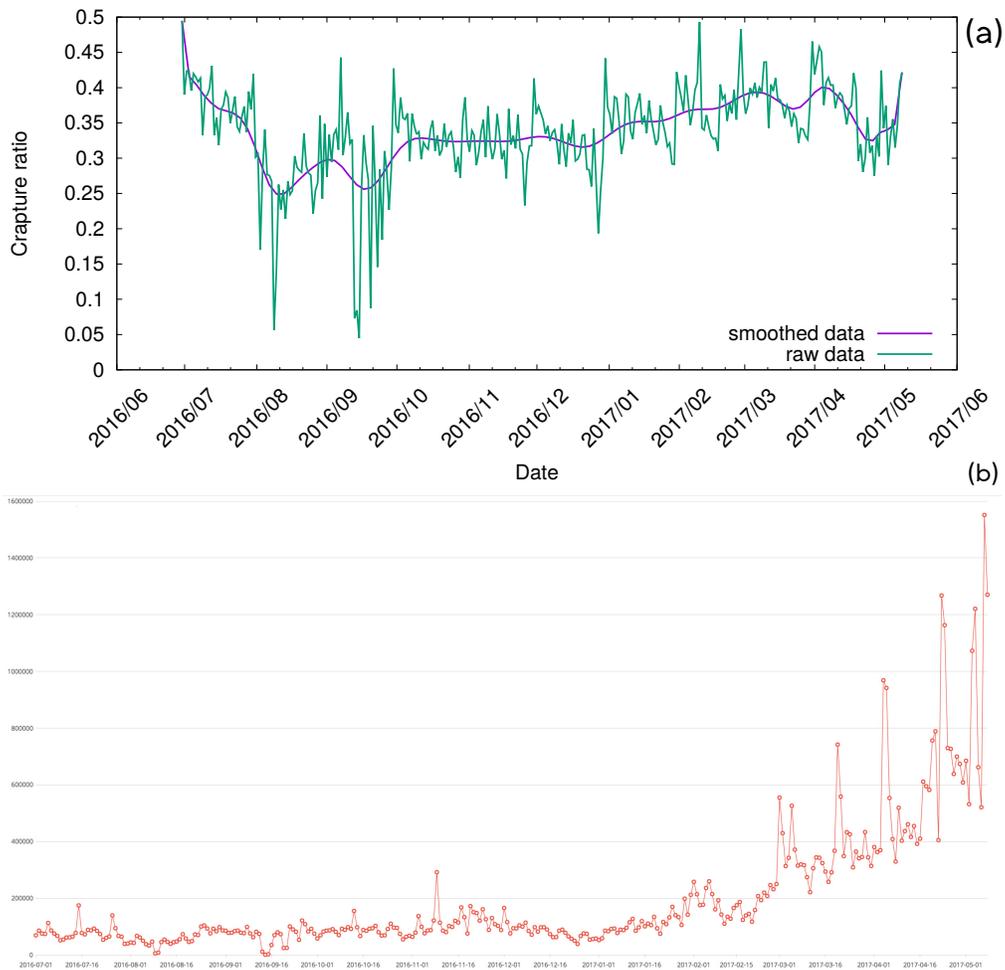


Fig A2. Statistique sur le nombre de retweet capturés (a) en haut: Évolution du taux de capture des retweets au cours du temps. Certains points bas correspondant à des pannes serveurs. (b) en bas: Nombre quotidien de tweets collectés.

Les républicains (droite). Les principaux candidats à ces deux primaires ainsi que l'ensemble des candidats à la présidentielle ont fait l'objet d'un suivi via les API "track" et "follow" (pour les ~ 3700 autres figures politiques, les données ont été collectées via l'API "follow"):

Candidats à la présidentielle (par ordre de score)	Principaux candidats aux primaires
1. Emmanuel Macron (EM)	<ul style="list-style-type: none"> • Yves Jadot (EEV - rallié à Hamon)
2. Marine Le Pen (FN)	<ul style="list-style-type: none"> • Emmanuel Valls (candidat à la primaire de la gauche et premier ministre sortant)
3. François Fillon (LR)	<ul style="list-style-type: none"> • Arnaud Montebourg (candidat à la primaire de la gauche , ancien ministre)
4. Jean-Luc Melenchon (FI)	<ul style="list-style-type: none"> • Vincent Peillon (candidat à la primaire de la gauche et ancien ministre)
5. Benoît Hamon (PS)	<ul style="list-style-type: none"> • François Hollande (Président sortant)
6. Nicolas Dupont-Aignan (DF) (UPR)	<ul style="list-style-type: none"> • Alain Juppé (candidat à la primaire de la droite, favori des sondages jusqu'à la primaire de la droite)
7. Jean Lassalle	<ul style="list-style-type: none"> • Nathalie Kosciusko-Morizet (candidate à la primaire de la droite)
8. Philippe Poutou (NPA)	<ul style="list-style-type: none"> • Nicolas Sarkozy (ex Président, candidat à la primaire de la droite)
9. François Asselineau	<ul style="list-style-type: none"> • Jean-Frédéric Poisson (candidat à la primaire de la droite)
10. Nathalie Arthaud (LO)	

2.4 Architecture de la plateforme

Ce dispositif de collecte a été mis en place le 1er juin 2016 à partir de l'infrastructure Multivac du CNRS/ISC-PIF (<https://multivac.iscpif.fr> cf. figure A3) et a opéré en continu jusqu'au lendemain de l'élection présidentielle à l'exception de rares interruptions pour maintenance de nos serveurs (entre le 13-09-2016 et 15-09-2016 et entre le 21-09-2016 et le 22-09-2016). Entre le 1 juin 2016 et le 7 mai 2017 (deuxième tour de l'élection présidentielle), nous avons ainsi collecté 60 326 216 tweets émis par 2 477 910 comptes uniques.

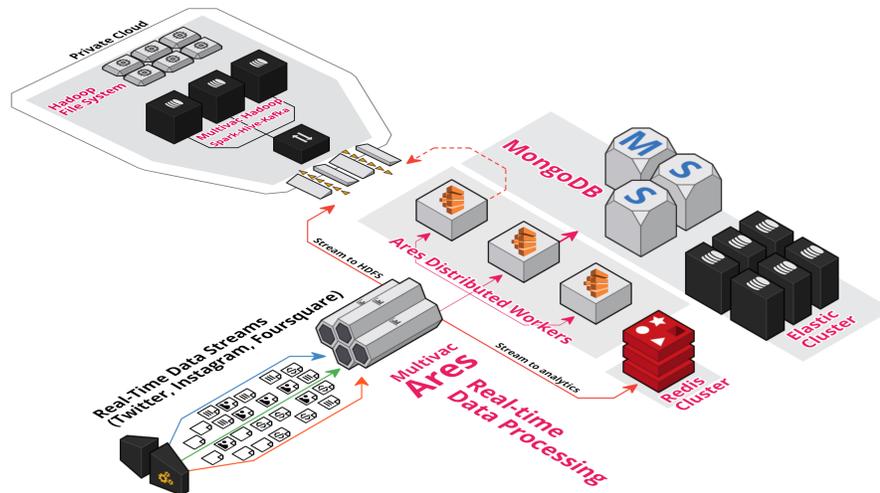


Fig A3. Description synthétique de l'infrastructure Multivac utilisée pour mettre en place le dispositif d'analyse en temps réel des tweets. Cette infrastructure mobilise des technologies telles que Elasticsearch, Apache Hadoop, RabbitMQ, MongoDB ou Redis.

TWITTER DATA ANALYTICS ARCHITECTURE (voir également la figure A3)

Real-time data processing and micro services

- Elasticsearch Cluster (highly available and distributed)
 - Primary database with over 6 billion indexed documents
 - Full-text search
 - Indexing keywords
 - Data aggregation (date histogram, significant, terms, nested, etc.)
- MongoDB
 - Secondary database with over 12 billions of data
- RabbitMQ
 - Distributing Tweets to workers for parallel computations
- Node.js
 - Streaming Tweets by tracking and following Twitter accounts
 - Scalable workers: annotating and tagging Tweets in realtime
 - Extracting Tweets content
- Redis
 - Caching
 - In-memory analysis

Offline and batch processing: Hadoop cluster

- HDFS (Hadoop distributed file system)
- Apache Hadoop YARN: Managing cluster resources
- Apache kafka: Streaming data for batch processing
- Apache Spark
 - ETL
 - Pre-processing data
- Apache Hive: Run SQL queries in Hadoop

2.5 Disponibilité des données

Les données ont été collectées selon les conditions d'utilisation de Twitter: *“If you provide Content to third parties, including downloadable datasets of Content or an API that returns Content, you will only distribute or allow download of Tweet IDs, Direct Message IDs, and/or User IDs.”* (cf.

<https://developer.twitter.com/en/developer-terms/agreement-and-policy#id34>)

Par conséquent pour les données brutes de Twitter, nous avons dû limiter la publication des données aux id des tweets avec un script qui permet de télécharger à partir de Twitter le contenu complet des tweets moins les tweets qui auraient été supprimés. Tout le matériel nécessaire est disponible dans le Harvard Dataverse (fichier A10) doi:10.7910/DVN/6739SP [24]. Il est donc possible de reconstituer la base de données Twitter complète utilisée pour cette étude, moins les tweets supprimés.

Comme mentionné dans les conditions d'utilisation de Twitter ², ces données sont disponibles sous certaines conditions, telles que : *“You may not distribute more than 1,500,000 Tweet IDs to any entity (inclusive of multiple individual users associated with a single entity) within any given 30 day period, unless you are doing so on behalf of an academic institution and for the sole purpose of non-commercial research or you have received the express written permission of Twitter.”*

²<https://developer.twitter.com/en/developer-terms/policy>

Pour faciliter la réutilisation des données, nous avons également construit une infrastructure basée sur l'API disponible sur <https://api.iscpif.fr/docs/#/datap01>. Sur demande et à usage académique uniquement, cette infrastructure donne accès aux données agrégées que Twitter permet de redistribuer. ζ

3 Détection des bots

Les robots ou *bots* sont souvent utilisés pour contrôler un compte Twitter et augmenter sa présence en ligne. Si cet usage était trop massif, il pourrait nous empêcher d’analyser le comportement des militants politiques. Pour s’assurer que la proportion de comptes animés par des robots est à un niveau raisonnable, nous avons utilisés l’API BotOrNot pour détecter les comptes qui pourraient ne pas être animés par des personnes ”réels”.

Nous avons créé deux échantillons d’environ 50 000 comptes chacun. Le premier contient tous les 49 995 comptes qui ont été au moins une fois dans une communauté tout en ayant au moins 3 liens dans le graphe de retweet. Cet échantillon représente les personnes intégrées dans le graphique de retweet. Le deuxième échantillon contient 50 000 utilisateurs qui ont été au moins une fois dans une communauté. Cet échantillon représente des comptes aléatoires dans les communautés.

L’API retourne pour chaque compte sa probabilité d’être un bot. Avec le même seuil de 0.5 utilisé par les concepteurs de l’API pour classer un compte comme un bot, il y a environ 5.3% des comptes qui appartiennent à une communauté qui seraient des bots. Ces comptes ont produit moins de 5% des retweets enregistrés dans les communautés (voir les tableaux A1 et A2).

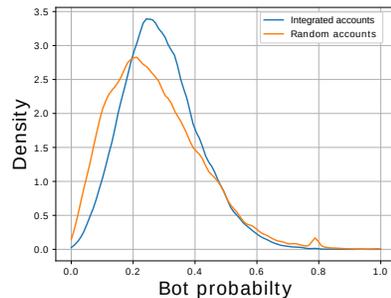


Fig A4. Distribution de la probabilité d’être un bot. Courbe bleue : comptes qui ont été au moins une fois dans une communauté tout en ayant au moins 3 liens dans le graphe de retweet. Courbe orange : comptes aléatoires qui ont été au moins une fois dans une communauté. Cette détection a été effectuée en juin 2016.

Nom	< comptes >	< bots >	% de bot
François Asselineau	213	14	6.5 %
François Fillon	1340	68	5.1 %
Benoit Hamon	561	31	5.6 %
Marine Le Pen	1560	70	4.5 %
Emmanuel Macron	994	44	4.4 %
Jean-Luc Mélenchon	804	30	3.7 %

Table A1. Présence des bot dans les communautés. Nombre moyen de comptes par période et par communauté (< comptes >) ainsi que le nombre moyen de bots détectés par période par communauté (< bots >).

Type de tweet	Bots politiques	Bots aléatoires
Original	7.2%	7.4%
Retweet	4.8%	3.7%
Response	0.9%	1.8%
Quote	3.4%	1.5%
Total	4.7%	3.9%
Proportion des comptes	5.1%	6.8%

Table A2. Proportion de tweets postés par des bots dans les communautés politiques et par des bots parmi les comptes aléatoires. Par exemple, les bots dans les communautés politiques ont émis 7,2% des tweets originaux émis par tous les comptes d'une communauté politique, alors qu'ils ne représentent que 5,1% des comptes.

4 Durée de vie des tweet et motif d'activité sur Twitter

L'activité politique sur Twitter a une certaine régularité qui est le couplage entre une activité endogène (jour/nuit, semaine/week-end, etc.) et une activité exogène qui dépend de l'actualité politique. La moyenne sur l'ensemble du corpus montre clairement une régularité hebdomadaire (voir figure A5).

En ce qui concerne le temps de retweet, 89.005% des retweets ont lieu dans les 24h suivant la parution du tweet original et 98.977% dans les 14 jours.

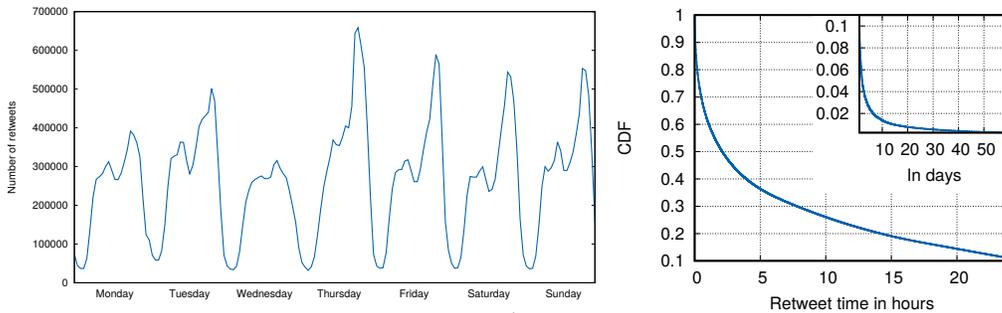


Fig A5. Motif d'activité sur Twitter. À gauche: Nombre de retweets observés par jour et par heure entre le 1er juillet 2016 et le 23 mai 2017; à droite: Distribution cumulative inverse du temps entre un tweet et ses retweets.

5 Stabilité des communautés

Nom	# comptes présents	Temps moyen en jours
Marine Le Pen	35,812	33.8
François Fillon	30,516	30.7
Jean-Luc Mélenchon	42,206	21.8
Emmanuel Macron	57,651	16.2
Benoît Hamon	37,748	17.7
François Asselineau	14,774	11.8
Philippe Poutou	51,908	11.6
Nicolas Dupont-Aignan	30 366	5.6
Nathalie Arthaud	45,450	4.5
Jacque Cheminade	29,153	4.4
Globale	187,619	55.8

Table A3. Comptes par communauté. Tableau indiquant, pour chaque candidat, le nombre de comptes présents au moins une fois dans sa communauté entre le 1er juillet 2016 et le 23 mai 2017. Les communautés de Le Pen et de Fillon sont clairement différentes des autres, en ce sens qu'elles sont les plus stables avec un temps moyen très élevé. Le fait que le temps moyen passé par un compte au sein d'une communauté (dernière ligne) est très élevé indique qu'une proportion significative des comptes changent d'une communauté à l'autre, sans disparaître de Twitter.

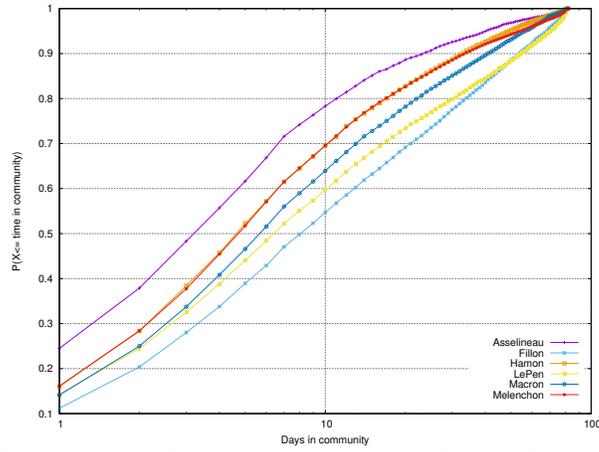


Fig A6. Distribution cumulative du temps passé dans les principales communautés entre juillet 2016 et le premier tour.

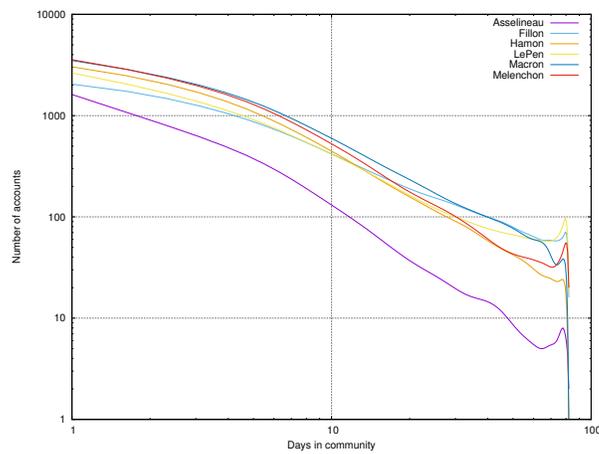


Fig A7. Distribution lissée du temps passé dans les principales communautés entre juillet 2016 et le premier tour.

