Document generated on 07/13/2025 2:19 p.m.

Sociologie et sociétés

L'analyse des grands réseaux évolutifs et la sociologie pragmatique des controverses

Croiser les méthodes face aux transformations des mondes numériques

Large evolutionary network analysis and the pragmatic sociology of controversies

How to combine methods in front of the transformations of digital worlds

Francis Chateauraynaud and David Chavalarias

Volume 49, Number 2, Fall 2017

La sociologie numérique
Digital Sociology

URI: https://id.erudit.org/iderudit/1054277ar
DOI: https://doi.org/10.7202/1054277ar

See table of contents

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0038-030X (print)
1492-1375 (digital)

Explore this journal

Cite this article

Chateauraynaud, F. & Chavalarias, D. (2017). L'analyse des grands réseaux évolutifs et la sociologie pragmatique des controverses : croiser les méthodes face aux transformations des mondes numériques. *Sociologie et sociétés*, *49*(2), 137–161. https://doi.org/10.7202/1054277ar

Article abstract

Based on an unprecedented experience of bringing together two logics of research, held to be contradictory according to epistemological norms in the social sciences, this article explores how to articulate a mathematized approach to large networks, calculated from web flows, and a socio-informatics approach of controversies, evolving through non-linear critical processes. Using new digital mediations, the authors suggest some pathways for a learning interface between human interprets and algorithms. The collective cognitive device that emerges from this cross-fertilization can create a new critical approach on how social processes unfold in digital worlds, from official sites to social media.

Tous droits réservés © Les Presses de l'Université de Montréal, 2018

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

L'analyse des grands réseaux évolutifs et la sociologie pragmatique des controverses

Croiser les méthodes face aux transformations des mondes numériques

FRANCIS CHATEAURAYNAUD
EHESS — GSPR
Courriel : chateau@ehess.fr

DAVID CHAVALARIAS
CNRS — ISC
Courriel : David.chavalarias@iscpif.fr

Les discussions épistémologiques qui traversent les sciences sociales n'ont eu de cesse de réengendrer les mêmes partages et les mêmes couples d'opposition, de sorte qu'aujourd'hui encore, tout effort pour ouvrir de nouvelles voies soulève vite de vieilles controverses, dont l'objet n'est autre que la scientificité des disciplines attachées à saisir les processus sociaux. Malgré des tentatives pour formaliser les modes de raisonnement propres aux sciences sociales, la pluralité des cadres épistémiques et le rôle majeur de l'interprétation historique s'imposent comme des maximes incontournables, surtout en sociologie (Abbott, 2001a ; Berthelot, 1990 ; Bouvier, 2009 ; Passeron, 1991). Le problème viendrait de la nature interprétative de la plupart des sciences sociales, fonctionnant moins selon des normes axiomatiques qu'au travers de séries de rapprochements et de recoupements, de mises en cohérence de fragments ou de séquences, dont la signification implique une attention aiguë à leurs contextes de production et de mise en circulation. Mal comprise, l'activité interprétative est trop souvent jugée incompatible avec la modélisation, ce qui rétroagit sur les manières dont les sociologues, les historiens ou les anthropologues considèrent la plupart des formes d'instrumentation : si les outils numériques sont éligibles comme supports d'archive ou de publication, dès lors qu'ils prennent appui sur des protocoles et des procédures informatiques capables de prolonger ou d'orienter le raisonnement humain, le soupçon

Sociologie et sociétés, vol. XLIX, nº 2, automne 2017, p. 137-161

est de mise. L'instrument est soupçonné de véhiculer une conception mathématisée des processus sociaux éloignant d'une approche compréhensive des jeux d'acteurs et d'arguments, des représentations et des pratiques qui ne peuvent être saisies qu'en situation. Par exemple, dans une contribution intitulée « Épistémologie du code et imaginaire des 'SHS 2.0' », Sebastien Broca met en garde contre les « tentations impérialistes de l'épistémologie du code », auxquelles, en s'inspirant de réflexions de Bruno Bachimont, il oppose deux caractéristiques qualifiées de « non computationnelles », à savoir l'argumentation et la narration (Broca, 2016). Selon lui, l'argumentation éloigne le travail des SHS du *problem solving*, puisqu'elle pointe sur des luttes pour la représentation légitime du monde social, ce qui implique la prise en compte de la confrontation raisonnée des points de vue. Quant à la narration, elle conduit à reconnaître la multiplicité des expériences singulières, qu'il s'agit de rendre accessibles par la redescription, sans les réduire. Le champ des humanités numériques est ainsi pensé comme un champ de luttes pour l'hégémonie, opposant d'un côté des savoirs fondés sur les humanités et de l'autre des sciences computationnelles. Il est vrai qu'une forme de défiance néoluddite, plutôt en vogue, conduit à relever la propension invasive des technologies numériques, qui se manifeste à travers les innombrables « solutions informatiques » proposées aux chercheurs[1]. Or l'argumentation et la narration, ce sont précisément les ressorts à partir desquels s'est élaborée la socioinformatique des controverses, dont les travaux ont montré à quel point il ne s'agissait en aucun cas d'opérer une réduction computationnelle de la diversité et de la complexité des mondes sociaux (Chateauraynaud et Debaz, 2017). La montée en puissance des « humanités numériques » peut être lue, non comme une expansion sans limites de la « logique du code », réduite au calcul numérique, mais comme l'occasion d'ouvrir de nouveaux champs d'expériences, renouant avec le raisonnement symbolique et conceptuel, et capables de faire bouger les lignes tout en comblant le fossé qui sépare les deux grandes logiques d'enquête, trop rapidement résumées sous l'opposition entre quantitatif et qualitatif.