6 Corrélation entre structure communautaire et motifs de retweets: une confirmation des chambres d'écho

Nous avons analysé le détail des motifs de retweets afin de valider l'hypothèse H_0 , intuitive d'après notre méthode de reconstruction, selon laquelle la diffusion des retweets est fortement contrainte par la structure des communautés calculées. Nous vérifions pour cela que les motifs de retweets dans les communautés sont très différents de ce qu'ils seraient s'ils avaient lieu de manière aléatoire au sein des communautés en analysant les motifs de diffusion pour des tweets ayant un nombre suffisant de retweets (≥ 50). Via notre dispositif, nous avons capté 4 124 888 tweets originaux qui ont totalisé 39 284 569 retweets. Parmi ces 4 millions de tweets, seulement 118 967 ont 50 retweets ou plus mais ils totalisent 21 288 537 des retweets soit 54% des retweets. En imposant que les retweets proviennent de comptes affiliés à une communauté, nous obtenons 82 112 tweets originaux totalisant 12 076 600 retweets soit 31% des retweets. Entre le 1er juin 2016 et le 7 mai 2017, nous avons calculé pour chaque jour j , les N^j communautés politiques fournies par notre méthodologie, notées $\mathcal{C}^j = \{C_i^j\}_{1 \leq i \leq N^j}$, et constitué un sous-corpus de tweets originaux \mathcal{D}_{50}^j composé des tweets ayant été retweetés au moins 50 fois par des personnes dont la communautés est connue le jour j . Les communautés des retweets ayant lieu après le jour j sont celle calculées le jour j , ainsi la détection de communauté et les motifs de retweets sont calculés sur des jeux de données indépendants. Cet approche diffère de celle utilisée dans le reste du papier dans laquelle la communauté calculée chaque jour est utilisée.

Sur cette période, nous avons calculé le nombre q_i^j de retweets de \mathcal{D}_{50}^j produits par l'ensemble des C_i^j . Nous obtenons ainsi la distribution $Q^j = \{q_i^j\}_{1 \leq i \leq N^j}$ des retweets de \mathcal{D}_{50}^j par les communautés C_i^j . À partir de cela, nous calculons la distance χ^2 entre cette distribution et la distribution de retweets $\mathcal{P}^j(tw) = \{p(tw)_i^j\}_{1 \leq i \leq N^j}$ de chacun des tweets tw de \mathcal{D}_{50}^j :

$$\chi^2(P(tw), Q) = \sum_i \frac{(P(tw)_i - Q_i)^2}{Q_i}$$

D'après cette analyse sur l'ensemble des périodes, pour 92% des diffusions, H_0 est vraie avec une p -value inférieure à 10^{-16} . En considérant un seuil de p -value de 1%, 99.8% des tweets émis par les communautés ont une diffusion de retweet très différente d'une diffusion aléatoire dans les communautés.

Cela valide notre hypothèse H_0 : *la propagation des retweets est fortement contrainte par la structure des communautés calculées.*

Afin d'avoir une vue plus précise de la manière dont les communautés structurent cette propagation, nous avons calculé, pour chaque jour j et chaque tweet $tw \in \mathcal{D}_{50}^j$, l'entropie \mathcal{H} de sa propagation en fonction des communautés \mathcal{C}^j , ainsi que, à des fins de comparaison, l'entropie des communautés pour l'ensemble des retweets ce jour-là \mathcal{D}_{50}^j : $\mathcal{H}(P^j(tw)) = -\sum_i P^j(tw)_i \log(P^j(tw)_i)$.

Comme le montre la figure A8, l'entropie de la distribution des retweet au sein des communautés est très faible, avec une valeur moyenne inférieure à 1, ce qui indique que la plupart des retweets d'une cascade de retweets se produisent dans une ou deux communautés³, alors que la distribution journalière des retweets couvre un ensemble

³Par exemple, si un tweet se propage dans une seule communauté, alors l'entropie est de 0. Si elle se propage uniformément dans 4 communautés, alors l'entropie est de 2

beaucoup plus large de communautés (entropie moyenne ⁴ de 2,707, ce qui correspond à la propagation dans 6,532 communautés).

L'analyse de la diffusion des retweet montre donc que la grande majorité des tweets sont retweetés au sein de leur communauté d'origine, comme par exemple pour le tweet de François Fillon montré dans la figure A8: *Il y a eu une étude Rexecode qui montre que mon programme permet la création d'un 1,5 million d'emplois. C'est celui qu'il faut appliquer.* (779 retweets, entropie=0.245).

Nous observons ainsi le rôle des chambres d'écho jouées par les communautés politiques [66], qui structurent la propagation de l'information avec, comme dans le cas d'étude, une configuration multipolaire.

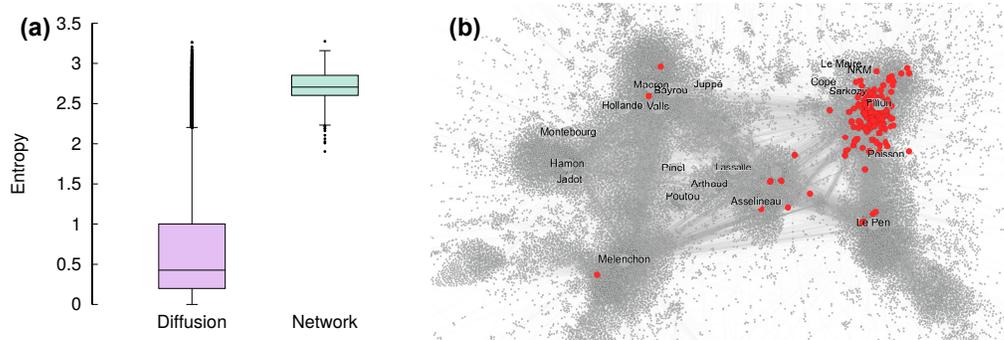


Fig A8. Motifs de diffusion des tweets. (a) à gauche : Boîte à moustache de l'entropie de la diffusion de chaque tweet dans les communautés et de l'entropie de la retweets chaque jour; (b) à droite : Exemple de la diffusion privilégiée dans certaines communautés d'un tweet de François Fillon. (La carte de référence est figure 4).

⁴Il y a une différence entre l'entropie moyenne et l'entropie pour tous les retweets, rapportée dans le tableau 9, car l'entropie moyenne pour chaque jour et l'entropie sur toute la période ne sont pas équivalentes.

7 Analyse du contexte sémantique des communautés

7.1 Définition des thèmes par requêtes

À l'aide des programmes présidentiels de 2017 et des corpus Twitter, nous avons sélectionné onze thèmes que nous avons identifiés comme particulièrement importants et représentatifs des débats politiques. Ces thèmes sont des groupes de termes identifiés par *Gargantext* dans les programmes présidentiels de 2017 (Fig. 3), auxquels nous avons enlevé des termes trop polysémiques susceptibles de générer des faux négatifs; et ajouté des mots-clés spécifiques identifiés dans les corpus Twitter tels que *frexit*, *muzz* ou *#cyberthreats*. Notre objectif n'était pas d'être exhaustif mais de mettre en évidence l'hétérogénéité de l'attention des communautés politiques sur les thèmes et leurs spécialisations respectives sur certains thèmes majeurs. L'objectif était d'obtenir une requête capable d'extraire de la base de données *Politoscope* un ensemble de tweets cohérents liés à ce thème. Une version csv de cette liste est disponible en ligne dans le Fichier S8.

Les requêtes suivantes ont été retenues :

- Emploi** : travailleur* OR emploi* OR "apprentissage" OR "chômage" OR chômeur* OR "rsa" OR "smic" OR "flexisécurité" OR "formation professionnelle" OR "insertion professionnelle" OR "lycées professionnels" OR "formation initiale" OR "heures supplémentaires" OR "temps partiel" OR "contrat de professionnalisation" OR "conditions de travail" OR "emplois précaires" OR "enseignement professionnel et technologique" OR "compensation salariale" OR "39 heures" OR "charges sociales salariales" OR "employeurs" OR "filières professionnelles" OR "35 heures" OR "tutorat" OR "pôle emploi" OR "création d'entreprise" OR "salaire horaire" OR "négociation salariale" OR "exonérations de cotisation sociale" OR "insertion des jeunes" OR "parcours professionnels" OR "licenciement*" OR "travail obligatoires" OR "jeune en alternance" OR "37 heures" OR "contrat de génération" OR "loi travail" OR "augmenter le salaire" OR "prime pour l'emploi" OR "code du travail" OR "formation des fonctionnaires" OR "dialogue social" OR "salaire des fonctionnaires" OR "loi el khomri" OR "sous-traitants" OR "devenir entrepreneurs" OR "souffrance au travail" OR "gel des traitements" OR "obligation pour eux d'accepter un emploi" OR "maîtrise de la masse salariale" OR "non remplacement de un fonctionnaire sur deux" OR "suppressions de postes" OR "artisanat" OR "réduction du temps de travail" OR "RTT" OR "cice" OR "mobilité professionnelle" OR "âge légal de départ à la retraite" OR "retraite solidaire" OR "pénibilité" OR "salaire moyen" OR "cdi" OR "cdd" OR "délocalisations" OR "relocalisation" OR "revenu universel" OR "revenuuniversel" OR "contrat de travail" OR "syndicalisation" OR "cotisations sociales" OR "détachement des travailleurs" OR "travailleurs détachés" OR "délocalisation" OR "relocalisation" OR "plans sociaux"
- Économie & fiscalité**: révolution fiscale" OR "coût de la vie" OR "pouvoir d'achat" OR "libre échange" OR "libéralisme économique" OR "économie sociale et solidaire" OR "entreprise" OR "pme" OR "pmi" OR "actionnaire*" OR "relocalisation*" OR "crédit impôt" OR "nationalisation*" OR "commerçant*" OR "plus-values" OR "délocalisation*" OR "scop" OR "niches fiscales" OR "gafa" OR "cice" OR "taux d'imposition" OR "retenue à la source" OR "crédit impôt-recherche" OR "impôt sur les sociétés" OR "services innovants" OR "marchés publics" OR "livret d'épargne industrie" OR "exportation" OR "impôt sur la fortune" OR "conseils d'administration" OR "small business act" OR "surplus" OR "capital-risque" OR "comités de rémunération" OR "représentants des salariés" OR "harmonisation fiscale" OR "complexité administrative" OR "reprise d'entreprises" OR "stock options" OR "start-ups" OR "pme européennes" OR "commandes publiques" OR "fiscalité européenne" OR "impôt minimal" OR "chiffre d'affaires mondial" OR "prélèvement à la source" OR "échanges commerciaux" OR "travailleurs handicapés" OR "commerce indépendant" OR "politique commerciale" OR "développement économique" OR "politique commerciale" OR "charges patronales" OR "tva anti-délocalisations" OR "allègements de charges" OR "cice" OR "alléger les charges patronales" OR "allègement des charges patronales" OR "centrales d'achat" OR "concurrence internationale" OR "exploitations familiales" OR "financiarisation" OR "travailleurs indépendants" OR "liberté syndicale" OR "austérité" OR "fisc*" OR "impôt" OR "imposition" OR "cotisations sociales" OR "retenue à la source" OR "tranche supplémentaire" OR "isf" OR "prélèvement à la source" OR "cotisations chômage" OR "exonérations de cotisation sociale" OR "tva" OR "charges patronales" OR "tva" OR "csg" OR "taxe" OR "allègements de charges" OR "cice" OR "trésor public" OR "droits de mutation" OR "dette des états" OR "dette de l'états" OR

"dette publique" OR banque* OR financi* OR "subprimes" OR bancaire* OR "épargne" OR spécul* OR "livret A" OR "agios" OR "taux d'usure" OR emprunt* OR "agence de notation" OR "crédit à la consommation" OR "surendettement" OR "assurance-vie" OR "loi sapin" OR "banque centrale européenne" OR "euro-obligations" OR "pacte de responsabilité" OR "taxe mondiale sur les transactions financières" OR "taxe Tobin" OR "dépôts des particuliers" OR "crise bancaire" OR "crise financière" OR trader* OR "ras-le-bol fiscal"

- **Démocratie:** front républicain" OR "ni-ni" OR "non-cumul des mandats" OR "conseil constitutionnel" OR "ve république" OR "6e république" OR "6eme république" OR "5e république" OR "vie politique" OR "données publiques" OR "%penelope%" OR "%pénélope%" OR "%pénélope%" OR "démocratie sociale" OR "cour des comptes" OR "proportionnelle" OR "parité" OR "assemblée nationale" OR "élections législatives" OR "fonction publique territoriale" OR "formation des fonctionnaires" OR "durée du quinquennat" OR "préfet" OR "établissements publics" OR "cour européenne des droits de l'Homme" OR "préférence nationale" OR "nationalistes" OR "patriotes" OR "progressistes" OR "%penelope%" OR "parlement" OR élu* OR "corruption" OR "ena" OR "révisions constitutionnelles" OR "réserve parlementaire" OR "statut pénal du chef de l'état" OR scandale OR (marine AND 300000) OR (marine AND "300 000") OR ("Le Pen" AND 300000) OR ("le pen" AND "300 000") OR "inéligibilité" OR "fonction présidentielle" OR "e-démocratie" OR "49.3" OR "déclaration de patrimoine" OR "égalité républicaine"
- **Recherche & Éducation:** lycée* OR "pisa" OR "éducation" OR apprentissage* OR "maternelle" OR "décrochage scolaire" OR "scolarité" OR "scolaire" OR "service civique" OR "savoirs fondamentaux" OR "méthodes pédagogiques" OR "enseignement professionnel et technique" OR "établissements scolaires" OR "tutorat" OR "échec scolaire" OR "écoles d'ingénieurs" OR "financer leurs études" OR "banque de la jeunesse" OR "écoles de commerce" OR "e-learning" OR "éducation artistique" OR "enseignement primaire" OR universit* OR "la recherche" OR chercheur* OR "R&D" science* OR scientifique* OR enseignant* OR professeur* OR "loi lru" OR "enseignants-chercheurs" OR "enseignement supérieur" OR "grandes écoles" OR "investissements d'avenir" OR "loi foraso" OR "crédit impôt-recherche" OR "conseil national des universités" OR "CNU" OR "autonomie des universités" OR "universités" OR "ZEP"
- **Immigration & multiculturalisme:** migr* OR immigr* OR "nationalité" OR clandest* OR "droit de séjour" OR "schengen" OR "pays d'origine" OR "contrôle des frontières" OR "réadmission" OR "régularisations" OR "regroupement familial" OR "naturalisation" OR "reconduites à la frontière" OR "titres de séjour" OR "procédures d'expulsion" OR "séjour des étrangers" OR "libre circulation des personnes" OR "connaissance préalable de la langue française" OR "intégration des étrangers" OR "citoyenneté française" OR "droit du sol" OR "droit du sang" OR islam* OR musul* OR muzz* OR salaf* OR "burkini" OR "barbarie" OR "laïcité" OR juif* OR juive* OR chrétien* OR "communautarisme" OR "multiculturalisme" OR "modèle républicain" OR "identité nationale" OR "civilisation française" OR "port du voile" OR "préférence raciale" OR race OR "demande d'asile" OR "submersion migratoire"
- **Santé:** santé" OR médecin* OR médic* OR hôpit* OR "sécurité sociale" OR "autisme" OR handicap* OR "mdph" OR "désamiantage" OR "cancer" OR "assurance-maladie" OR "hébergement en établissement" OR "maintien à domicile" OR "soins d'urgence" OR "urgences sanitaires" OR "maladies" OR "perturbateurs endocriniens" OR "accès aux soins" OR "hospitalisation" OR "soins dispensés" OR "parcours de soins" OR "centre de soin" OR "malentendants" OR "malvoyants" OR "dépassements d'honoraires" OR "transports sanitaires" OR "arrêt de travail" OR "numerus clausus" OR "carte vitale" OR "perte d'autonomie" OR "ehpad" OR "agences de sécurité sanitaire et alimentaire" OR "tiers payant" OR "recours aux génériques" OR "lunettes"
- **Écologie & environnement:** écolo* OR "biodiversité" OR "protection environnementale" OR "protection de l'environnement" OR pollution* OR "diesel" OR "ogm" OR "changement climatique" OR "fessenheim" OR "flamanville" OR "gaz de schiste" OR "pesticides" OR "gaz à effet de serre" OR "ressources naturelles" OR "quotas de pêche" OR "agroécologique" OR "énergies marines renouvelables" OR "pêche illégale" OR "performance énergétique" OR "renovation thermique" OR "isolation de l'habitat" OR "précarité énergétique" OR "isolation de l'habitat" OR "forêts" OR "sylvicultures" OR "changements environnementaux" OR "services écosystémiques" OR "déforestation" OR "aires protégées" OR "émissions de carbone" OR "émission de CO2" OR "empreinte carbone" OR "gaz à effet de serre" OR "politique climatique" OR "durabilité" OR "obsolescence programmée" OR "niveau des mers" OR "niveau de la mer" OR "vagues de chaleur"
- **Politique énergétique:** politique énergétique" OR "électricité" OR "nucléaire" OR "énergies renouvelables" OR "éolien" OR "énergies marines renouvelables" OR "énergie renouvelable" OR

"énergie solaire" OR "epr" OR "flamanville" OR "charbon" OR "centrales thermiques" OR "diesel" OR "fessenheim" OR "flamanville" OR "éol*" OR "performance énergétique" OR "fukushima" OR "rénovation thermique" OR "isolation de l'habitat" OR "précarité énergétique" OR "gaz de schiste" OR "énergies marines renouvelables" OR "isolation de l'habitat" OR "sources d'énergie" OR "transition énergétique" OR "conversion écologique" OR "énergies alternatives" OR "dépendance énergétique" OR "combustibles fossiles" OR "énergies fossiles" OR "efficacité énergétique"

- **Politique étrangère:** europ* OR souverain* OR "CETA" OR "TAFTA" OR "bruxelles" OR "russie" OR "Poutine" OR "les russes" OR "pacte de stabilité" OR "directive européenne" OR "traité franco-allemand" OR "trump" OR "washington" OR "maison blanche" OR "moscou" OR "crise syrienne" OR "libye" OR "iran" OR "irak" OR "syrie" OR "OTAN" OR "traité transatlantique" OR "traités internationaux"
- **Sécurité intérieure:** terror* OR "attentat" OR "cyber-sécurité" OR "guerre civile" OR "surveillance" OR "délinqu*" OR "espion*" OR "loi taubira" OR "vidéo-surveillance" OR "vidéo-protection" OR "sécurité des français" OR "laxisme judiciaire" OR "prescription" OR "peines planchers" OR "remises de peine" OR "Daech" OR "cyber-attaques" OR "cyber-attaque" OR "#cyberthreats" OR "#cyberthreat" OR "cyber-guerre" OR "cyber guerre" OR "cyber-menace" OR "#cybersecurity" OR "cyber-sécurité" OR "cyber sécurité" OR "cybersécurité" OR "cyberdéfense" OR "cyberarmée" OR "daech" OR "daesh"
- **Agriculture:** agricult* OR agro* OR "pac" OR "alimentation" OR "engrais" OR "circuits courts" OR "transition agricole" OR "surplus" OR intrant* OR "bio" OR "pesticides"

7.2 Distribution des thèmes par communautés politiques

Nous avons étudié la diversité des communautés discutant de chaque thème. Pour cela, nous avons calculé l'entropie des communautés parlant sur chaque thème (voir Tableau A4).