DEUX FAMILLES D'INSTRUMENTS POUR LES SCIENCES SOCIALES

En s'appuyant sur des expérimentations récentes de croisement de perspectives méthodologiques, nous exposons dans ce texte les linéaments d'un nouvel espace de recherche placé au cœur des humanités numériques, en proposant des médiations permettant la convergence, pour certaines classes d'objets, de deux logiques épistémiques généralement tenues pour opposées et même incompatibles. Si tout semble opposer les calculs opérés sur de grands réseaux construits à partir du web et une approche argumentative abordant les corpus sous l'angle de la sociologie pragmatique des controverses, il y a

1. La sociologie des controverses a examiné en profondeur les formes de critique radicales portées contre les sciences contemporaines. Les chercheurs sont suspectés d'accepter une formes de servitude volontaire relativement aux pouvoirs politiques et économiques, en contribuant à un « déferlement technologique » face auquel se développent des pratiques néoluddites, mêlant des stratégies de hacking à des attaques pamphlétaires, en passant par le sabotage ou l'opacification (*obfuscation*) (Jones, 2006 ; Brunton et Nissenbaum, 2015).

moyen d'explorer une troisième voie. La notion de réseau souvent considérée, avec raison, comme une bonne médiation pour développer des approches quali-quantitatives (Venturini, 2012, 2014) est-elle suffisante pour saisir les processus sociaux contemporains ? Revenons dans un premier temps sur les deux familles d'instruments, liés à des logiques d'enquête différentes, en rappelant de manière synthétique leurs propriétés marquantes. Cet article n'ayant pas pour visée de comparer terme à terme les deux familles d'instruments, nous proposons plutôt d'explorer les zones de contact et les points d'articulation, dans le but de faire advenir de nouveaux espaces de raisonnement et d'enquête dans les mondes numériques.

Une analyse fine des jeux d'acteurs et d'arguments

Les travaux socioinformatiques menés autour des logiciels Prospéro et Marlowe ont déjà une longue histoire (Chateauraynaud, 2003). Issus d'une prise de distance vis-à-vis des formes usuelles d'analyse du discours, ils incarnent d'autres manières de construire et d'interroger des corpus évolutifs. Les dispositifs initiés par la socioinformatique des controverses ont en effet placé au cœur de leurs protocoles la caractérisation des jeux d'acteurs et d'arguments, des événements reconfigurateurs, des points de rupture ou de bifurcation, appelés classiquement des *turning points* (Abbott, 2001b). Ces objets sont saisis à la fois comme des propriétés marquantes des univers textuels étudiés et comme des nœuds de significations qui importent pour les acteurs (Keller, 2013). Ces significations sont attribuées par les acteurs eux-mêmes aux processus dans lesquels ils agissent : si des énoncés engagent par exemple des segments répétés autour de Fukushima, c'est avant tout parce que la catastrophe de la centrale nippone de mars 2011 a rebattu les cartes dans les scénarisations des futurs du nucléaire civil à l'échelle mondiale, en donnant une puissance d'expression aux alertes et aux contestations de cette énergie. Sans redéployer le détail des travaux réalisés à ce jour, plusieurs grandes applications ont concerné précisément le dossier nucléaire, mais aussi les OGM ou les pesticides, et plus généralement l'évolution des alertes et des controverses environnementales (Chateauraynaud et Debaz, 2017).