Bien que la plupart des thèmes aient une entropie assez similaire, il semble que les thèmes *immigration et multiculturalisme* et *sécurité intérieure* soient discutés par un plus petit nombre de communautés, car leur entropie est faible. Inversement, le thème *emploi* est celui qui est le plus uniformément discuté par toutes les communautés. En plus de l'entropie, nous montrons aussi, pour chaque thème, les deux communautés parlant le plus spécifiquement sur ce thème. Comme les communautés sont de tailles très diverses, les communautés choisies sont celles ayant le score de *tf-idf* le plus élevé sur le thème et non celles mentionnant le plus le thème.

Ainsi, si une petite communauté ne parle que d'un seul thème, elle peut toujours faire partie des principales communautés sur ce thème même si, en valeur absolue, elle participe moins que les autres. Le score *tf-idf* d'une communauté c_j pour un thème t_i est calculé comme suit :

$$tf-idf(c_j, t_i) = \frac{tweets(c_j, t_i)}{\sum_k tweets(c_k, t_i)} \log \left(\frac{\sum_{k,l} tweets(c_k, t_l)}{\sum_l tweets(c_j, t_l)} \right),$$

où $tweets(c_j, t_i)$ est le nombre de tweet de la communauté c_j sur le thème t_i .

Nous avons également étudié la diversité des thèmes discutés dans chaque communauté en calculant l'entropie des thèmes dans chaque communauté (cf. tableau 8).

Sur ce point, il y a des disparités entre les candidats. Cependant, les thèmes que nous avons définis ne sont pas exhaustif et ne permettent pas de conclure sur la diversité des thèmes abordés par les candidats.

Afin de mieux comprendre la représentativité de ces thèmes et la présence de chaque candidat, la proportion de tweets appartenant à un thème et les proportions de chaque thèmes sont présentées dans les figures A9 à A14.

Thème	Entropie	# communautés	Communautés principales
Agriculture	2.37	5.2	François Fillon, Marine Le Pen
Économie & fiscalité	2.32	5.0	François Fillon, Jean-Luc Mélenchon
Emploi	2.41	5.3	Jean-Luc Mélenchon, Benoît Hamon
Démocratie	2.32	4.9	Emmanuel Macron, Marine Le Pen
Écologie & environnement	2.14	4.4	Benoît Hamon, Jean-Luc Mélenchon
Immigration & multiculturalisme	1.58	3.0	Marine Le Pen, François Fillon
Recherche & Éducation	2.41	5.3	François Fillon, Emmanuel Macron
Santé	2.38	5.2	François Fillon, Benoît Hamon
Sécurité intérieure	2.07	4.2	Marine Le Pen, François Fillon
Politique énergétique	2.19	4.6	Jean-Luc Mélenchon, Benoît Hamon
Politique étrangère	2.37	5.2	Marine Le Pen, Emmanuel Macron
Tous thèmes	2.31	5.0	Marine Le Pen, François Fillon

Table A4. entropie des 3–communautés discutant chaque thème. Les 2 communautés principales selon le $TF - IDF$ sont également présentées pour chaque thème. L'entropie maximale est de 2.585 lorsque les 6 communautés ont le même volume de tweet sur ce thème.

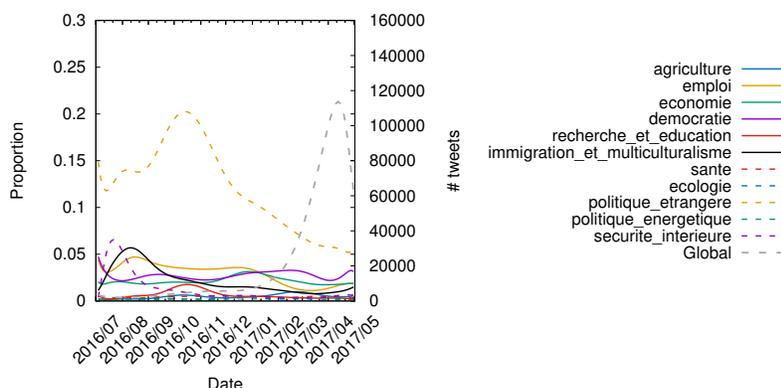


Fig A9. Évolution des thèmes discutés par la communauté Asselineau entre le 01 Juin 2016 et le 08 Mai 2017.

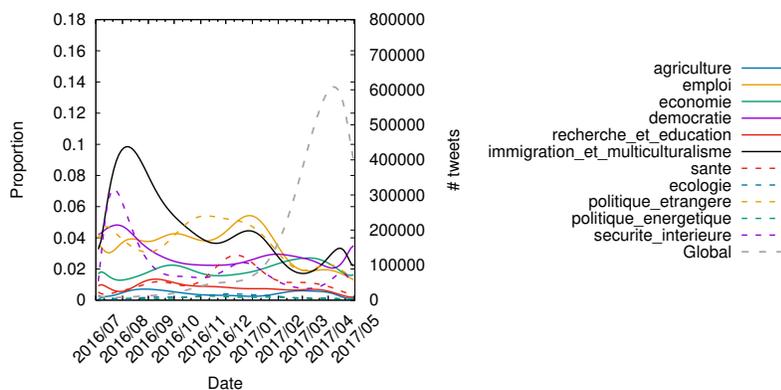


Fig A10. Évolution des thèmes discutés par la communauté Fillon entre le 01 Juin 2016 et le 08 Mai 2017.

7.3 Évolution sémantique des communautés

La plate-forme accompagnant cet article, le *Politoscope*.

(<https://presidentielle2017.politoscope.org/dashboard>) permet de suivre l'utilisation des

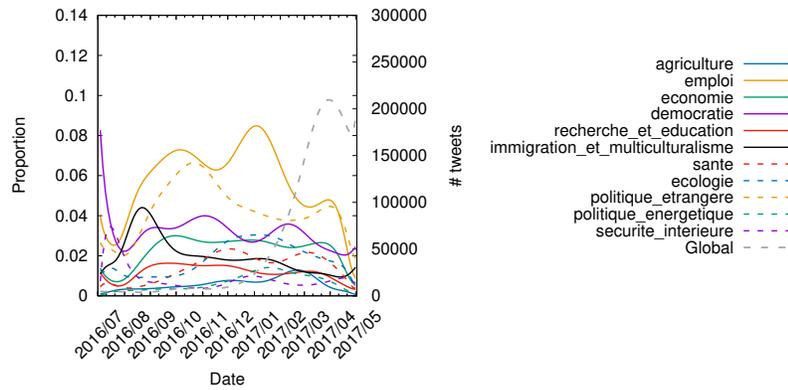


Fig A11. Évolution des thèmes discutés par la communauté Hamon entre le 01 Juin 2016 et le 08 Mai 2017.

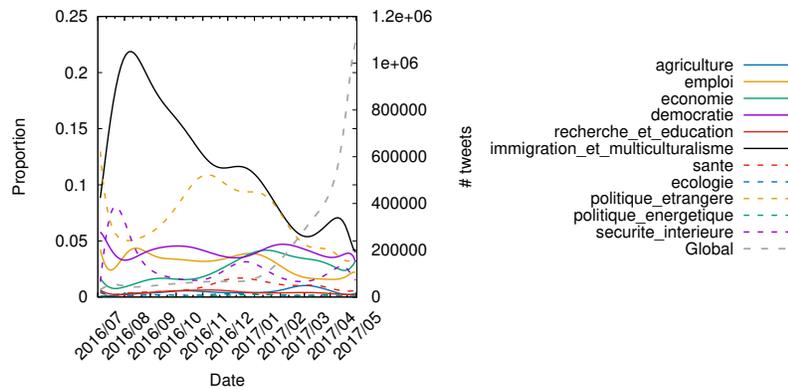


Fig A12. Évolution des thèmes discutés par la communauté Le Pen entre le 01 Juin 2016 et le 08 Mai 2017.

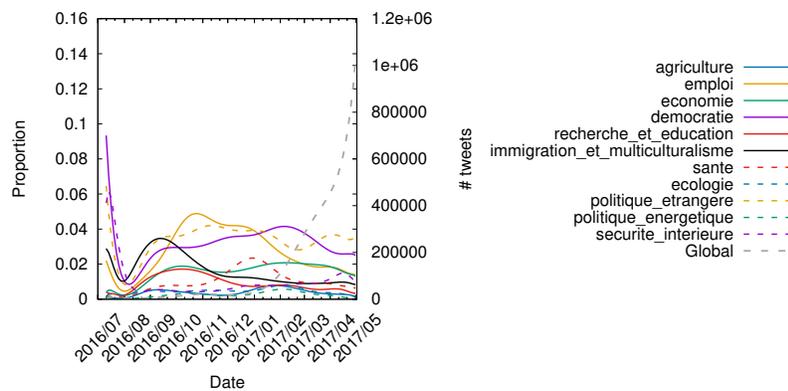


Fig A13. Évolution des thèmes discutés par la communauté Macron entre le 01 Juin 2016 et le 08 Mai 2017.

15 mots-clés les plus spécifiques à chacune des principales communautés politiques lorsqu'ils traitent de chacun des onze thèmes sélectionnés. Ces mots-clés ont été

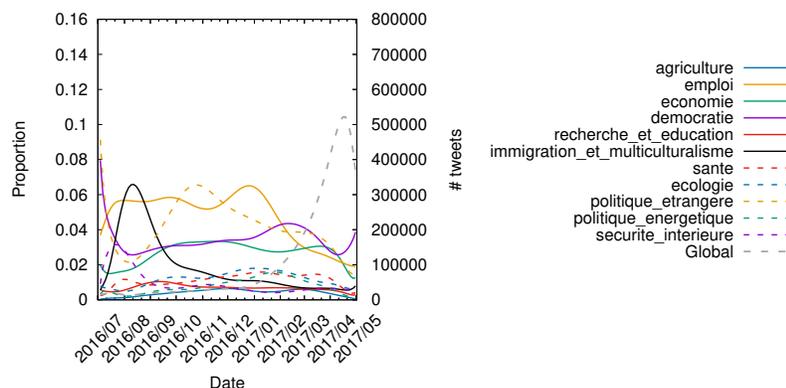


Fig A14. Évolution des thèmes discutés par la communauté Mélenchon entre le 01 Juin 2016 et le 08 Mai 2017.

identifiés à l'aide du test du χ^2 implémenté dans Elasticsearch⁵.

On peut observer qu'en plus des spécificités sémantiques locales au niveau des communautés, l'utilisation du vocabulaire évolue au fil du temps pour les différentes communautés, certains termes apparaissant, se renforçant ou disparaissant au fil du temps.

Voici quelques exemples concernant le thème de l'emploi. Le lecteur peut se référer à la plateforme pour une exploration exhaustive.



Fig A15. Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Macron pour le thème de l'emploi.

⁵Voir <https://www.elastic.co/guide/en/elasticsearch/reference/current/search-aggregations-bucket-significantterms-aggregation.html> et Manning et Manning et al, Chapitre 13.5.2 - <https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>



Fig A16. Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Le Pen sur le thème de l'emploi.



Fig A17. Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Fillon sur le thème de l'emploi.



Fig A18. Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Mélenchon sur le thème de l'emploi.



Fig A19. Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Hamon sur le thème de l'emploi.



Fig A20. Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Asselineau sur le thème de l'emploi.

8 Nettoyage des données sur les fausses nouvelles

Afin de suivre les fausses nouvelles et les débunks dans le *Politoscope*, nous avons normalisé toutes les urls contenues dans les tweets. Pour éviter les faux négatifs causés par l'utilisation d'un service de raccourcisseur d'url tel que *bit.ly* ou *fb.me*, nous avons remplacé les urls raccourcies par les urls réelles. Nous nous sommes également assurés d'ignorer les paramètres d'url, tel que *?utm_source=*, ajoutées pour suivre les personnes et nous n'avons également pas pris en compte l'utilisation de *http* ou *https* comme ces deux protocoles se réfèrent au même site Web.

Nous avons examinés manuellement la liste des 5 148 urls de fausses nouvelles fournies par le *Decodex*. Nous avons supprimé 4 urls que nous ne considérons pas comme contenant de fausses nouvelles à cause de leur présentation ouvertement parodique par le site web ou du contenu peu clair au moment de notre examination (entre crochets, le nombre de tweets mentionnant le lien dans \mathcal{D}):

- [158] <https://www.buzzbeed.com/macron-veut-faire-payer-un-loyer-aux-proprietaires>,
- [3,642] <http://www.legorafi.fr/2017/04/23/la-belgique-annonce-la-construction-dun-mur-pour-repousser-les-vagues-dimmigrants-francais>,
- [98] <http://www.legorafi.fr/2016/06/01/emmanuel-macron-quand-je-serre-la-main-dun-pauvre-je-me-sens-sale-pour-toute-la-journee>,
- [54] <http://candidat-2017.fr>

Cela n'a pas changé le nombre d'histoires.

Nous avons vérifié que les mentions dans les tweets d'URLs liées aux fausses nouvelles (resp. debunks) participent à la diffusion des fausses nouvelles (resp. debunks) associées, au sein des communautés politiques en ligne. Pour ce faire, nous avons sélectionné au hasard un échantillon de 100 tweets contenant des liens vers des fausses nouvelles et un échantillon de 100 tweets contenant des liens vers des débunk, afin de les évaluer manuellement. 100% des tweets avec un lien vers une fausse nouvelle diffusaient cette fausse nouvelle - 95% $CI = [96.3\% ; 100\%]$ ⁶ et 98% des tweets avec un lien de débunk participaient au débunk - $CI = [93.0\% ; 99.8\%]$.

⁶L'intervalle de confiance est Clopper-Pearson.

9 Vocabulaire du discours politique lors des élections présidentielles de 2017

Les termes considérés comme équivalents ont été regroupés et apparaissent dans les analyses suivantes sous une étiquette qui est le premier terme du groupe. Une version csv de cette liste de termes est fournie par sous le DOI:10.7910/DVN/AOGUIA dans le fichier A9.

Liste des groupes de termes extraits à l'aide du logiciel Gargantext (<https://gargantext.org>).

- #frexit OR frexit
- 18-25 ans
- 300 000 euros OR 300 000 €OR 300 000€
- 35 heures OR 35h
- 37 heures
- 39 heures
- 6ème république OR 6emerepublique OR 6erepublique OR 6érépublique
- abattements fiscaux
- abattoirs
- abroger OR abrogation
- abus de position dominante
- accès à la propriété
- accès aux soins
- accessibilité des personnes handicapées
- accord de libre-échange
- accords bilatéraux
- accords de schengen
- accords du touquet
- accueil des réfugiés
- accusé
- actifs
- action économique
- action publique
- actionnaires
- actionnariat salarié
- actions d'entreprises
- activités de conseil
- administration centrale
- administration fiscale
- administration pénitentiaire
- administratives OR administration OR administrations
- aéroports
- affaires criminelles
- afghanistan
- afp
- afrique
- âge légal de départ à la retraite
- agence publique européenne de notation
- agences de sécurité sanitaire et alimentaire
- agents publics
- agglomération
- agios
- agriculture OR agriculteurs OR agricole OR agricoles OR agriculteur
- agriculture biologique OR agriculture bio
- agroécologique
- aide à la construction
- aide médicale d'état OR aide médicale d'etat
- aide sociale à l'enfance
- aides publiques
- aides sociales OR aide sociale
- airfrance
- algérie
- alimentation
- allègement
- allègements de charges
- allègements fiscaux
- alléger les charges patronales
- allemagne
- allocation adulte handicapé OR aah
- allocation aux adultes handicapés
- allocations OR allocs
- allocations chômage
- allocations familiales OR allocation familiale
- allocations sociales
- alpes-maritimes
- alstom
- alternance
- alternance forte
- aluminium
- alzheimer
- aménagement du territoire
- amendes
- anciens combattants
- animaux
- ankara
- apa
- apl
- apologie du terrorisme
- apologie du terrorisme
- apple
- apprentis
- apprentissage OR apprentissages
- l'arabie saoudite OR l'arabie saoudit
- argent public
- armée d'afrique
- armées OR armée
- arrangements fiscaux
- arrêt de travail
- artisanat local
- artisans OR artisanat
- artistes OR artiste
- aspa
- assange
- assassinat politique
- assassiner OR assassins
- assemblée constituante
- assemblée nationale
- assimilation
- assistance médicalisée

- assistant parlementaire OR assistante parlementaire
- associations OR association
- assurance-chômage
- assurance-maladie OR assurance maladie
- assurance-vie
- atteindre l'équilibre budgétaire
- attentatnice
- attentats OR attentat
- augmenter
- augmenter le salaire
- austérité
- autisme
- autonomie aux communes
- autonomie fiscale
- autorité
- autorité de l'état
- autorité hiérarchique
- autorité indépendante
- autorités françaises
- autoroutes
- auxiliaire de vie scolaire
- avions ravitailleurs
- avocats
- avortement
- avs
- ayant-droits
- baccalauréat
- bachar al-assad
- baisse d'impôts OR baisses d'impôts OR baisse d'impôt
- baisse des dépenses interventions de l'état
- baisser
- baisser les prix
- balkany
- banlieues OR banlieue
- banque centrale
- banque centrale européenne OR bce
- banque de détail
- banque de données
- banque de la jeunesse
- banque postale
- banque publique d'investissement
- banques OR banque
- banques d'affaires
- banques de dépôt OR banques de dépôts
- barbarie OR barbare OR barbares
- barbarie terroriste
- bas revenus
- bâtiments
- bâtiments publics OR bâtiments nationaux
- bayrou
- bénéficiaires du rsa
- bénévoles
- benoît hamon
- bi-langues
- bibliothèques
- bien-être animal
- bien-être des animaux
- biens communs
- bio
- biodiversité
- biogaz
- blessés
- blindés
- blocages
- bois
- bonus
- bonus-malus
- bonus/malus
- branches professionnelles
- brexit OR #brexit
- bricolage
- bruxelles
- budget européen
- budgets participatifs
- bureaucratie
- burkini OR #burkini
- burnout
- burqa
- buy european act
- bygmalion
- cadeaux fiscaux
- caf
- caisse des dépôts et consignations
- caissière
- calais
- cancers OR cancer
- cannabis
- cantine
- cantines bio
- cantines scolaires
- capital-risque
- caricature OR caricatures
- carnées
- carrières judiciaires
- carte de séjour
- carte vitale biométrique
- casier judiciaire
- caution solidaire
- caution universelle
- cavous
- cdd
- cdi OR contrat à durée indéterminée
- censure
- centrales OR central OR centrale
- centrales d'achat
- centrales nucléaires
- centres éducatifs fermés
- ceseda
- ceta
- chaînes publiques de télévision
- chambres régionales des comptes
- champ d'honneur
- changement climatique
- charges OR charge
- charges patronales
- charges salariales
- charges sociales salariales
- chef d'entreprise OR chefs d'entreprise OR chefs d'entreprises
- chef de l'état
- chefs d'état OR chef d'état OR chefs d'état
- chiffre d'affaires mondial
- chine
- chirurgie ambulatoire
- chômeur de longue durée
- chômeurs OR chômeur OR chômage
- chrétiens d'orient OR chretiensdorient
- cice
- circuits courts
- citoyenneté française

- citoyens
- civilisation européenne
- civilisation française
- civilisations OR civilisation
- clandestine
- clandestins OR clandestin
- classe politique
- classes moyennes
- clause générale de compétence
- clients des banques
- climat
- clubs professionnels
- cnsfp
- cocaïne
- code civil
- code de l'entrée et du séjour des étrangers
- code de la nationalité française
- code du travail
- cohésion nationale
- cohésion sociale
- collectivités OR collectivité
- collectivités d'outre-mer
- collectivités locales
- collectivités territoriales OR collectivité territorial
- collègues OR collègue
- combattants
- comités de rémunération
- commandes publiques OR commande publique
- commerçants
- commerçants indépendants
- commerce
- commerce indépendant
- commission européenne
- communautarisme OR communautarismes
- communautés urbaines
- communiste OR communistes
- compassion
- compensation salariale
- compétitivité
- complexité administrative
- complot OR complotiste
- compte bancaire
- compte individuel de formation
- comptes publics
- concertation
- concurrence OR concurrents
- concurrence déloyale
- concurrence internationale
- concurrence loyale
- concurrence sociale déloyale
- condamnations OR condamnation OR condamnés
- conditions de travail
- confession juive OR juif OR juifs
- confiance confiance
- conflits d'intérêt
- congé parental
- conjoints survivants
- connaissance préalable de la langue française
- conseil constitutionnel
- conseil de sécurité
- conseil national des universités
- conseil supérieur de l'audiovisuel
- conseil supérieur de la magistrature OR csm
- conseiller territorial
- conseils d'administration
- conseils généraux
- conseils régionaux
- consommations OR consommer OR consommation OR consommateurs
- consommation responsable
- constitution
- construction
- construction européenne
- contrat de génération
- contrat de professionnalisation
- contrats d'assurance
- contribution fiscale
- contrôle de légalité a priori
- contrôle des frontières OR contrôle de ses frontières
- contrôles d'identité
- convention de dublin
- convention européenne des droits de l'homme
- conventions citoyennes
- conversion écologique
- conversion écologique et énergétique
- convocation
- convocation judiciaire OR convocation des juges
- cop21
- corporatisme
- correctionnalisation
- corruption
- cotisation vieillesse
- cotisations OR cotisation OR cotiser
- cotisations chômage
- cotisations sociales
- coup d'État institutionnel
- coupables
- couples homosexuels
- cour des comptes
- cours d'assises
- coût de la vie
- coût du travail
- couverture mobile
- création artistique
- création d'entreprise
- crèches
- crédit à la consommation
- crédit impôt OR crédit d'impôt
- crédit impôt réparation
- crédit impôt-recherche OR crédit d'impôt recherche OR crédit impôt recherche
- crédits OR crédit
- crimes OR crime
- criminel OR criminels
- crise
- crise bancaire
- crise bancaire économique
- croissance
- csg
- culture OR cultures
- culture française
- cumul

- cyber-attaques OR cyber-attaque OR #cyberthreats OR #cyberthreat OR cyber-guerre OR cyber-guerre OR cyber guerre OR cyber-menace
- #cybersecurity OR cyber-sécurité OR cyber sécurité OR cybersécurité
- cyberdéfense OR cyberarmée
- daech OR daesh
- décentralisation
- déchets
- décisions de remise en liberté
- déclassement
- découverts bancaires
- décrochage scolaire OR déscolarisé
- décrocheurs
- déductibilité fiscale des dons
- défense de la laïcité
- défense européenne commune
- défense nationale
- déficit
- défiscalisations OR défiscaliser OR défiscalisation OR défiscalisées
- dégradation du mobilier urbain
- délais de jugement
- délinquants étrangers
- délit OR délits OR délinquance OR délinquante OR délinquants
- délocalisations OR délocalisent OR delocalisees OR délocalisé
- demandes d'asile OR demandeurs d'asile
- dématérialisation OR dématérialisée
- démocratie OR démocraties OR democracy OR democracies
- démocratie sociale
- dentaires
- départements OR département
- dépassements d'honoraires
- dépendance
- dépendance au pétrole
- dépenses administratives
- dépenses culturelles
- dépenses d'investissement
- dépenses de l'état OR dépenses publiques OR dépenses de fonctionnement et d'intervention de l'état OR dépense publique
- dépôts des particuliers
- dérives jurisprudentielles
- désamiantage
- désendettement
- désengorger les urgences
- déserts médicaux
- déserts universitaires
- détachement des travailleurs
- détourné d'argent public OR détournement d'argent public
- dette OR dettes
- dette des états OR dette des états OR dette publique
- développement durable
- développement économique
- devenir entrepreneurs
- dialogue social
- dictatures OR dictature
- diesel
- dieu
- diffusion de l'art
- dignité
- diplomatie
- directive européenne
- directives de libéralisation
- discrimination
- discrimination positive
- discriminations
- dislocation
- dispositif de renseignement
- dissuasion nucléaire
- distribution d'électricité
- diversité culturelle
- dividendes
- djihadistes OR djihad
- dmto
- doctorants
- documents officiels
- donald trump OR d.trump OR donaldtrump OR d..trump OR d trump
- donations
- données publiques
- dons
- dotations de l'état OR dotations de l'état
- double nationalité
- drapeau français
- drogue OR drogues
- droit à l'oubli
- droit d'auteur OR droits des créateurs
- droit d'être syndiqué
- droit de faire appel des décisions
- droit de mourir dans la dignité
- droit de séjour
- droit de veto
- droit de vote
- droit de vote des étrangers
- droit du sol
- droit européen
- droit français
- droits à construire
- droits civiques
- droits culturels
- droits de douane
- droits de mutation OR droit de mutation
- droits de succession
- droits humains
- drones
- dumping social
- durée du quinquennat
- e-learning
- eau OR eaux
- eaux
- échanges commerciaux
- échec scolaire
- éco-conception
- école OR écoles

- école nationale de la magistrature OR ecole nationale de la magistrature OR ENM
- écoles d'ingénieurs
- écoles de commerce
- écoles primaires OR école primaire
- écologie OR écologique OR écologiques OR ecologie
- écologistes
- économie OR économies OR economie
- économie maritime OR economiedelamer
- économie numérique
- économie réelle
- économie sociale et solidaire
- économies sur les dépenses
- économiste OR économistes
- edf
- éducatives OR éducation
- éducation artistique
- éducation nationale
- eelv
- effectifs de l'état OR effectifs de l'état
- effort de défense
- effort national
- égalité OR egalite
- égalité républicaine
- égalité salariale
- ehpad
- el khomri
- électeurs
- élection présidentielle OR élections présidentielles
- élections OR élection
- élections législatives
- élections locales
- électricité
- élevage
- élevage intensif
- élèves
- éleveur OR éleveurs
- élu
- élucidation des affaires
- élus condamnés
- élus locaux
- embaucher
- émissions de gaz à effet de serre
- emotion
- empêcher les abus
- emploi OR emplois OR embauches OR embauche
- emploi fictif
- emploi présumé fictif
- emploi stable
- emplois précaires
- employés
- employeurs
- emprunts
- ena OR école nationale d'administration
- endettement de l'état
- énergie OR énergies
- énergies marines renouvelables
- énergies renouvelables OR énergie renouvelable OR éolien OR énergie solaire
- éolien
- enfants OR enfant
- enfants récidivistes
- engrais
- enmarche
- enquêtes OR enquête
- enr
- enseignants OR enseignant
- enseignants-chercheurs
- enseignement élémentaire
- enseignement primaire
- enseignement professionnel et technologique
- enseignement supérieur
- enseignements OR enseignement OR enseigner
- ententes
- entrepreneuriat culturel
- entrepreneurs
- entreprises OR entreprise
- entreprises innovantes
- entreprises publiques
- épargne
- épargne bancaire
- épargne populaire
- epr
- équilibre budgétaire
- erasmus
- espaces maritimes
- espaces naturels protégés
- établissements hospitaliers
- établissements privés
- établissements publics
- établissements scolaires
- établissements universitaires
- etat de droit
- état fort
- état libre
- état palestinien
- état-major
- éthique OR éthiques
- étudiants
- étudiants étrangers
- euro-atlantistes
- euro-obligations
- europe OR union européenne OR ue
- exception culturelle
- exode des rapatriés
- exonérations OR exonération
- exonérations de cotisation sociale
- exonérer de charges patronales
- expérimentations animales
- expertise citoyenne
- exploitations familiales
- exportation
- exportations agricoles
- expulsions
- faceauxchrétiens
- facebook
- faciès
- falsification
- familles
- familles françaises
- fan-zones OR fan zone
- favoriser
- femme-homme OR femmes-hommes
- féodalités locales

- fermes-usines
- fessenheim
- fichés s
- fierté d'être français
- filières du travail clandestin
- filières industrielles
- filières professionnalisantes
- filières professionnelles
- fin de vie
- finance spéculative
- financement de la recherche
- financement étranger
- financement public
- financer leurs études
- finances publiques
- finances publiques locales
- financiarisation
- financiers
- fiscalité OR fiscal OR fiscales OR fiscale
- fiscalité européenne
- fiscalité locale
- fixation des tarifs
- flamanville
- florange
- flotte
- flux migratoires
- fnsea
- fonction présidentielle
- fonction publique OR fonctions publiques
- fonction publique hospitalière
- fonction publique territoriale
- fonctionnaires OR personnel de l'état
- fonds européen de défense
- fonds régionaux
- forces armées
- forêts
- formation OR formations
- formation de nos jeunes
- formation des fonctionnaires
- formation initiale
- formations politiques
- formations professionnelles OR formation professionnelle
- fossiles
- frais bancaires
- françaisfrrique
- france du désastre
- france métropolitaine
- franchises OR franchise
- françois asselineau
- françois hollande
- francophonie
- frapor
- fraternité
- fraude OR frauduleux OR fraudes
- fraude aux prestations sociales
- fraude fiscale
- fraude sociale
- fraudes commerciales
- frenchtech
- frères musulmans
- frontière OR frontières
- frontières extérieures
- frustration
- fuite des cerveaux
- g20
- gafa
- gains de productivité
- garde à vue
- garde d'enfants
- garde nationale
- garde-frontières
- gauche irréconciliable OR gauches irréconciliables
- gaulois OR gauloi
- gaz
- gaz à effet de serre
- gaz de schiste
- gel des traitements
- gel des transferts
- gendarmes OR gendarmerie
- général de gaulle
- génération
- gestion planifiée
- google
- gouvernement algérien
- gpa
- grand emprunt
- grand patronat
- grand remplacement
- grande bretagne OR grande-bretagne
- grande distribution OR grandes surfaces OR grands distributeurs
- grande guerre
- grande réforme fiscale
- grandes écoles
- grandes entreprises OR grands groupes OR grosses entreprises
- grèce
- gros patrimoines
- gros travaux
- guadeloupe
- guerre civile
- guerre mondiale
- guerres
- guyane
- gwada
- hacker
- hadopi
- harcèlement
- harki OR harkis
- harmonisation fiscale
- hébergement d'urgence
- hébergement en établissement
- héritage
- heures supplémentaires
- histoire nationale
- hlm
- hommage aux victimes
- hôpital OR hôpitaux hospital OR hopitaux
- hôpital public
- hôpitaux de proximité
- hospitalisation
- identité
- identité de genre
- identité nationale
- imams
- immigration de masse
- immigration illégale OR migrants clandestins OR immigration clandestine OR immigrés clandestins
- immigration légale
- immobilier

- importation OR importations
- impôt OR impôts OR imposition
- impôt minimal
- impôt sur la fortune
- impôt sur le revenu
- impôt sur les sociétés
- impôts locaux
- incivilités
- incubateurs
- indemnisation
- indemnités des élus locaux
- indemnités journalières
- indépendance de la justice
- indépendance stratégique
- industrie OR industriels
- industries stratégiques
- inéligibilité
- infirmières
- inflation
- ingénieurs OR ingénieur
- ingénierie
- injustices OR injustice
- innovation OR innovations
- insertion des jeunes
- insertion professionnelle
- inspection générale de l'état
- inspections générales interministérielles
- insultes
- intégration des étrangers
- intercommunalités OR intercommunalité
- interdiction OR interdire
- intérêt général
- intérêts économique OR intérêts économiques
- intérêts nationaux OR intérêt national
- intermittent du spectacle
- internet
- intrants
- investissement privé
- investissement public
- investissements d'avenir
- investissements étrangers
- iran
- irlande
- isf
- islam OR l'islam
- islamistes OR islamiste OR islamique OR islamiques
- isolation de l'habitat
- israël
- istanbul
- ivg
- jaures OR jaurès
- jdd
- jean-luc mélenchon
- jeune en alternance
- jeunesse OR jeunes
- journalistes
- juges OR juge
- jungle OR jungles
- juppé
- jury de cour d'assise
- jurys citoyens
- justice
- justice des mineurs
- justice fiscale
- justice sociale
- kinésithérapeutes
- égalité femmes-hommes OR égalité femmes/hommes
- etat islamique OR état islamique OR ei
- les étrangers
- excellence environnementale
- extrême-droite OR extrême droite
- identité heureuse
- immigration
- islam politique
- facho OR fachos OR fachisme OR fachistes OR fachiste
- l'union européenne
- la poste
- la recherche OR chercheurs
- la république
- la russie
- labels de qualité
- lagarde
- laïcité
- langue française
- langues régionales
- latin
- laxisme
- laxisme judiciaire
- le pen
- légitime défense
- les collectivités
- les étrangers
- les jeunes OR jeunes OR la jeunesse
- libéralisation des services
- libéralisme économique
- libération conditionnelle
- liberonsnosconvictions
- liberté d'expression
- liberté religieuse
- liberté syndicale
- librairies indépendantes
- libre circulation
- libre circulation des personnes
- libre-échange
- libye
- licence globale
- licences professionnelles
- licenciement OR licenciements
- licenciements collectifs OR licenciement collectif
- lieux de culte
- litiges
- livret a
- livret d'épargne industrie
- locataires
- logements OR logement
- logements sociaux OR logement social
- loi d'orientation
- loi de 1905
- loi de moralisation de la vie publique
- loi el khomri
- loi fioraso
- loi hadopi
- loi lru
- loi religieuse
- loi sapin
- loi sru
- loi sur la protection des sources