Lorsqu'au début des années 2000, la socioinformatique des controverses a renoué avec l'intelligence artificielle en doublant le logiciel Prospéro (dédié à l'analyse de corpus textuels évolutifs) par le logiciel Marlowe (conçu comme un interlocuteur virtuel fonctionnant en mode dialogique), le sens et la portée de cette expérimentation ont été difficilement perçus, y compris dans l'environnement intellectuel le plus proche. Il faut dire que cette expérience était menée bien avant l'avènement des humanités numériques[2]. Après plus de 15 ans d'expérience, et des évolutions considérables dans les mondes numériques, de nouveaux chemins s'offrent à ce qui prend désormais la forme d'une « contre-intelligence artificielle ». Un des enjeux est en effet de reconnecter les

2. On ne manquera pas de noter que les « humanités numériques » sont passées, en quelques années, entre 2010 et 2014 environ, de l'état d'avant-garde critique des machines dominantes de l'internet à celui d'alignement quasi général des disciplines « littéraires » face à un « impératif numérique », dont l'énonciation a rapidement pris la forme d'une injonction managériale.

activités autonomes de Marlowe, qui s'exprime tous les jours sur son blogue, en n'étant que la partie émergée d'un réseau d'artefacts cognitifs et d'interprètes humains, avec une forme d'« écosystème numérique » fondé sur d'autres conceptions des structures de données, des algorithmes et des interfaces.

La plupart des recherches récentes ont donné lieu à des discussions méthodologiques sur le carnet de recherches *Socio-informatique et argumentation*[3]. Les questions ont principalement porté sur les différentes manières de modéliser les épreuves argumentatives dans le temps long des controverses. Une des caractéristiques fortes de Prospéro et Marlowe, comparés aux nombreux outils d'analyse de corpus, est de diriger l'attention vers les modes et modalités, les marques et les marqueurs qui, au-delà des classes lexicales habituellement étudiées (comme dans le cas du logiciel Alceste), rendent manifestes les orientations narratives ou argumentatives que donnent les auteurs-acteurs à leurs discours, et les interprétations qu'en font leurs lecteurs, soit pour les reprendre à leur compte, soit pour les contredire. Toute argumentation engendre une contre-argumentation, ce qui renvoie aux dimensions agonistiques des processus étudiés (Hayer, 1995), lesquels n'excluent pas des moments ou des séquences de coopération argumentative (van Eemeren et Garsen, 2008 ; Mercier et Sperber, 2011). Si des notions de liens et de réseaux sont présentes et installées dans les algorithmes de Prospéro, ce sont plutôt les catégories et les modes qui dominent dans la manière d'explorer les corpus, d'en extraire les propriétés marquantes et d'en fournir des représentations évolutives. Le logiciel Marlowe ne fait que prolonger les élaborations sémantiques de Prospéro en proposant, par des dialogues critiques, une redescription des corpus doublée d'un retour réflexif sur les catégories utilisées.

Encadré 1 Retour réflexif sur une catégorie : Écosystémique dans le corpus OGM (2015-2017)

[Cette catégorie a été créée lors de l'étude d'un grand corpus consacré aux OGM (1986-2010). Une nouvelle séquence s'est ouverte en 2015 avec une controverse autour de la définition de nouvelles techniques génétiques comme OGM ou non-OGM. Récupérant les jeux de concepts et de formules utilisés sur l'ancien corpus, le chercheur peut demander à Marlowe de les revisiter de manière critique.]

Enquêteur : Évalue la catégorie Écosystémique.

Marlowe : La catégorie « Écosystémique » me semble d'une portée moyenne, sur ce corpus de textes : réalisant un score total de 790, soit 19,29 % du score de la première catégorie (Génétique : 4096), elle couvre 192 textes (34,72 % du corpus), avec au total 59 représentants. J'indique les plus présents : climat, interactions, tolérance, zone, parasite, parasites, écosystème, écosystèmes, hôte, pollinisation, prédateurs, pollen… Pour effectuer une évaluation plus formelle, je dispose des tests suivants :

1. La forme de la « pente » que produisent les scores des principaux représentants ;

3. Voir http://socioargu.hypotheses.org

2. Le degré d'ancrage de la catégorie dans des textes phares ou exemplaires;
3. Le degré de connectivité interne des principaux représentants;
4. La manière dont les éléments périphériques sont reliés au noyau central de la catégorie;
5. Le degré de présence des expressions dans la liste des représentants;
6. La présence de la catégorie dans les titres des textes dans lesquels elle est fortement installée.