- loi sur le renforcement de la démocratie et des libertés locales
- loitaubira OR loi taubira
- loitravail OR loi travail
- loyers
- lunettes
- lutte antiterroriste
- lutte contre l'évasion fiscale
- lutte contre l'immigration clandestine
- lutte contre l'optimisation fiscale
- lutte contre la contrefaçon
- lutte contre la discrimination
- lutte contre la fraude sociale
- lutte contre la précarité OR lutterai contre la précarité
- lutte contre les discriminations
- lutte contre les fraudes
- lycéens
- lycées OR lycée
- lycées professionnels
- machines OR machine
- madeinfrance
- magistrats OR magistrat OR magistrature
- maintien à domicile
- maisons de santé
- maisons de soins
- maîtrise de la langue française
- maîtrise de la masse salariale
- majorité silencieuse
- malades OR malade
- maladies OR maladie
- maladies graves ou contagieuses
- malentendants
- malvoyants
- mandat
- manifpolice
- marché commun
- marché unique européen
- marchés publics
- marges des producteurs
- mariage
- marine nationale
- maritime
- marseillaise
- massacre
- maternelle
- mdph
- mécénat populaire
- médecine
- médecins
- médecins généralistes
- medef
- médias
- médicaments
- méditerranée
- membres du parquet
- menace OR menaces
- menace terroriste OR menace terroristes
- ménages
- mensonges OR mensonge
- méritocratie
- mesures emblématiques
- mesures fiscales
- méthodes pédagogiques
- métiers de l'entrepreneuriat
- métiers de la finance
- métiers du droit
- microentreprise
- mieux rémunérer
- migrants
- migratoires
- militaires
- millefeuille administratif
- mineurs
- minimum vieillesse
- ministère de l'intérieur, de l'immigration et de la laïcité
- ministère des souverainetés
- ministres du culte
- mis en examen
- mixité sociale
- mlp_officiel
- mobilisation OR mobilise
- mobilité professionnelle
- modèle républicain
- modem
- modernisation
- moins de 18 ans
- monde carcéral
- monde judiciaire
- monde rural
- mondialisation
- monnaie
- monnaie nationale
- montauban
- montée de la délinquance
- moraliser
- moratoire
- mosquées OR mosquée
- moyen-orient OR moyen orient
- moyenne européenne
- moyens de nos armées
- multiculturalisme
- multilatéralisme
- multinationales
- musique OR musiques
- musulman OR musulmans OR musulmanes
- mutation digitale
- mutuelles
- nation OR nations
- nationalisation OR nationaliser
- nationalité
- nationalité française
- naturalisations OR naturalisation
- négociation salariale
- neutralité
- niceattentat OR attentat de nice
- niches fiscales OR niche fiscale
- nicolassarkozy
- non remplacement de un fonctionnaire sur deux OR non-remplacement d'un fonctionnaire sur deux
- non-cumul des mandats
- norme nationale
- normes OR norme
- normes environnementales
- normes européennes
- nos soldats

- notation sociale
- notre sécurité
- notredémocratie
- nouvelle calédonie OR nouvelle-calédonie
- nouvelles technologies
- nucléaire OR industrie nucléaire OR filière nucléaire OR part du nucléaire
- nuitdebout
- nullité de procédure
- numérique OR numériques OR numérisation
- numerus clausus
- obligation pour eux d'accepter un emploi
- obligations légales
- officines
- offre de logements
- offres en ligne
- ogm
- omc
- onema
- ong
- onu OR nations unies
- opa
- open data OR open-data
- opérateurs téléphoniques
- opérations extérieures OR opex
- opérations spéculatives
- opinion publique
- optimisation fiscale
- ordonnances
- organisations syndicales
- orthophonistes
- otan OR l'otan
- outil de défense
- outre-mer OR outre-mers
- ouvriers
- pac OR politique agricole commune
- pacs
- pacte de confiance
- pacte de responsabilité
- pacte de stabilité
- paix
- palestine
- paradis fiscaux
- parcours de soins
- parcours européens
- parcours professionnels
- parcours scolaire
- parents
- parisiens OR parisien
- parité
- parlement
- parlement européen
- parlementaire
- partenaires européens
- partenaires sociaux
- partenariats public-privé
- patrimoine
- patriotisme OR patriotes
- patriotisme économique
- patrons OR patron
- pays d'afrique
- pays d'origine
- pays de l'union OR état membre
- paysans OR paysanne OR paysannes
- pêche
- pêche illégale
- peines OR peine
- bi-nationalité OR binationalité OR binationaux
- déchéance de nationalité
- peines de prison OR peines d'emprisonnement
- peines de réclusion criminelle
- peines planchers
- penelope fillon OR fillongate
- pénibilité
- pensions OR pension
- performance énergétique
- période de sûreté
- periscope
- perpétuité
- personnels soignant OR personnel soignant OR personnels soignants
- personnes âgées
- personnes de plus de 55 ans
- personnes handicapées OR handicap OR handicaps OR handicapés OR personnes handicapés
- perte d'autonomie
- perturbateurs endocriniens
- pesticides
- petite enfance
- petites retraite OR petites retraites
- pétrodollars
- peuple
- peurs OR peur
- pib
- pisa
- places de prison
- places en crèche
- plan carcéral
- plan d'investissement
- plan de rénovation
- planète
- planification
- planification écologique
- planification stratégique
- plein emploi OR plein-emploi
- plus de 45 ans
- plus-value des terrains à bâtir
- plus-values OR plus-value
- pma
- pme
- pme européennes
- pmi
- pnru
- pôle emploi
- pôles de santé de proximité
- police OR policiers
- politique agricole commune
- politique commerciale OR politiques commerciales
- politique d'application des peines
- politique de coopération
- politique de défense
- politique de la ville
- politique étrangère
- politique industrielle de défense
- politique nataliste
- politiques culturelles
- politiques publiques

- pollution atmosphérique
- pollutions OR pollution OR polluants OR polluant
- polynésie
- pompiers
- porte-avions
- position dominante
- postes supplémentaires
- poursuites disciplinaires
- poutine OR putin
- pouvoir d'achat
- pré-apprentissage
- précarité
- précarité énergétique
- préfet
- prélèvement à la source
- prélèvements indexés sur le revenu
- prélèvements obligatoires
- premier ministre
- prescription à 10 ans
- prescriptions OR prescription
- présomption d'innocence
- prestations sociales
- prêts aidés
- prévention
- primaire à droite
- primaires citoyennes OR primairescitoyennes
- prime à la détention longue de terrains constructibles
- prime annuelle
- prime pour l'emploi
- primes OR prime
- principe de précaution
- priorité d'accès
- priorité nationale
- prison OR prisons
- prison ferme
- prison supplémentaires
- privatisation OR privatiser OR privatisations
- privilège
- prix des médicaments
- prix planchers
- procédures d'expulsion
- procès
- processus judiciaire
- producteurs
- production OR productions
- production agricole
- production d'électricité
- produit intérieur
- produits agricoles
- produits agricoles et alimentaires
- produits de santé
- produits dérivés spéculatifs
- produits financiers
- professionnels de santé
- professions judiciaires
- professions libérales
- programmes jeunesse
- progrès social
- progressistes OR progressiste
- proportionnelle
- propriétaires
- protection animale
- protection du territoire national
- protection sociale
- protectionnisme
- prothèses auditives
- ptz
- publicité
- puissance européenne
- puissances d'argent
- pupilles de la nation
- qatar
- qualité alimentaire
- quartiers OR quartier
- quartiers populaires
- quartiers prioritaires
- quinoa
- quotas OR quota
- quotas d'importation
- quotient familial
- racines chrétiennes
- racisme OR racism OR raciste
- anti-sémitisme OR antisémitisme OR antisémites OR antisémitisme
- xenophobe OR xénophobe
- racisme anti-français
- radicalisation
- rationalisation
- reinformation OR réinformation OR ré-information
- réadmission
- réarmement
- recettes
- recettes publiques
- recherche agronomique
- recherche publique
- récidivistes
- reconduites à la frontière
- reconnaissance
- recours aux génériques
- rectorats
- redistributif citoyen
- redressement OR redresser
- réduction
- réduction de la déductibilité des dépenses de gros travaux en revenus fonciers
- réduction de la dépense
- réduction des déficits
- réduction du taux d'endettement public OR réduction de la dette publique
- réduction du temps de travail
- réduire de moitié l'immigration
- réduire le nombre de parlementaires
- réduire les surcoûts
- référendum
- réforme
- réforme de la dépendance
- réforme de la fiscalité
- réfugiés
- régalien
- régime dérogatoire
- régime fiscal
- régime général
- régime social des indépendants
- régions OR région
- réglementation européenne

- règles du détachement
- regroupement familial
- régularisations
- régulariser des clandestins
- réhabilitation des logements
- réindustrialisation OR réindustrialisation OR ré-industrialisation
- religion OR religions OR religieux
- relocalisation
- rembourser OR remboursements OR remboursable OR remboursement
- remises automatiques de peine
- remises de peine
- remplacement OR remplacer
- rémunération du président de la république
- rémunération forfaitaire
- rémunérations OR rémunération
- renégociation
- renouvelables OR renouvelable
- renouvellement urbain
- rénovation
- rénovation thermique
- rénovation urbaine
- renseignement intérieurs
- rentrées fiscales OR recettes fiscales
- représentants de l'état
- représentants des salariés
- reprise d'entreprises
- république
- république française OR république françaises
- réseaux sociaux
- réserve parlementaire
- réservée aux migrants clandestins
- réservistes
- résidences principales
- résistance OR résistances
- responsabilité pénale
- responsable politiques OR responsable politique OR responsables politiques
- ressources naturelles
- restaurants d'entreprise
- restauration collective OR restaurations collectives
- restauration scolaire
- rétention administrative
- retenue à la source
- retraite OR retraités OR retraites OR retrait
- retraite solidaire
- revalorisation OR revaloriser
- revenir à l'équilibre de nos comptes publics
- revenu universel
- revenus de l'assistance
- revenus du capital
- revenus modestes
- révisions constitutionnelles
- révolution citoyenne
- révolution fiscale
- révolution numérique
- rgpp OR révision générale des politiques publiques
- richesse nationale
- robots OR robotisation
- route migratoire
- royaume-uni
- rsa
- rsi
- ruralité
- russes OR russe
- russie
- rythmes scolaires
- sacrifices OR sacrifice
- salafisme OR salafiste OR salafistes
- salaire des fonctionnaires OR point d'indice
- salaire horaire
- salaire minimum
- salaire moyen
- salaires OR salaire
- salarié OR salariés OR salariées OR salarié.e.s
- sanctions
- santé
- santé publique
- savoirs fondamentaux
- schengen
- sciences OR science
- scolarité
- scolarité obligatoire
- scop
- secret de l'instruction
- secteur amateur
- secteur bancaire
- secteur privé
- secteur public
- secteurs régaliens
- secteurs stratégiques
- sécurité
- sécurité collective
- sécurité des français
- sécurité intérieure
- sécurité nationale
- sécurité sociale
- seniors OR senior OR séniors
- sens de l'état
- séparation des pouvoirs
- service civique
- service militaire
- service national
- service public OR services publics
- service sanitaire
- services de l'état
- services innovants
- services publics marchands
- simplification OR simplifier
- situation irrégulière
- small business act
- smart nation
- smic
- smicards
- sncf
- snowden
- société civile
- société du travail
- soins d'urgence
- soins dispensés
- solaire
- soldats
- solidarité OR solidarités OR solidaire OR solidaires
- sols

- solutions de garde
- sommetelevage
- sondage OR sondages
- sortir du nucléaire
- souffrance au travail
- souffrances OR souffrance
- sources d'énergie
- sous-traitants
- souveraineté
- souverainetés démocratiques
- souverainetés nationales OR souveraineté nationale
- spectacle vivant
- spéculation OR spéculateurs
- sport
- sportifs amateurs
- stages OR stage
- start-ups OR startups
- statut pénal du chef de l'état
- stock options OR stock-options
- stratégie de défense
- structures parapubliques
- subprimes
- subvention publique
- successions
- suicides OR suicide
- suppression
- suppression des cotisations minimales
- suppressions de postes
- supprimant OR supprimer
- surendettement
- sûreté
- surplus
- surveillance
- suspension des allocations
- sylvicultures
- syndicat OR syndicats OR syndicaliste OR syndical OR syndicalistes OR syndicalisation
- système de santé
- système scolaire
- système universel
- tafta
- tarification progressive
- tarifs réglementés
- taux d'élucidation
- taux d'imposition
- taux d'usure
- taux de chômage
- taxe OR taxes OR taxation OR taxer OR taxerai
- taxe d'habitation
- taxe foncière
- taxe mondiale sur les transactions financières
- technologies
- technologies numériques
- télé
- téléphones portables
- temps de travail
- temps partiel
- terrains constructibles
- territoire national
- territoires OR territoire
- territoires ruraux
- terrorisme
- terrorisme islamiste
- tgv
- thales
- théorie du complot
- tiers payant
- tisa
- titres de séjour
- tolérance zéro
- totalitarisme islamique
- tourisme
- toxiques
- tpe OR très petites entreprises
- traçabilité
- traité franco-allemand
- traités
- traités européens OR traité européen
- traités internationaux
- traités transatlantiques OR traité transatlantique
- tranche supplémentaire
- transactions financières
- mariage pour tous OR mariage gay
- transactions immobilières
- manif pour tous OR manifpourtous
- transformation économique
- transition agricole
- transition écologique
- transition énergétique
- transition numérique
- transparence
- transports OR transport
- transports régionaux
- transports sanitaires
- travail
- travail clandestin
- travail détaché OR travailleur détaché
- travail obligatoires
- travailleurs
- travailleurs handicapés
- travailleurs indépendants
- très haut débit
- trésor public
- tribunal OR tribune
- tribunal administratif
- tribunal de première instance
- tribunaux correctionnels
- tribunaux pénaux
- tribunaux pénaux de proximité
- tunisie
- turquie
- tutorat
- tva
- tva à 5,5
- tva anti-délocalisations
- uber
- ultra-marins
- souverainiste
- une france puissante
- une honte
- une nouvelle société
- une république forte
- une société du travail
- uniforme à l'école
- universités OR université
- urgences sanitaires
- urssaf
- usines
- vacances
- valeurs républicaines OR valeurs de la république

- valls OR vall
- ve république
- véhicule neuf
- vente à l'étranger
- victime OR victimes
- vidéo-surveillance
- vie politique
- vieux véhicules
- village OR villages
- violencesfaitesauxfemmes OR violences faites aux femmes
- vivre-ensemble OR vivre ensemble
- voile
- voitures OR voiture
- vote utile
- votées OR votés
- washington
- web
- zone euro
- zone inondable
- zone rurale
- zones de sécurité prioritaires
- zones désertifiées
- zones prioritaires OR zone prioritaire
- zones sans réseau

10 Analyse des retweets tardifs

Plusieurs auteurs [?, ?] ont émis l'hypothèse que les retweets pourraient être négatif, en particuliers lorsqu'ils interviennent longtemps après le tweet original.

Nous avons étudié le temps entre les retweets et le tweet original pour les comptes que nous suivons ⁷. Nous considérons qu'un retweet est tardif lorsqu'il est émis plus de 14 jours après le tweet original.

Nous avons analysé les tweets originaux ayant au moins 10% de retweets tardif et nous avons gardé les 100 tweets ayant le plus de retweets (109 retweets minimum). La liste ci-dessous montre pour ces 100 tweets le nom de l'auteur, le nombre de retweets, le nombre de retweets tardifs, le texte du tweet ⁸, la date du tweet original, la date médiane des retweets en retard et enfin l'url du tweet. Les tweets que nous avons catégorisés avec certitude comme ayant reçu des retweets négatifs sont indiqués par un point rouge. La liste est triée par proportion décroissante de retweets tardifs.

De ces 100 tweets, seul 30% ont reçu des retweets négatifs ou ironique. Cette proportion de retweets négatifs élevés corrobore le lien entre la négativité et le temps de retweet. Cependant, ces derniers semblent être limités aux cas extrêmes de tweets ayant beaucoup de retweets tardifs.

- ● **@Fillon2017_fr** 2848 RT (1): *"Moi, si j'étais mis en examen, je ne serais pas candidat à l'élection présidentielle. Question de morale."* @FrancoisFillon #BourdinDirect - **29/08/2016**
01/03/2017
https://twitter.com/Fillon2017_fr/status/770150054480998400
- ● **@bayrou** 5864 RT (1): *Emmanuel #Macron est le principal responsable de la politique économique de François #Hollande depuis quatre ans. Pour quel résultat ? #QDI* - **21/09/2016**
25/03/2017
<https://twitter.com/bayrou/status/778677784184971265>
- ● **@bayrou** 297 RT (1): *La première mesure annoncée par E. #Macron est de proposer que les jeunes travaillent plus pour gagner moins. #le79inter* - **07/12/2016** 22/02/2017
<https://twitter.com/bayrou/status/806405379605274624>
- ● **@bayrou** 1252 RT (1): *Je ne me reconnais pas ds ce qu'E. #Macron incarne. Son projet de société est proche de celui défendu par N. #Sarkozy en 2007.* #QuestionsPol - **11/09/2016**
22/02/2017
<https://twitter.com/bayrou/status/774926260266999808>
- ● **@bayrou** 116 RT (0.99): *À propos d'Emmanuel #Macron : Il ne suffit pas de dire qu'on est ni de droite ni de gauche pour être centriste #E1soir* - **22/09/2016**
22/02/2017
- ● **@BrunoLeRoux** 214 RT (0.97): *Il n'y a pas de complot, François #Fillon doit s'expliquer devant la justice #LaMatinaleInfo #cnews*
<https://t.co/SVdDDqT2OU> - **02/03/2017**
21/03/2017
<https://twitter.com/BrunoLeRoux/status/837201126332194816>
- ● **@Fillon2017_fr** 686 RT (0.96): *"Si j'étais mis en examen je ne me présenterais pas à l'élection présidentielle. C'est évident."* @FrancoisFillon #8h30Aphatie - **02/09/2016**
01/03/2017
https://twitter.com/Fillon2017_fr/status/771600966655234048
- ● **@bayrou** 2782 RT (0.96): *Derrière Emmanuel #Macron il y a des gds intérêts financiers incompatibles ac l'impartialité exigée par la fonction politique #BourdinDirect* - **07/09/2016**
21/04/2017
<https://twitter.com/bayrou/status/773413492963573760>
- ● **@JeanLucRomero** 1046 RT (0.93): *Malgré les gesticulations de la #ManifPourTous,65% des Français pour le maintien de la loi sur le #mariagepourtous L'amour triomphe tjrs!* - **14/09/2016**
12/12/2016
<https://twitter.com/JeanLucRomero/status/776005126901424128>
- ● **@FrancoisFillon** 254 RT (0.93): *Pas d'autorité de l'Etat sans honnêteté ! Le*

⁷Nous limitons l'étude à ces comptes, car ce sont les seuls pour lesquels nous collectons les retweets les concernant, voir **Dispositif de collecte des données** sur la collecte de données.

⁸Certaines émoticônes ont dû être supprimées du texte

- pouvoir doit être exemplaire pour être vraiment légitime. #Fillon2017 - 21/09/2016 01/02/2017*
<https://twitter.com/FrancoisFillon/status/778684351013789696>
- **@benoithamon** 266 RT (0.92): *Triste pour @Cristiano Ronaldo et le Portugal. Énorme joueur. #PORFRA - 10/07/2016 22/01/2017*
<https://twitter.com/benoithamon/status/752222252909924353>
 - **@jeanlassalle** 384 RT (0.87): *Et si la France se remettait à marcher sur l'eau ? #Presidentielle2017 #Résistons*
<https://t.co/6oKyVtBLzg> - **30/12/2016 07/04/2017**
<https://twitter.com/jeanlassalle/status/814781999625936896>
 - **@ramayade** 207 RT (0.85): *Pour que les femmes ne soient pas représentées que par les extrêmes (Artaud, Le Pen), @ramayade, une candidature nécessaire*
#7novembre16h34 - 07/11/2016 29/01/2017
<https://twitter.com/ramayade/status/795654465374195713>
 - **@VigierPhilippe** 327 RT (0.83): *#Macron en marche aux frais des contribuables. Il ne sert pas la Fr. , il se sert. Nous allons saisir la Haute Auto. . .*
<https://t.co/auOhcikwLV> - **24/01/2017 07/03/2017**
<https://twitter.com/VigierPhilippe/status/823851682626990080>
 - **@manuelvalls** 419 RT (0.8): *La primaire est un formidable moyen pour recréer l'unité. #CeQuiNousRassemble - 05/12/2016 29/03/2017*
<https://twitter.com/manuelvalls/status/805830310155972608>
 - **@bayrou** 273 RT (0.78): *Je ne sais pas - à cette date - qui est E. #Macron, quels sont son projet et son positionnement, avec qui il veut gouverner. #BourdinDirect - 01/02/2017 22/02/2017*
<https://twitter.com/bayrou/status/826699689038905344>
 - **@FrancoisFillon** 460 RT (0.75): *Pour gouverner un pays, ma conviction est qu'il faut être irréprochable. Je pose le principe de l'exemplarité du Président et des ministres. - 18/09/2016 04/03/2017*
<https://twitter.com/FrancoisFillon/status/777431613923885056>
 - **@BrunoLeMaire** 625 RT (0.75): *Que les choses soient claires une bonne fois pour toutes : je ne me rallierai à AUCUN candidat au 2nd tour et je ga. . .*
<https://t.co/yihPCmneMN> - **02/10/2016 20/11/2016**
<https://twitter.com/BrunoLeMaire/status/782567180063813634>
 - **@bayrou** 532 RT (0.73): *J'ai lu avec beaucoup de soin le discours d'EmmanuelMacron mais je n'y ai pas trouvé beaucoup de substance. #LeGrandJury - 05/02/2017 22/02/2017*
<https://twitter.com/bayrou/status/828208897261445121>
 - **@Bernard_Monot** 193 RT (0.72): *L'islamisme, ce cancer qui détruit tout. #Fillon, #Valls, #Sarkozy ne l'ont pas arrêté. @MLP-officiel le fera !. . .*
<https://t.co/iCvsxdKUKj> - **04/01/2017 21/03/2017**
https://twitter.com/Bernard_Monot/status/816686129118375936
 - **@bayrou** 213 RT (0.72): *Je ne suis pas pour que le pouvoir de l'argent prenne le pas sur la politique.#Macron #BourdinDirect @BFMTV @RMCinfo*
<https://t.co/DittqSBegJ> - **07/09/2016 23/02/2017**
<https://twitter.com/bayrou/status/773515676711399425>
 - **@Marion_M_Le_Pen** 3706 RT (0.68): *En avant vers la victoire ! #AssisesMLP*
<https://t.co/4SYKqNoI01> - **04/02/2017 23/04/2017**
https://twitter.com/Marion_M_Le_Pen/status/827991739327713280
 - **@EmmanuelMacron** 17669 RT (0.65): *I have a message for you guys. #ScienceMarch*
<https://t.co/ZnkFIksdx> - **10/02/2017 08/05/2017**
<https://twitter.com/EmmanuelMacron/status/829849353867128832>
 - **@benoithamon** 660 RT (0.63): *Oui pour le moral j'écoute Still - Dr Dre*
<https://t.co/oTSNCazlkJ> #Punchline #Hamon2017 - **23/10/2016 22/01/2017**
<https://twitter.com/benoithamon/status/790145245262544896>
 - **@MAZUEL_2017** 128 RT (0.62): *@KingSalman Sir, may Your Majesty pardon young #AliMohammedAlNimr. Philippe Mazuel, candidate to French Presidency*
<https://t.co/RP3xShqLiB> - **10/09/2016 26/11/2016**
https://twitter.com/MAZUEL_2017/status/774722675361087489
 - **@lepenjm** 496 RT (0.62): *Ni Droite, Ni Gauche : #Macron ! Droits d'auteur ! Mon pronostic : 3 à 4%. - 16/11/2016 07/05/2017*
<https://twitter.com/lepenjm/status/798846004556546049>
 - **@GLarrive** 112 RT (0.62): *La France n'a besoin, ni de la brutalité extrémiste de Marine Le Pen, ni de l'irréalité bougiste d'Emmanuel Macron. . .*
<https://t.co/NlPgZ1YSUB> - **23/02/2017 10/04/2017**

- <https://twitter.com/GLarrive/status/834761651383566337>
- **@LydiaGuirous** 411 RT (0.61):
Retrouvez dans @Valeurs ma tribune face à l'islamisme, 2017 sera "Le quinquennat de la dernière chance"
<https://t.co/z1Au8CEkz9> - **09/12/2016**
08/03/2017
<https://twitter.com/LydiaGuirous/status/807204636641861632>
 - **@FrancoisFillon** 708 RT (0.61):
Beaucoup de Français ont l'impression de travailler pour ceux qui ne travaillent pas. #PrimaireLeDebat - 17/11/2016
01/02/2017
<https://twitter.com/FrancoisFillon/status/799368039552708609>
 - **@Simonnet2** 377 RT (0.58): *En vidéo l'arnaque de l'ubérisation, paupérisation organisée par des multinationales hors-la-loi ! La vidéo complet...*
<https://t.co/sEKuPuXxr4> - **23/12/2016**
04/03/2017
<https://twitter.com/Simonnet2/status/812410976989564929>
 - **@ThomasScuderi** 110 RT (0.57):
#JouerCollectif pour éviter les extrêmes, tel est le défi #citoyen qui réinventera la #Politique ! Merci...
<https://t.co/MYeodTN1OH> - **13/11/2016**
20/12/2016
<https://twitter.com/ThomasScuderi/status/797771376094564352>
 - **@marseille** 149 RT (0.57): *Suivez toute l'actualité de Marseille sur Twitter -*
20/01/2017 **09/02/2017**
<https://twitter.com/marseille/status/822396311722807296>
 - **@DamienObrador** 113 RT (0.57): *La #Patriosphere mobilisée pour la victoire de @MLP_officiel en 2017 ! #AvecMarine*
<https://t.co/wmpzFFa35g> - **16/10/2016**
16/04/2017
<https://twitter.com/DamienObrador/status/787727421898391553>
 - **@OuchikhKarim** 436 RT (0.54): *La République est laïque, mais la France est chrétienne. Une vérité qui s'impose à tous, y compris en politique*
<https://t.co/jHpI0u6Izu> - **04/09/2016**
05/11/2016
<https://twitter.com/OuchikhKarim/status/772411000637317121>
 - **@JLMelenchon** 349 RT (0.54):
@LiemHoangNgoc : Nous allons revaloriser les petites retraites et augmenter le minimum vieillesse. #JLMChiffre...
<https://t.co/HS2gwbHopW> - **19/02/2017**
16/03/2017
<https://twitter.com/JLMelenchon/status/833320688048099328>
 - **@FrancoisFillon** 372 RT (0.54): *Il faut une justice rapide et ferme.*
 - L'impunité zéro doit être la règle !*
#PalaisDesCongrès
<https://t.co/puyJLiyuLj> - **18/11/2016**
01/03/2017
<https://twitter.com/FrancoisFillon/status/799690695883657216>
 - **@Simonnet2** 155 RT (0.53): *Pourquoi le Gvt n'engage pas d'enquête sur Lafarge après les révélations du Monde sur sa collaboration avec #Daech ?*
<https://t.co/q6paaQz4uP> - **06/07/2016**
13/08/2016
<https://twitter.com/Simonnet2/status/750711159737741312>
 - **@FrancoisFillon** 133 RT (0.53): *De nous, les Français attendent transparence et intégrité : pour rétablir ordre et confiance, l'exemple doit venir d'en haut. #CNLR -*
02/07/2016 **01/02/2017**
<https://twitter.com/FrancoisFillon/status/749233025184108544>
 - **@PhilippePoutou** 256 RT (0.52): *Vu sur un mur : "Pour une abstention utile, votez Poutou"* <https://t.co/UE3XKkODuh> - **24/11/2016** **04/04/2017**
<https://twitter.com/PhilippePoutou/status/801892217803722752>
 - **@EmmanuelMacron** 157 RT (0.52): *C'est un livre où j'explique la vision que j'ai du pays. #Révolution*
<https://t.co/zwyXTULNaN> - **24/11/2016**
27/12/2016
<https://twitter.com/EmmanuelMacron/status/801850774594736128>
 - **@Marion_M_Le_Pen** 663 RT (0.49): *"#Fillon est pour le droit du sol, le regroupement familial, contre la restauration des frontières, contre la prior..."* <https://t.co/uYH3mZxDtJ> - **15/12/2016** **13/03/2017**
https://twitter.com/Marion_M_Le_Pen/status/809324258312781826
 - **@PhdeVilliers** 151 RT (0.48): *Mathieu Bock-Côté : "Un certain catholicisme se mue en utopie multiculturaliste"*
<https://t.co/pg5sXONbU7> - **24/01/2017**
05/03/2017
<https://twitter.com/PhdeVilliers/status/823894913678974976>
 - **@manuelvalls** 234 RT (0.46): *Le revenu universel que je défends, c'est une grande innovation sociale : garantir un revenu à tous et à tout momen...*
<https://t.co/zfbbsZswUV> - **27/10/2016**
24/01/2017
<https://twitter.com/manuelvalls/status/791653817544826881>
 - **@lesRepublicains** 155 RT (0.46): *Trouvez le bureau de vote le plus proche de chez vous pour voter à la #Primaire2016 des 20 & 27 novembre !*
<https://t.co/Eb1SFDuU5e> - **13/10/2016**
09/11/2016

- <https://twitter.com/lesRepublicains/status/786494987080990720>
- **@OuchikhKarim** 277 RT (0.45): *Pour #2017, je confirme ma position au #Scan : "J'appelle à voter pour #MarineLePen". "Mon coeur ne balance pas" !*
<https://t.co/4Qv1vfodst> - **14/10/2016 03/11/2016**
<https://twitter.com/OuchikhKarim/status/786965898540220416>
 - **@lepenjm** 1749 RT (0.44): *Un éléphant, ça TRUMP, ça TRUMP, un éléphant ça TRUMP énormément. . . (L'éléphant est l'animal emblème du Parti Républicain) -* **20/07/2016 09/11/2016**
<https://twitter.com/lepenjm/status/755809512649547776>
 - **@alainhoupert** 503 RT (0.43): *"La seule activité continue de l'Elysée, c'est l'hostilité à Sarkozy" Jean d'Ormesson*
<https://t.co/U3HBfMrze7> - **02/10/2016 14/11/2016**
<https://twitter.com/alainhoupert/status/782483601225113600>
 - **@PhdeVilliers** 150 RT (0.41): *Alexandre Devecchio : "Derrière l'affaire Théo, les banlieues en sécession"*
<https://t.co/mPmz4ZXqre> - **14/02/2017 27/03/2017**
<https://twitter.com/PhdeVilliers/status/831510972661641216>
 - **@Paris** 118 RT (0.41): *.@RElementaires organise une collecte de tampons et serviettes pour la #JournéeInternationaleDesDroitsDesFemmes ...*
<https://t.co/Mje4mQdWg3> - **08/03/2017 25/03/2017**
<https://twitter.com/Paris/status/839509879316443137>
 - **@SIEL_Off** 146 RT (0.4): *Le #SIEL apporte son soutien à #MarineLePen aux #Presidentielle2017 et aux candidats SIEL aux législatives* <https://t.co/ELAHohyhjzY> - **26/02/2017 29/03/2017**
https://twitter.com/SIEL_Off/status/835950123855282176
 - **@FrancoisFillon** 303 RT (0.4): *Au sommet de l'Etat, l'intégrité du Président et de ses ministres doit être irréprochable car il n'y a pas d'autorité sans exemplarité.* - **18/11/2016 31/01/2017**
<https://twitter.com/FrancoisFillon/status/79969641555551232>
 - **@Paris** 552 RT (0.39): *Parc #RivesDeSeine, venez admirer Paris au fil de l'eau* <https://t.co/9gfz5Uy1ln>
<https://t.co/8e5jQFSFL1> - **22/03/2017 20/04/2017**
<https://twitter.com/Paris/status/844513445823303680>
 - **@EmmanuelMacron** 1680 RT (0.39): *Non, nous ne devons pas accepter qu'en Tchétchénie des camps pour homosexuels soient réouverts. #MacronPrésident*
<https://t.co/wriwoHdvoc> - **17/04/2017 02/05/2017**
<https://twitter.com/EmmanuelMacron/status/854021146924240896>
 - **@DLF_Officiel** 113 RT (0.39): *.@dupontaignan "Je suis un gaulliste social. Chez DLF, il n'y a pas d'arrière boutique d'extrême droite comme au #FN" #LEntretienPolitique -* **15/03/2017 29/04/2017**
https://twitter.com/DLF_Officiel/status/842096496417898503
 - **@alainjuppe** 163 RT (0.39): *Je crois à l'esprit Bordelais. Une ville fidèle à son esprit girondin, à la tempérance, à la modération. Tout ne vient pas d'en haut. -* **12/01/2017 06/02/2017**
<https://twitter.com/alainjuppe/status/819500637083881473>
 - **@FrancoisFillon** 146 RT (0.37): *Un délit : une sanction ! Voilà la règle qui doit être parfaitement claire pour les délinquants. #PalaisDesCongrès -* **18/11/2016 31/01/2017**
<https://twitter.com/FrancoisFillon/status/799690849311490048>
 - **@OuchikhKarim** 206 RT (0.36): *L'#islam est incompatible avec la #République : il doit être mis sous tutelle ! Ma tribune sur @BVoltaire*
<https://t.co/LvvLYDoelJ> - **16/11/2016 10/12/2016**
<https://twitter.com/OuchikhKarim/status/798787380496863232>
 - **@FrancoisFillon** 196 RT (0.36): *Pour être Président de la République, il faut être soi-même irréprochable. #BourdinDirect -* **18/11/2016 29/01/2017**
<https://twitter.com/FrancoisFillon/status/799519082538209280>
 - **@UPR_Asselineau** 482 RT (0.35): *Frédéric Robert, maire de Monétier-Allemont : "Non Mr @EmmanuelMacron, La commune, La France ne sont pas à vendre !..."*
<https://t.co/DmQTXqqPDN> - **26/03/2017 04/05/2017**
https://twitter.com/UPR_Asselineau/status/846108489479000064
 - **@FrancoisFillon** 249 RT (0.35): *Nous devons entendre les Français. Nous devons entendre leur demande d'honnêteté, de courage et de vérité. #CNlesRépublicains -* **14/01/2017 01/02/2017**
<https://twitter.com/FrancoisFillon/status/820233545046171648>
 - **@FrancoisFillon** 152 RT (0.35): *C'est le cri de désespoir des Français lassés de travailler pour ceux qui ne travaillent pas qui a guidé la rédaction de mon programme. -* **03/09/2016 25/01/2017**
<https://twitter.com/FrancoisFillon/status/772024442377670656>