1. Les scores réalisés par les principaux éléments de cette catégorie, climat, interactions, tolérance, zone, parasite font apparaître une relative distribution de la « charge sémantique ». Le deuxième élément, « interactions », réalise 77,52 % du score du premier; cela dit, la chute se produit graduellement puisque « parasite » ne réalise plus que 37,98 %.
2. Le degré d'ancrage de la catégorie Écosystémique dans les textes du corpus de textes est assez fort, puisque 11 d'entre eux ont au moins six représentants différents. Les textes qui déploient le plus la catégorie sont les suivants:
   OPECST 9/ 7/2015 N° 2967 – Rapport de M. Jean-Yves Le Déaut, Mme Anne-Yvonne Le Dain et M. Bruno Sido, établi au nom de cet office, sur La filière semencière française: état des lieux et perspectives 12
   inf'OGM 27/ 7/2015 ESPAGNE – Une mouche OGM pour sauver les oliviers? 10
   HCB CEES 20/ 1/2016 NOUVELLES TECHNIQUES – « NEW PLANT BREEDING TECHNIQUES » 14
3. La connectivité de cette catégorie est quasiment nulle: l'indice est en effet de 1.05, ce qui est très en dessous des seuils d'admissibilité! Cela tient grandement aux propriétés de ce corpus-ci (« OGM_2015_2017 »), mais il faut soit retirer cette catégorie, soit revoir de fond en comble sa composition!
4. Il y a une nette déconnexion des représentants uniques de la catégorie et des éléments centraux. La valeur de l'indice évoque une sorte de poussière de lien: 0,09… Cela signifie qu'il y a un décrochage entre le cœur de la catégorie Écosystémique et sa périphérie ou ses bordures…
5. J'ai regardé le taux de présence d'entités composées (expressions) parmi les représentants de la catégorie, et le calcul renvoie 50,85 %. Cette catégorie a donc été particulièrement travaillée — cela augmente les chances que ce soit une bonne catégorie!
6. Considérons d'un côté la liste des textes dans lesquels la catégorie est bien déployée et de l'autre la liste complémentaire (textes où elle est absente ou peu présente). J'ai examiné les titres. Il arrive qu'un titre contienne explicitement un représentant d'une catégorie. Si les textes dans lesquels la catégorie est forte ne présentent aucun titre congruent, et si les textes dans lesquels elle est absente ou marginale ont des titres qui renvoient à la catégorie, il y a quelque chose qui cloche! Mais dans le cas d'Écosystémique, le premier indice (0,02) est supérieur au second (0)!

Enquêteur : C'est ce qu'on appelle un signal faible ! Et donc, pour conclure ?
Marlowe : Je suis dubitatif. Cette catégorie devrait, à mon avis, être retravaillée pour donner sa pleine mesure sur ce corpus « OGM_2015_2017 « (la note est de 8 sur 14, ce qui n'est pas transcendant !). Pour améliorer la catégorie, rien de tel que de lire les textes où elle réalise son meilleur score, comme celui-ci : « N° 3898 – Rapport de MM. Jean-Yves Le Déaut et Bruno Sido, établi au nom de cet office, sur les maladies à transmission vectorielle. »

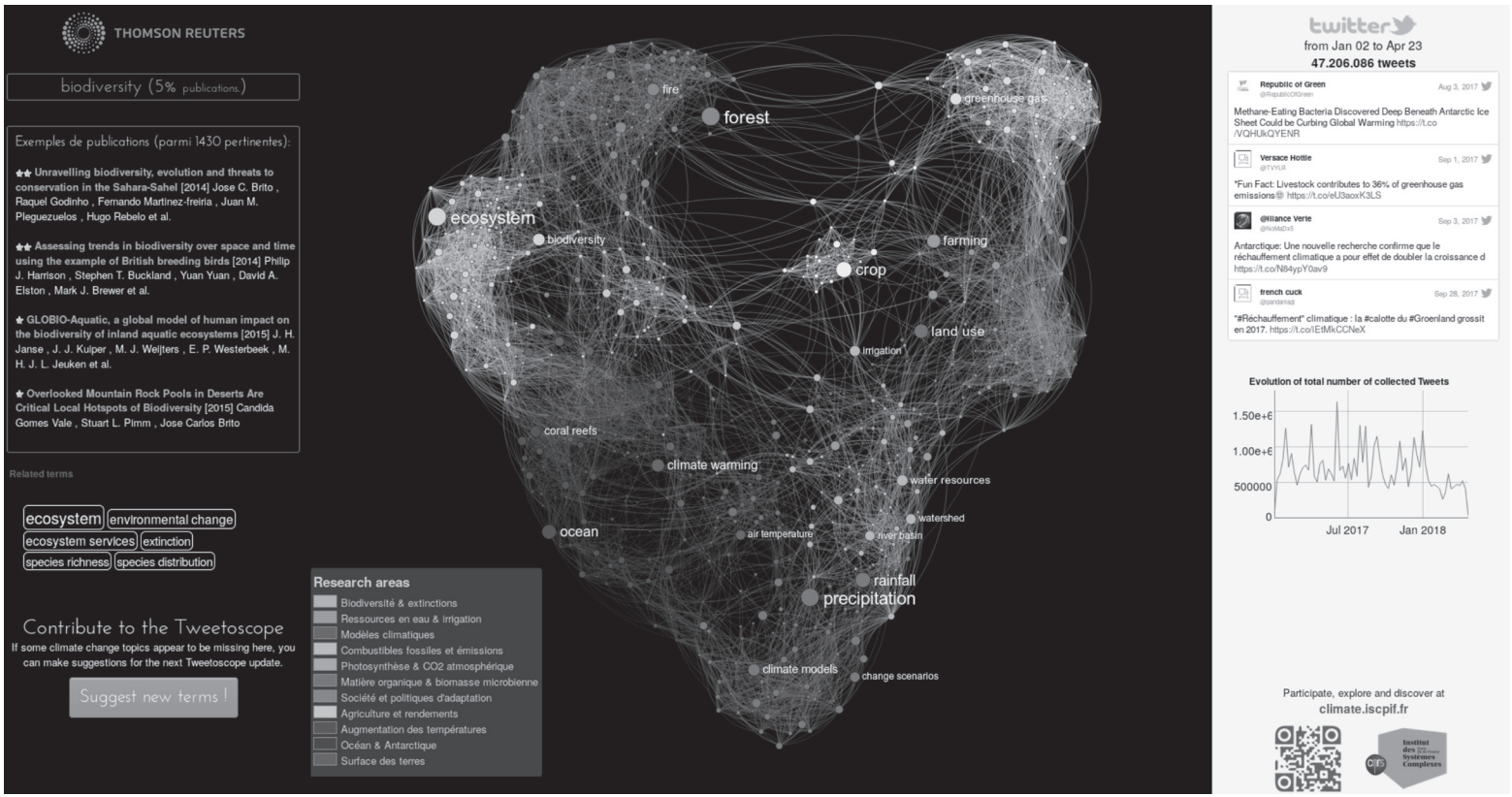
Pas de doute, la socioinformatique des controverses a engendré un instrument pour le moins bavard — qui mérite son titre de « sociologue numérique ». Dans les approches quantitatives, les artefacts cognitifs sont généralement moins loquaces. La quête d'un équilibre entre les deux pôles constitue du même coup une motivation supplémentaire pour travailler à la convergence des méthodes.