- **@dupontaignan** 534 RT (0.35): *Retrouvez mon clip officiel de campagne pour la #présidentielle A découvrir et à partager ! #NDA2017 #DeboutLaFrance*
<https://t.co/sfffjfbkjc> - **28/03/2017**
<https://twitter.com/dupontaignan/status/846759051505651712>
- **@RafikSmati** 155 RT (0.33): *Ce que cache vraiment la candidature Macron...*
<https://t.co/RIwz4VnirU> - **15/01/2017**
<https://twitter.com/RafikSmati/status/820611357456404480>
- **@LydiaGuirous** 166 RT (0.33): *Ce n'est pas à la République de s'adapter à l'islam mais à l'islam de s'adapter à la République. #Laïcité ne se négocie pas #Valls ! @leJDD* - **31/07/2016** 26/08/2016
<https://twitter.com/LydiaGuirous/status/759687863072071680>
- **@LaurencePache** 269 RT (0.33): *Merci Mr #Vidalies, pour l'ouverture à la concurrence des #TER, pour l'augmentation des péages autoroutiers, pour l...*
<https://t.co/orfp5jUWJm> - **19/09/2016** 10/01/2017
<https://twitter.com/LaurencePache/status/777993712546811904>
- **@LGRANDGUILLAUME** 131 RT (0.32): *#taxis #VTC #loti Je serai vers 18h Gare de Lyon pour commencer la distribution pour la loi*
<https://t.co/ZOJA9GDXJe> - **23/10/2016**
<https://twitter.com/LGRANDGUILLAUME/status/790191309764038657>
- **@LellouchePierre** 109 RT (0.32): *À libre absolument dans @lemondefr, le réquisitoire d'@Anne_Hidalgo contre @EmmanuelMacron. La meilleure attaque à...*
<https://t.co/eGjmsRsa76> - **12/01/2017** 02/05/2017
<https://twitter.com/LellouchePierre/status/819611464168308739>
- **@JeanLucRomero** 1248 RT (0.32): *Tweetons tous cette campagne de prévention #VIH que fustigent tous les homophobes et bigots d'un autre temps !...*
<https://t.co/vho592qDfv> - **18/11/2016**
<https://twitter.com/JeanLucRomero/status/799681071214428160>
- **@FrancoisFillon** 130 RT (0.3): *Il n'y a pas d'autorité sans prestige, sans une certaine dignité et une intégrité certaine. #PalaisDesCongres* - **18/11/2016**
<https://twitter.com/FrancoisFillon/status/799684116941676544>
- **@benoithamon** 372 RT (0.3): *Saisissons la chance de ne pas subir le dérèglement climatique, les mutations de la société, de l'industrie, mais d'e...*
<https://t.co/omCWDiNrA> - **19/03/2017**
<https://twitter.com/benoithamon/status/843493085770334208>
- **@pierredeniziot** 399 RT (0.29): *Ce moment gênant où tu vas sur le site de #Macron et où il n'y a pas d'onglet #projet...*
<https://t.co/mHx0WJVxSn> - **21/01/2017** 05/02/2017
<https://twitter.com/pierredeniziot/status/822908822608023552>
- **@PhdeVilliers** 776 RT (0.29): *Philippe de Villiers pense à s'afficher aux côtés de Marine Le Pen*
<https://t.co/xUd3iiVfdF> - **27/03/2017** 29/04/2017
<https://twitter.com/PhdeVilliers/status/846416458594832384>
- **@LellouchePierre** 125 RT (0.29): *Je le dis avec gravité, la question de l'application de l'art. 68 de la Constitution (destitution pour manquement à ses devoirs) est posée* - **19/10/2016** 07/11/2016
<https://twitter.com/LellouchePierre/status/788754826712719361>
- **@UPR_Asselineau** 158 RT (0.27): *ÉVÉNEMENT : Grand rassemblement de l'UPR le 25 mars à Paris ! Réservez votre place sur notre site ! #Asselineau2017*
<https://t.co/hHAEPnOfyt> - **17/02/2017**
https://twitter.com/UPR_Asselineau/status/832525210729451520
- **@Marion_M_Le_Pen** 771 RT (0.27): *#LesRépublicains : "Plutôt que de vous contenter d'un plan B, tournez-vous vers le plan M, le plan Marine" #Fillon*
<https://t.co/r3rV34j7fB> - **06/02/2017**
https://twitter.com/Marion_M_Le_Pen/status/828571665315532801
- **@Fillon2017_fr** 434 RT (0.27): *"Je veux une classe politique exemplaire, avec des ministres qui ne sont pas mis en examen." @FrancoisFillon #AvecFillonElevonsLeDébat* - **24/11/2016** 01/03/2017
https://twitter.com/Fillon2017_fr/status/801882003536805890
- **@najatvb** 137 RT (0.26): *Avec ce livre, "La vie a plus d'imagination que toi", j'ai voulu parler de la France telle que je la connais, telle...*
<https://t.co/zLStE8DUCN> - **31/03/2017**
<https://twitter.com/najatvb/status/847764078609260544>
- **@JLMelenchon** 919 RT (0.26): *#Juppé et #Fillon ont parlé de l'ISF et des riches. Mais pas un mot sur les 9 millions de pauvres en France...*
<https://t.co/0pJbZHRpxr> - **24/11/2016**
<https://twitter.com/JLMelenchon/status/80104042017>

- <https://twitter.com/JLMelenchon/status/801905686296657920>
- **@DidierTauzin** 174 RT (0.26): *Exigeons un casier judiciaire vierge à nos hommes politiques ! Ils doivent être exemplaires.* <https://t.co/VrvvAui83D> - **05/11/2016 11/12/2016**
<https://twitter.com/DidierTauzin/status/794954967840854020>
 - **@valerieboyer13** 163 RT (0.25): *Porter le voile est perçu comme un signe d'allégeance à notre ennemi. Je proposerai un texte pour y mettre un terme, luttons c/ le salafisme* - **11/09/2016 22/11/2016**
<https://twitter.com/valerieboyer13/status/774908438417670144>
 - **@NicolasSarkozy** 3203 RT (0.25): *Je souhaite le meilleur pour mon pays et pour celui qui aura à conduire la France que j'aime tant - NS*
<https://t.co/xNYy61YUVV> - **20/11/2016 30/03/2017**
<https://twitter.com/NicolasSarkozy/status/800451750045970432>
 - **@MLP_officiel** 248 RT (0.25): *"L'immense majorité des homosexuels ne demandent pas le mariage gay. Ils ne réclament qu'un PACS amélioré."* **#BFMPolitique** - **16/10/2016 28/04/2017**
https://twitter.com/MLP_officiel/status/787604569593356289
 - **@MLP_officiel** 648 RT (0.24): *"On doit renvoyer les clandestins chez eux et maîtriser nos frontières nationales, sinon les camps se reconstituero..."*
<https://t.co/HEZ64kypQn> - **24/01/2017 25/03/2017**
https://twitter.com/MLP_officiel/status/823916093064548355
 - **@MLP_officiel** 611 RT (0.24): *"Nous vivons la fin d'un monde et la naissance d'un autre ! C'est le retour des États-Nations !"* **#Koblenz**
<https://t.co/IHH2wFSMz4> - **21/01/2017 20/03/2017**
https://twitter.com/MLP_officiel/status/822762303925981184
 - **@manuelvalls** 1483 RT (0.24): *Je suis candidat à la présidence de la République.* **MV** - **05/12/2016 29/01/2017**
<https://twitter.com/manuelvalls/status/805827675482951680>
 - **@LydiaGuirous** 562 RT (0.24): *Message aux "féministes" démissionnaires de gauche qui défendent #burkini voile burqa...au nom d'une fausse Liberté*
<https://t.co/ku378P1PNf> - **07/09/2016 20/10/2016**
<https://twitter.com/LydiaGuirous/status/77358642534222337>
 - **@Simonnet2** 281 RT (0.23): *Il y a pas écrit La Poste, là ! Ne laissons pas les supérettes faire main basse sur la ville, et sur le service pub...*
<https://t.co/9AJZgojfl4> - **20/11/2016 16/12/2016**
<https://twitter.com/Simonnet2/status/800321460078252033>
 - **@MLP_officiel** 447 RT (0.23): *"Les pompiers et les policiers sont immobilisés chaque jour pour régler les problèmes liés aux jungles de #migrants..."*
<https://t.co/ck0qUBXXhG> - **24/01/2017 26/03/2017**
https://twitter.com/MLP_officiel/status/823916442613612544
 - **@GilbertCollard** 223 RT (0.23): *Et vous demandez comment on peut faire des économies :* <https://t.co/5UtaUJDxLV> - **06/01/2017 22/03/2017**
<https://twitter.com/GilbertCollard/status/817302224040853504>
 - **@ECiotti** 1013 RT (0.23): *Je propose à @EliseLucet d'enquêter sur le financement de la campagne de M. #Macron, elle qui a enquêté sur M...*
<https://t.co/ABahs9ombb> - **06/02/2017 11/03/2017**
<https://twitter.com/ECiotti/status/828542977815494656>
 - **@valerieboyer13** 245 RT (0.21): *L'assassinat de #ZhangChaolin passe sous silence ... !* <https://t.co/BbGwDjK86w> - **15/08/2016 05/09/2016**
<https://twitter.com/valerieboyer13/status/765314139971526657>
 - **@valerieboyer13** 184 RT (0.21): *Jamais nous ne laisserons prendre en otage ni la démocratie ni le destin de notre pays. Soutien à @FrancoisFillon*
<https://t.co/VCbOpTfO6Q> - **02/02/2017 01/03/2017**
<https://twitter.com/valerieboyer13/status/827165016684130304>
 - **@Paris** 198 RT (0.21): *#ParisPlages revient du 20/07 au 04/09. Le programme complet*
<https://t.co/hQoJzM430L>
<https://t.co/60ZVJhVPT2> - **19/07/2016 17/08/2016**
<https://twitter.com/Paris/status/755342759968530432>
 - **@OuchikhKarim** 308 RT (0.21): *Qui peut croire un seul instant que #Macron soit capable de garantir aux Français une majorité pour gouverner !*
<https://t.co/nr4cDS3HpA> - **09/03/2017 27/03/2017**
<https://twitter.com/OuchikhKarim/status/839875214079057922>
 - **@MLP_officiel** 476 RT (0.21): *"Il n'y a rien pour nous de plus beau que la France, il n'y a rien pour nous de plus grand que la France !"*... <https://t.co/1jeyJvAeka> - **05/02/2017 21/03/2017**

https://twitter.com/MLP_officiel/status/828249040450101250

- **@GilAverous** 335 RT (0.21): *Si on me demande pourquoi je soutiens @NicolasSarkozy, la réponse est là ↓ #ToutPourLaFrance*
<https://t.co/u2lSs9ohE9> - **16/09/2016**
15/10/2016
<https://twitter.com/GilAverous/status/776765712232747008>
- **@francoisbaroin** 163 RT (0.21): *Rassemblement total autour de @FrancoisFillon pour la présidentielle. 2/2*
- **27/11/2016** **31/01/2017**
<https://twitter.com/francoisbaroin/status/802974514191728641>
- **@alainjuppe** 668 RT (0.21): *Je me sens plus proche du Pape François que de Sens Commun et de la #ManifPourTous #JuppéPrésidentCest sur @France2tv*

#20H - 21/11/2016 **16/04/2017**

<https://twitter.com/alainjuppe/status/800781567090130944>

- **@OuchikhKarim** 560 RT (0.2): *#Macron est aussi l'homme qui a vendu #Alstom aux Américains. La France bradée, c'est lui !*
<https://t.co/B3H50PsoFH>
<https://t.co/6nAy6ipJte> - **05/02/2017**
05/04/2017
<https://twitter.com/OuchikhKarim/status/828193741504651264>
- **@Marion_M_Le_Pen** 2499 RT (0.2): *"Tous les musulmans ne sont pas djihadistes, mais tous les djihadistes sont musulmans." #BBRGrandSud*
<https://t.co/Kq7TfXxIJW> - **09/07/2016**
29/10/2016
https://twitter.com/Marion_M_Le_Pen/status/751776519538745344

11 Listes des tables et des figures

List of Figures

1	Graphe des 5-communautés de la twittosphère politique pré-électorale française calculée sur la période du 1er août au 31 décembre 2016. Le paysage politique multipolaire est clairement décrit avec une disposition sur un axe gauche-droite de la plupart des partis politiques. Il faut noter que les alliances observées ultérieurement lors de la campagne politique se sont déroulées entre des personnalités politiques proches sur cette carte : Hamon-Jadot, Macron-Bayrou, Le Pen-Dupont-Aignan. La droite (<i>Les Républicains</i> , bleu) est divisée en plusieurs courants qui sont reflétés par différentes communautés. La communauté Juppé est à gauche de la droite tandis que les communautés Fillon et Sarkozy sont proches l'une de l'autre et beaucoup plus à droite. Ces deux dernières ont fusionné après les primaires de la droite tandis que la communauté Juppé a pour partie progressivement migré vers la communauté Macron. Le <i>Parti socialiste</i> était également composé d'une "gauche" (communauté Hamon, Montebourg et Pinel) et d'une "droite" (Hollande, Valls). Ce graphique a été spatialisé avec le logiciel Gephi [36] en utilisant l'algorithme ForceAtlas2 [37]. Voir File A1 sur DOI:10.7910/DVN/AOGUIA pour la version anonymisée de ce graphique.	8
2	Variation du nombre de nœuds et de liens pour les graphes de G_{T_i}.	9
3	Carte des thèmes couverts par les programmes présidentiels de 2017, établie à partir de <i>Gargantext</i>. La carte a été construite à partir des mesures politiques extraites des programmes des candidats. Les nœuds de la carte sont des labels pour des groupes de termes jugés équivalents dans le discours politique. Le lien entre un label A et un label B indique que la probabilité que A et B soient mobilisés conjointement dans la même mesure politique est élevée. <i>Gargantext</i> applique l'algorithme de Louvain pour identifier les groupes de labels qui relèvent de la même thématique et leur assigne la même couleur. Pour améliorer la lisibilité, la carte a été éditée dans le logiciel Gephi (https://gephi.org) pour définir la taille des nœuds et des labels selon une fonction monotone de leur PageRank [50]. Le Fichier A3 disponible sur DOI:10.7910/DVN/AOGUIA fournit une version éditable de cette carte (gexf).	11

4	<p>3—communautés du paysage politique français le jour du 1er tour (T=[9 avril 2017 - 23 avril 2017]). Le paysage est très morcelé et multipolaire, les principales forces politiques étant représentées (les pourcentages indiqués correspondent aux scores obtenus au premier tour) : Mélenchon et la <i>France Insoumise</i> (rouge vif - 19.58%), Hamon et le <i>Parti Socialiste</i> (rouge clair - 6.36%), Macron et <i>En Marche !</i> (rose - 24.01%), Fillon et <i>Les Républicains</i> (bleu - 20.01%), Le Pen et le <i>Front National</i> (marron - 21.30%), Dupont-Aignan et <i>Debout la France</i> (lilas - 4.7%). Pour les “petits” candidats, Asselineau (0.92%) et Arthaud (0.64%) se retrouvent dans la même communauté (violet), Poutou (1.09%) se retrouve avec Mélenchon mais n’a pas vraiment de communauté Twitter. On remarquera que non seulement Bayrou est dans la communauté Macron, ce qui est normal après sont ralliement le 23 février 2017, mais que Juppé, ex-candidat à la primaire <i>Les Républicains</i>, est beaucoup plus proche de la communauté Macron que de celle de Fillon et Sarkozy. Son “lieutenant” Edouard Philippe, deviendra premier ministre de Macron. Deux communautés non labellisées par un leader politique sont néanmoins importantes. En orange à gauche, proche de la communauté Mélenchon on retrouve une communauté identifiée comme proche de la <i>Discorde Insoumise</i>, communauté du jeu vidéo qui a joué un rôle important dans la campagne de Mélenchon. En jaune en bas à droite, proche de la communauté Le Pen, nous retrouvons une communauté de comptes populistes et nationalistes anglophones qui a fortement soutenu Le Pen. Elle est elle-même composée de sous-communautés, en particulier des supporters anglophones de Le Pen, des supporters de Trump et des nationalistes UK pro-Brexit. On relèvera également des différences importantes avec la carte pré-campagne présidentielle de 2016 (figure 1) : en plus de l’axe “gauche-droite” qui était alors dominant, un nouvel axe semble important (“haut-bas” sur la carte) qui distingue les partis à tendance nationaliste ou protectionniste des partis adhérant à la mondialisation et à la globalisation des échanges économiques (le fait que les petits candidats soient au “centre” de la carte ne doit pas être interprété comme une forme de centralité. Seules les contiguités topologiques sont significatives sur ce type de graphe. Les communautés des “petits” candidats étant relativement peu connectées aux autres se retrouvent au centre du fait de l’algorithme de spatialisation du graphe) Le Fichier A4 disponible sur DOI:10.7910/DVN/AOGUIA donne une version éditable de ce graph.</p>	16
5	<p>3—communautés du paysage politique français entre les deux tours de scrutin (T=[27 avril 2017 - 7 mai 2017]). Avec seulement deux candidats restants (Macron et Le Pen), les motifs du militantisme politique sur Twitter ont radicalement changé avec deux pôles correspondant aux partisans des deux candidats restants (issus de la recomposition des anciennes communautés) et un troisième pôle constitué par ceux qui refusaient de prendre parti. Ces derniers se positionnaient pour être la prochaine force d’opposition du futur gouvernement. Il est à noter que le parti <i>Les Républicains</i> s’est scindé dans l’entre-deux tours, les ténors de ce parti étant dispersés entre les sphères informationnelles de Macron et de Le Pen. Le Fichier A5 disponible sur DOI:10.7910/DVN/AOGUIA fournit une version éditable de ce graphique.</p>	17
6	<p>Représentation de l’évolution des communautés par un diagramme alluvial</p>	18
7	<p>Reconfigurations des communautés politiques entre juin 2016 et avril 2017. Certains grands titres de journaux ont été ajoutés pour apporter un éclairage sur les bifurcations observées. Les labels des communautés encore actives lors du second tour ont été ajoutés en fontes larges pour une meilleure visibilité. Chaque barre verticale correspond à une communauté politique sur Twitter labellisée par les comptes des candidats qui y participent. La hauteur est proportionnelle au nombre de comptes Twitter faisant partie de cette communauté. Celles-ci sont calculées tous les lundis. .</p>	19