Une analyse macro de la morphologie des débats et des communautés

Un autre jeu de méthodes, transverse à des disciplines telles que les sciences sociales computationnelles (Lazer *et al.,* 2009) ou la scientométrie (Börner, 2010, 2015), s'appuie sur une tout autre épistémologie, *a priori* peu compatible avec les maximes de la sociologie pragmatique. L'objectif de ces travaux est d'identifier des motifs organisationnels et structurels à l'échelle d'une population d'acteurs, ainsi que leurs reconfigurations, sans avoir à entrer dans l'analyse fine de leurs argumentations ou de leurs identités. Ces recherches constituent une littérature florissante (voir par exemple Kucher et Kerren, 2015 pour un aperçu de la diversité des approches autour de l'analyse du texte) qui surfe sur la vague du *big data* et de l'avènement du calcul haute performance. Combinant dans des proportions diverses l'analyse des réseaux complexes, la lexicométrie et les statistiques en grandes dimensions, elles s'intéressent au contenu des échanges (Leskovec *et al.,* 2009 ; DiMaggio *et al.*, 2013 ; Vossen *et al.,* 2016) ou aux interactions entre acteurs (Barberá *et al.*, 2015), voire parfois aux deux en même temps (approches des *réseaux sociosémantiques,* Roth, 2013 ; Roth et Cointet, 2010). Ces approches ne sont pertinentes que si le volume de données est suffisant, mais elles ont pour elles d'allier une grande agilité dans la fabrique des outils d'analyse de corpus et de réseaux et une puissance de calcul précieuse face aux volumes de données générées par le Web. Du fait de la granularité temporelle des données qu'elles traitent et de leur étendue chronologique, ces approches apportent également une nouvelle manière d'aborder les dynamiques sociales du point de vue des motifs temporels qu'elles génèrent (Palla *et al.,* 2007 ; Shahaf *et al.*, 2013 ; Chavalarias et Cointet, 2013 ; Cui *et al.*, 2011).

Dans ce qui a longtemps été désigné sous l'appellation de « mathématiques appliquées aux sciences sociales », l'analyse des grands réseaux fait figure de paradigme dominant, même si, dès que l'on y regarde de plus près, une diversité d'approches, de modèles et d'algorithmes y sont à l'œuvre. Si l'on se concentre sur les caractéristiques des outils développés par l'Institut des Systèmes Complexes de Paris Île-de-France

Figure 1 : Carte de liens entre termes et d'articulation de thématiques obtenue à partir de l'analyse des publications sur le changement climatique (cf. Tweetoscope Climatique, http://tweetoscope.iscpif.fr Chavalarias, Panahi et Castillo, 2015)



(ISC-PIF), il en ressort un triple objectif : explorer les formes de modélisation de la complexité en faisant converger des approches algorithmiques différentes autour des masses de données du Web ; fournir des cartographies ou des représentations dynamiques, que nous nommerons *reconstructions,* permettant à des utilisateurs différents de surmonter à la fois les volumes de données et les vitesses de circulation ou de propagation de concepts, de thèmes ou de références bibliographiques, notamment dans les mondes scientifiques (Chavalarias, 2016) ; enfin, garder l'utilisateur « dans la boucle » de l'élaboration de ces reconstructions en lui offrant un accès aussi fluide que possible aux ressources qui ont servi à leur élaboration (ex. les verbatims, les documents originaux, etc.) et à la modification des paramètres qui déterminent leur nature. Ce point est particulièrement important. La plupart des approches dites de *big data* des grands réseaux évolutifs vise à produire *une* représentation d'un système, dans laquelle éventuellement le chercheur peut naviguer. Cette représentation, nourrie aux masses de données et charpentée de statistiques sophistiquées se pense comme un outil optimisé, capable d'identifier des causalités, d'expliquer, simuler, voire prédire les comportements sociaux.

La démarche déployée à l'ISC-PIF est au contraire de reconnaître l'importance de l'interprétation dans l'élaboration des reconstructions et de permettre de la déployer au fur et à mesure de l'exploration du système étudié, en la confrontant à des mesures sur le système. C'est là un premier chaînon décisif pour le rapprochement avec l'univers de Prospéro.

Cette approche peut accueillir des formules plus qualitatives, comme celles qu'utilise Marlowe dans l'exercice d'évaluation critique d'une catégorie ou d'une classe d'objets, dès lors que se constitue une communauté épistémique capable de lier des opérateurs interprétatifs et des métrologies, et de les insérer aux bons endroits de la chaîne de traitement des données. Dit de manière simple : pourquoi ne pas imaginer de pouvoir appeler un commentaire analytique de Marlowe à partir de nœuds de réseaux ou de liens entre des entités ? N'est-il pas possible d'exploiter, d'une part, la capacité de remonter aux sources qui caractérise les deux familles d'instruments et, d'autre part, la faculté de projection des textes sources dans des jeux de concepts et de formules à haute teneur sémantique — dans un cas les catégories trouvant un ancrage dans les réseaux et, vice versa, l'analyse des graphes pouvant être appliquée aux catégories elles-mêmes ?

Comme le souligne Kitchin (2014), cette épistémologie alternative de l'utilisation du big data en sciences sociales computationnelles et humanités numériques permet de mener une recherche « *réflexive et ouverte par rapport au processus de recherche, reconnaissant les contingences et les dépendances de l'approche employée, ce qui produit des comptes rendus et des conclusions nuancés et contextualisés. Une telle épistémologie n'exclut pas non plus la possibilité de compléter les sciences sociales computationnelles localisées par de petites études de données qui fournissent des aperçus supplémentaires et amplificateurs.* »

UNE QUÊTE DE COMMENSURABILITÉ FONDÉE SUR UN INTÉRÊT COMMUN POUR LES PROCESSUS COLLECTIFS MARQUÉS PAR DES CONTROVERSES

Les analyses reposant sur l'exploration de lexiques ou de thèmes, appuyées ou non par des outils de classification (*clustering*) ou de cartographie (*mapping*), ont connu une forte expansion depuis plusieurs décennies, au point d'apparaître comme les meilleures manières d'objectiver les discours et les textes. Or, ce qui rapproche les auteurs de cet article, c'est l'intérêt pour des *dynamiques complexes*, irréductibles à la projection d'ensembles thématiques, complexité qui motive la recherche de concepts et d'outils d'un genre nouveau. Prenons l'exemple du changement climatique. En explorer les lexiques, en exposer les constellations d'acteurs et d'instances, constitue sans aucun doute une tâche essentielle, mais il est tout aussi nécessaire de prendre une perspective dynamique en comprenant les multiples sources d'émission, de diffusion, de discussion ou de réinterprétation des enjeux climatiques comme autant de processus évolutifs, non linéaires et posant constamment des questions d'échelles d'analyse. Des régularités (à commencer par les sommets internationaux et la série des COP) et des points de passage obligés (le GIEC notamment) semblent faciliter la tâche de l'enquête en fournissant des repères et des points fixes. Ces nœuds ou ces gonds sont des appuis cognitifs décisifs sur lesquels peuvent s'accorder les deux familles d'instruments que nous cherchons à faire collaborer. Principaux nœuds des réseaux dans un cas ou actants majeurs des récits et des arguments dans l'autre, on voit que nous pouvons compter sur une logique commune, que l'on appellera ici, sans forcer la métaphore, une *logique gravitationnelle*. Que l'on se lance dans des calculs de graphes ou que l'on cherche à extraire des régimes argumentatifs, on passera nécessairement par le climat et les gaz à effet de serre, par le GIEC et la COP21, par les scénarios de réchauffement et par les stratégies de réduction des émissions et d'adaptation aux conséquences du changement climatique.