- 8 Reconfiguration des communautés politiques au moment des primaires de la droite. Les primaires de la droite ont été rythmées par trois débats et deux votes. On peut observer les renforcements progressifs et réguliers de la communauté Fillon semaines après semaines, alors que personne ne le pronostiquait au second tour de la primaire, Juppé étant le favori des sondages. A l'issue du premier tour, de la victoire surprise de Fillon et de l'élimination également surprenante de Sarkozy (président du parti *Les Républicains* et ex-Président de la République), seule une infime partie de la communauté Sarkozy a rejoint la communauté Fillon. La communauté Sarkozy traînera des pieds pour se rallier à la communauté Fillon et Sarkozy négociera pendant plusieurs jours la place de ses membres dans le dispositif Fillon. La communauté Juppé, en revanche, prendra immédiatement ses distances par rapport aux nouvelles orientations politiques prises par le parti. Cette ligne de fracture de la droite était apparente dans le paysage politique pré-électoral (cf. figure 1). Après l'échec du retour de Juppé en mars 2017 (cf. figure 11), une partie de la communauté Juppé rejoindra Macron. Les bifurcations observées ici sont bien résumées par le titre d'un article de BFM TV du 12 décembre 2017 : *La vie après la primaire : Sarkozy recase ses proches, Juppé "abattu"*. Chaque barre verticale correspond à une communauté politique sur Twitter labellisée par les comptes des candidats qui y participent. La hauteur est proportionnelle au nombre de comptes Twitter faisant partie de cette communauté. Celles-ci sont calculées tous les lundis. 20
- 9 Reconfiguration des communautés politiques au moment où le Président de la République François Hollande annonce qu'il ne se représentera pas. Cette déclaration a eu lieu peu après que Macron ait annoncé sa candidature. Il apparaît clairement sur le graphe alluvial que la non candidature de Hollande a dynamisé la communauté de Macron au détriment de celle de Valls. Alors que la taille de la communauté Macron était restée stable au moment de l'annonce de sa candidature (427 comptes la semaine précédente, 561 la semaine de l'annonce, 472 la semaine suivante), elle a bondit à 749 comptes la semaine du renoncement de Hollande, puis 1,479 comptes la semaine suivante. Elle a dès lors été en croissance régulière jusqu'aux élections. Cette évolution est concomitante avec une perte de vitesse de la communauté Valls-Hollande, aile droite du *Parti Socialiste* la plus proche de Macron d'un point de vue idéologique. La communauté Hamon, plus à gauche, est en revanche restée relativement stable sur cette période. Chaque barre verticale correspond à une communauté politique sur Twitter labellisée par les comptes des candidats qui y participent. La hauteur est proportionnelle au nombre de comptes Twitter faisant partie de cette communauté. Celles-ci sont calculées tous les lundis. 21

10 Reconfiguration des communautés politiques au moment des primaires de la gauche. Benoît Hamon (36,03% au 1er tour et 58,69% au second tour) et Manuel Valls (31,48% au 1er tour et 41,31% au second tour) étaient les deux favoris de ces primaires et représentaient deux courants du parti socialiste qualifiés par Valls de “gauches irréconciliables” lors de sa campagne. Bien que François Hollande ne se soit pas exprimé officiellement en faveur d’un candidat particulier, Valls affirmait que celui-ci le soutenait (cf. par exemple *Le Monde* du 4 janvier 2017 “Selon Valls, Hollande le soutient”). En effet, sur notre reconstruction, la communauté Hamon est, pendant toute notre période d’observation, bien distincte de la communauté Valls. Cette dernière est quant à elle, la plupart du temps également la communauté de Hollande, notamment au moment des primaires, indiquant une proximité idéologique entre les deux personnalités politiques. Après la victoire de Hamon, une polémique éclata sur l’engagement pris par les candidats à cette primaire de se ranger derrière le vainqueur alors que certains affichaient clairement leur préférence pour Macron. Ce flottement est clairement visible sur le graphe alluvial avec des mouvements importants de militants de la communauté Valls-Hollande vers la communauté Macron dès le lendemain du deuxième tour de la primaire. Sur les 2075 comptes que comportait la communauté Valls-Hollande juste avant sa fusion avec la communauté Hamon le 13 février (Valls rejoignant également la communauté Hamon), seuls 959 comptes resteront. Des mouvements importants de militants continueront d’avoir lieu entre ces deux communautés jusqu’au premier tour de la présidentielle. Si Valls finira par soutenir officiellement Macron le 29 mars, il n’apparaissait déjà plus dans la communauté Hamon dès le 13 mars. Il apparaît dans la communauté Macron à partir du 17 avril et jusqu’au deuxième tour, indiquant un soutien actif de sa part. On remarquera également sur ce zoom le ralliement de l’écologiste Jadot à Hamon le 23 février 2017, qui se manifeste immédiatement par la fusion de sa communauté avec celle de Hamon. Chaque barre verticale correspond à une communauté politique sur Twitter labellisée par les comptes des candidats qui y participent. La hauteur est proportionnelle au nombre de comptes Twitter faisant partie de cette communauté. Celles-ci sont calculées tous les lundis.

11	Reconfiguration des communautés politiques au moment du Pénélope Gate. Le Pénélope Gate arrive à un moment où le rassemblement autour de Fillon est fragilisé en raison d’une polémique sur les investitures aux législatives, mal vécues par les sarkozystes. Une partie de ceux-ci avait déjà commencé à se détacher de la communauté <i>Les Républicains</i> dès le 23 janvier 2017 et formeront une communauté à part pendant toute la polémique. Les différentes révélations du <i>Canard Enchaîné</i> provoquent de larges mouvements de militants entre différentes communautés, les communautés autour de Juppé (potentielle alternative à la candidature Fillon ou “plan B”) et Sarkozy se renforçant progressivement. Au 1er mars, jour de la convocation de Fillon en vue d’une mise en examen, une grande partie des Sarkozystes soutiennent le retour de Juppé dans un ultime effort pour pousser Fillon à la démission, ce qui se manifeste par un flux inhabituel de comptes de la communauté Sarkozy vers la communauté Juppé. Le rassemblement du Trocadéro le 5 mars met fin aux perspectives de plan B. La communauté Fillon gagne beaucoup de nouveaux militants suite à ce coup de force mais une partie des sarkozistes et des juppéistes restent à l’écart et rejoindront respectivement, Dupont-Aignan et Macron. Sarkozy quant à lui, restera plusieurs semaines en retrait, ce qui se traduit par son absence dans la plupart des communautés Fillon entre le 16 janvier et le 10 avril. Ce n’est qu’à partir du 7 avril, qu’il soutiendra à nouveau activement la candidature Fillon, comme le mentionne par exemple le titre du <i>Parisien</i> de ce jour “Présidentielle : Sarkozy soutient (enfin) Fillon” et ce qui se manifeste par sa présence permanente dans la communauté Fillon à partir du 10 avril 2017. Comme on peut le constater, les bifurcations observées dans le Politoscope sont concomitantes et même parfois anticipent les commentaires de la presse sur l’activité des communautés politiques. Chaque barre verticale correspond à une communauté politique sur Twitter labellisée par les comptes des candidats qui y participent. La hauteur est proportionnelle au nombre de comptes Twitter faisant partie de cette communauté. Celles-ci sont calculées tous les lundis.	23
12	Reconfiguration des communautés politiques au moment du “Grand Débat” télévisé du 4 avril 2017. Philippe Poutou a créé le buzz lors de cette confrontation en dénonçant ouvertement la triche et la corruption des candidats Fillon et Le Pen et en critiquant les positions anti-système : “Quand nous on est convoqué par la police, on n’a pas d’immunité ouvrière, on y va.” Son intervention a été relayée par la plupart des médias et a recueilli plus d’un million de vues en ligne. Sa communauté Twitter, quasi inexistante jusque là et souvent fusionnée avec celle de Mélenchon, s’est développée considérablement (> 13,000 comptes) sur une courte période suite à cette intervention.	24
13	Diagramme radar de la propension normalisée à propager des fausses nouvelles (en pointillé rouge) et des réfutations (en bleu) par communauté.	34
A1	Description du dispositif de capture centré sur les leaders politiques. Celui-ci est complété par un dispositif de capture centré sur les mots-clé (Twitter track API).	52
A2	Statistique sur le nombre de retweet capturés (a) <i>en haut</i> : Évolution du taux de capture des retweets au cours du temps. Certains points bas correspondant à des pannes serveurs. (b) <i>en bas</i> : Nombre quotidien de tweets collectés.	53
A3	Description synthétique de l’infrastructure Multivac utilisée pour mettre en place le dispositif d’analyse en temps réel des tweets. Cette infrastructure mobilise des technologies telles que Elasticsearch, Apache Hadoop, RabbitMQ, MongoDB ou Redis.	55

A4	Distribution de la probabilité d’être un bot. Courbe bleue : comptes qui ont été au moins une fois dans une communauté tout en ayant au moins 3 liens dans le graphe de retweet. Courbe orange : comptes aléatoires qui ont été au moins une fois dans une communauté. Cette détection a été effectuée en juin 2016.	58
A5	Motif d’activité sur Twitter. À gauche: Nombre de retweets observés par jour et par heure entre le 1er juillet 2016 et le 23 mai 2017; à droite: Distribution cumulative inverse du temps entre un tweet et ses retweets.	60
A6	Distribution cumulative du temps passé dans les principales communautés entre juillet 2016 et le premier tour.	61
A7	Distribution lissée du temps passé dans les principales communautés entre juillet 2016 et le premier tour.	61
A8	Motifs de diffusion des tweets. (a) à gauche : Boîte à moustache de l’entropie de la diffusion de chaque tweet dans les communautés et de l’entropie de la retweets chaque jour; (b) à droite : Exemple de la diffusion privilégiée dans certaines communautés d’un tweet de François Fillon. (La carte de référence est figure 4).	63
A9	Évolution des thèmes discutés par la communauté Asselineau entre le 01 Juin 2016 et le 08 Mai 2017.	67
A10	Évolution des thèmes discutés par la communauté Fillon entre le 01 Juin 2016 et le 08 Mai 2017.	67
A11	Évolution des thèmes discutés par la communauté Hamon entre le 01 Juin 2016 et le 08 Mai 2017.	68
A12	Évolution des thèmes discutés par la communauté Le Pen entre le 01 Juin 2016 et le 08 Mai 2017.	68
A13	Évolution des thèmes discutés par la communauté Macron entre le 01 Juin 2016 et le 08 Mai 2017.	68
A14	Évolution des thèmes discutés par la communauté Mélenchon entre le 01 Juin 2016 et le 08 Mai 2017.	69
A15	Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Macron pour le thème de l’emploi.	69
A16	Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Le Pen sur le thème de l’emploi.	70
A17	Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Fillon sur le thème de l’emploi.	70
A18	Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Mélenchon sur le thème de l’emploi.	70
A19	Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Hamon sur le thème de l’emploi.	71
A20	Utilisation mensuelle des 15 mot-clés les plus spécifiques dans la communauté Asselineau sur le thème de l’emploi.	71

List of Tables

1	Correspondance entre les leaders politiques et leur parti au 6 février 2017.	12
2	Rappel pour les comptes suivis qui étaient actifs entre le 23 janvier et le 6 février 2017 (au moins 10 tweets émis).	14
3	Précision pour les comptes suivis qui étaient actifs entre le 23 janvier et le 6 février 2017 (au moins 10 tweets émis).	14
4	Rappel pour les comptes de l’échantillon aléatoire qui étaient actifs entre le 23 janvier et le 6 février 2017 (au moins 10 tweets émis).	14

5	Précision pour les comptes de l'échantillon aléatoire qui étaient actifs entre le 23 janvier et le 6 février 2017 (au moins 10 tweets émis).	14
6	Statistiques détaillées pour les communautés des 6 candidats ayant l'utilisation la plus intensive de Twitter dans leur campagne ($T_k \in [\text{Feb. 1er 2017} - \text{22 avril 2017}]$) : temps moyen passé par compte dans la communauté ($\langle \textit{jours} \rangle$), nombre total de tweets de cette communauté ($ Tweets $), nombre de comptes présents au moins une fois dans la communauté ($ comptes $) et nombre de comptes présents au moins une fois dans la communauté pondéré par l'engagement ($ comptes ponderes $). Les valeurs extrêmes sont en gras pour une meilleure lisibilité.	27
7	Statistiques détaillées pour les communautés des 6 candidats ayant l'utilisation la plus intensive de Twitter dans leur campagne ($T_k \in [\text{1er Fév. 2017} - \text{22 Avr. 2017}]$) : engagement moyen des comptes dans la communauté ($\langle \textit{engagement} \rangle$) ; quantités pondérées par l'engagement : nombre moyen de tweets par compte ($\langle \textit{activite} \rangle$), intégration politique ($\langle \textit{integration} \rangle$) et politisation ($\langle \textit{politisation} \rangle$) Les valeurs extrêmes sont en gras pour une meilleure lisibilité.	27
8	Entropie des thèmes abordés par chaque candidat et sa 3-communauté. Les deux thèmes dominants ainsi que les pourcentages qui leur sont associés sont également présentés. L'entropie maximale est de 3,46 lorsque les 11 thèmes sont discutés de manière uniforme. En ce qui concerne les thèmes retenus, Benoît Hamon est le candidat qui a abordé le plus grand nombre de thèmes, tandis que François Asselineau est le candidat qui s'est le plus concentré sur un sous-ensemble de thèmes (dont la politique étrangère et la démocratie). Les variations de l'attention accordée par les candidats à ces thèmes sont représentées sur les Figures I à N dans S1 File.	30
9	Entropie et nombre moyen de communautés [entre crochets] atteintes en moyenne par les fausses nouvelles, les réfutations et les tweets avec ou sans mention d'URL sur la période. Nous présentons également les résultats correspondant à l'agrégation de toutes les communautés en une seule à l'exception des 6 principales communautés politiques en ligne identifiées lors de l'élection présidentielle française. Les tweets sans communauté (43,44%) ont été considérés comme une seule communauté pour le traitement de l'entropie. Pour comparaison avec les données sur les réfutations et les fausses nouvelles, identifiées par les URL qu'elles diffusent, nous fournissons également les statistiques globales pour le sous-ensemble de tweets qui mentionnent des URLs, quelle qu'elles soient.	32
10	Nombre d'utilisateurs, nombre de tweets mentionnant un lien référant à une fake news et proportion de premiers relais pour les tweets contenant des liens à des "fake news". Les premiers relais sont définis comme des utilisateurs qui ont sont dans le premiers 1/30 percentile d'utilisateurs à partager une fausse nouvelle.	33
11	Nombre d'utilisateurs, nombre de tweets et proportion de premiers relais pour les tweets contenant une réfutation. Les premiers relais sont des utilisateurs qui ont été parmi les premiers 1/10 percentile d'utilisateurs à partager une réfutation.	33
A1	Présence des bot dans les communautés. Nombre moyen de comptes par période et par communauté ($\langle \textit{comptes} \rangle$) ainsi que le nombre moyen de bots détectés par période par communauté ($\langle \textit{bots} \rangle$).	58

A2 **Proportion de tweets postés par des bots dans les communautés politiques et par des bots parmi les comptes aléatoires. Par exemple, les bots dans les communautés politiques ont émis 7,2% des tweets originaux émis par tous les comptes d’une communauté politique, alors qu’ils ne représentent que 5,1% des comptes.** 59

A3 **Comptes par communauté.** Tableau indiquant, pour chaque candidat, le nombre de comptes présents au moins une fois dans sa communauté entre le 1er juillet 2016 et le 23 mai 2017. Les communautés de Le Pen et de Fillon sont clairement différentes des autres, en ce sens qu’elles sont les plus stables avec un temps moyen très élevé. Le fait que le temps moyen passé par un compte au sein d’une communauté (dernière ligne) est très élevé indique qu’une proportion significative des comptes changent d’une communauté à l’autre, sans disparaître de Twitter. 60

A4 **entropie des 3–communautés discutant chaque thème.** Les 2 communautés principales selon le $TF - IDF$ sont également présentées pour chaque thème. L’entropie maximale est de 2.585 lorsque les 6 communautés ont le même volume de tweet sur ce thème. 67

iiiiii 68c79b77b6a3062a04dfd8af17732e1c7c46cd92