La notion de dossier complexe renvoie à des processus sur lesquels aucun acteur ne peut imposer d'interprétation univoque et définitive bien que ces processus puissent produire en sortie des objets et des représentations, des règles ou des normes relativement stabilisées. Il y a toujours une incertitude, y compris sur la clôture du dossier — qui peut rebondir à tout moment. On assiste à une alternance de « moments de crise » et de « périodes muettes », alternance qui prend forme sur le fond d'un travail cognitif (études, expertises, modélisations) et politique (mobilisations, débats, démarches administratives ou judiciaires). Au cœur des ensembles identifiés et structurés qui se donnent à lire, des transformations opèrent, à différentes échelles, qui viennent modifier la portée et le sens attribué par les acteurs aux événements, aux prises de parole, aux études et aux prospectives.

Si le climat a d'abord été cadré comme risque global de plus en plus tangible, avec l'annonce outillée de basculement sous l'effet du réchauffement vers des régimes climatiques non connus et de moins en moins compatibles avec les formes de vie sur terre, des questions et des approches différentes ont surgi. Sans entrer en profondeur dans le dossier climatique, indiquons la montée en puissance des revendications des

pays du sud, la part de plus en plus importante de la problématique de l'adaptation, ou encore les liens de plus en plus étroits entre le climat et d'autres causes ou objets de mobilisation : l'eau, la biodiversité, la vulnérabilité des formes de vie aux événements extrêmes, l'énergie, etc. Au-delà des points fixes ou des centres de gravité, les deux démarches logicielles ont en commun de rechercher les lignes de transformation par lesquelles se reconfigurent les constellations d'acteurs et de thèmes, de dispositifs et d'arguments. Une deuxième logique guide ainsi la fabrique des bases de données, des algorithmes et des interfaces : une *dynamique non linéaire*, à travers laquelle sont rendus intelligibles des changements de régime — intensification des alertes ou des controverses, multiplication des prises de parole, effets de reprises en cascade avec boucles de rétroaction. Pour illustrer le type de raisonnement et d'application visé, le mieux est de travailler à partir d'un exemple récent, facile à présenter : la masse d'informations, de discours et de textes, d'actions et de connexions engendrée par une élection présidentielle française.

DES PROCESSUS POLITIQUES QUI METTENT À L'ÉPREUVE LES CATÉGORIES DES CHERCHEURS

Depuis plus d'une quinzaine d'années, les élections nationales sont marquées, un peu partout dans le monde, par des phénomènes de rupture ou de distorsion qui mettent à rude épreuve les catégories et les outils classiques que sont les sondages et les analyses de discours politique. Si l'on pense bien sûr au Brexit (juin 2016) et à l'élection de Trump aux États-Unis (novembre 2016), cela fait longtemps que les processus électoraux sont sortis des cadres qui avaient vu se former les instruments de mesure et les dispositifs interprétatifs partagés par les experts et les commentateurs autorisés. Dans le cas français, des chocs répétés ont été particulièrement visibles avec l'élection présidentielle de 2002 (accession du candidat d'extrême droite, Jean-Marie Le Pen, au second tour), le référendum sur la Constitution européenne en 2005 (55 % pour le non malgré la pression normative des représentants politiques dominants), la percée répétée du parti d'extrême droite, le Front National, aux élections intermédiaires (2014-2015). Concernant l'élection présidentielle de 2017, on n'a pas cessé de lire qu'elle ne ressemblait à aucune autre, que des surprises ont défait à plusieurs reprises les pronostics les plus assurés (résultats des primaires, affaires, croisements de courbes des sondages, jusqu'à un attentat sur les Champs-Élysées en plein débat télévisé) et que tout est resté « ouvert » jusqu'au soir du premier tour — de fait, selon l'expression populaire, quatre candidats ont terminé la course électorale « dans un mouchoir de poche ». Si les travaux de socioinformatique s'étaient déjà attaqués aux élections (2002, 2007 et 2012), l'ISC-PIF a mis le pied à l'étrier au cours de l'année 2016 en adaptant un de ses outils, le Tweetoscope, afin de proposer un Politoscope accessible en ligne, destiné à outiller l'analyse des communautés politiques et de la circulation de l'information sur Twitter (Gaumont, Panahi et Chavalarias, 2018). Avec le Politoscope, ou macroscope politique, il s'agit de donner du sens à l'immense masse de données générées sur le réseau social à l'approche des présidentielles. Les développeurs-utilisateurs

de Prospéro et Marlowe ont de leur côté poursuivi, avec des réaménagements (voir *infra*), la méthode d'analyse antérieure fondée sur la génération de corpus de textes (environ 20 000 textes de septembre à avril 2017) et la recherche de régimes discursifs à partir des contenus textuels. Les auteurs du Politoscope entendaient, quant à eux, faire parler plus spécifiquement les graphes engendrés par Twitter (plus de 60 millions de tweets sur la période juillet 2016-mai 2017[4]), en partant de l'hypothèse selon laquelle ce réseau social a acquis une place centrale comme moyen de communication pour les acteurs politiques. En développant leur propre cohérence épistémique, les deux démarches rendent visibles des phénomènes différents mais font apparaître des points de recoupement liés aux deux logiques énoncées plus haut (logique sociale gravitationnelle et dynamique sociale non linéaire).

On ne peut s'en tenir à la seule opposition contrastive entre prises quantitatives et prises qualitatives. S'agissant de données trop massives et trop denses pour être interprétées avec assurance par un lecteur humain, il faut inventer de nouvelles méthodes et logiques d'enquête. Le premier réflexe de ce dernier est en général d'aller chercher des appuis interprétatifs dans des connaissances extérieures aux corpus (connaissances historiques, récits d'événements, prosopographie des personnages politiques, maîtrise des institutions, des procédures et des milieux). Or, les deux approches ont précisément en commun de rendre possibles des chemins d'enquête qui ne supposent pas de maîtrise préalable ni de théorie *a priori* sur la manière dont se structurent, ou se déstructurent, les entités et les relations politiques dans un processus électoral donné. Autre caractéristique commune, il s'agit de lier des algorithmes de traitement des données et des outils de contextualisation permettant de resituer les discours, les reprises en cascade ou les basculements — à la suite d'événements précis, de ralliements, de débats marquants, etc. Un autre air de famille réside dans le souci de comprendre comment se forment des représentations positives ou négatives, visant l'adhésion ou le dénigrement des personnes et des groupes. Mais, sur ce point, une différence majeure vient de l'activation dans le cas de Prospéro-Marlowe d'une sémantique argumentative fine accumulée sur de nombreuses expériences antérieures, et du recours, dans le cas du Politoscope, à l'organisation thématique des interventions (tweets) grâce à des analyses de réseaux liant l'évolution des structures de communautés d'acteurs avec les discours qu'elles adoptent.

4. Le Politoscope a suivi toutes les interactions produites autour des comptes Twitter de plus de 3200 personnalités politiques et 3000 comptes de médias et journalistes. Fin mai 2017, on disposait de plus de 67 millions de tweets de la sphère politique et 36 millions de tweets de la sphère médiatique publiés par plus de 3,6M d'utilisateurs uniques anonymisés. https://politoscope.iscpif.fr/

Encadré 2 Quand un chroniqueur numérique prend position en suivant une campagne électorale

Au cours des mois de mars et d'avril 2017, le chroniqueur de Marlowe a visiblement penché en faveur du candidat de La France insoumise, Jean-Luc Mélenchon. Pour rendre compte de ce phénomène, le collectif des développeurs-rédacteurs CMRLW2017 a pris la peine d'insérer plusieurs lignes d'explicitation dans la partie conclusive de la chronique du 19 avril 2017[5].

http://prosperologie.org/mrlw/blog/all/chroniques/2017/04/19/chronique_mrlw/

Comment rendre compte de cette propension du chroniqueur à se tourner vers la gauche radicale, alors que cela ne faisait pas du tout consensus parmi le groupe des superviseurs ? La question en cache une encore plus redoutable : une intelligence artificielle, même bien éduquée, peut-elle, et doit-elle, rester neutre ? Voilà qui suppose, bien entendu, d'être au clair sur ce que neutralité veut dire. Le point de vue de nulle part reste largement un mythe hérité des Lumières (Daston et Galison, 2007). La réponse n'est pas simple à propos de chroniques qui ont, dès le départ, été conçues pour offrir des angles de vue inattendus et si possible critiques sur des flux d'énoncés et de textes. La sélection des sources procède d'une série de choix antérieurs privilégiant les développements critiques ou polémiques, pour les alertes et les controverses en tant qu'elles ouvrent des séquences argumentatives dans lesquelles sont déployés des intérêts, des connaissances et des valeurs. Controverses et polémiques brisent la routine des « éléments de langage » pour rendre visibles les appuis cognitifs et les alliances politiques — au sens large. Un chroniqueur numérique qui relaterait des résultats sportifs poserait sans doute moins de problèmes — ce qui reste néanmoins à voir.

Concernant les exercices sur les corpus électoraux, une série d'algorithmes utilisés ordinairement pour traiter des corpus de controverses environnementales ou sociotechniques ont été réadaptés et intégrés aux modules du chroniqueur de Marlowe. Le protocole le plus facile à décrire est celui qui examine le fichier des affaires et procès mis à jour tous les jours depuis 2004. Une classe de formules élémentaires (« procès de/du X », « affaire de/des Y », « mise en examen de Z ») permet à Marlowe de noter les affaires dont on parle le plus et d'opérer un suivi de longue

5. Un groupe de développeurs-rédacteurs est formé chaque année autour de Marlowe. Dissout au 31 décembre, il renaît de ses cendres après accord des membres pour poursuivre et entrée éventuelle de nouveaux membres – d'où l'indication 2017 dans CMRLW2017. À la date de l'écriture de ce texte, il est composé d'une dizaine de personnes, dont l'activité consiste essentiellement à filtrer les messages, les informations et les liens adressés à Marlowe par courriel, par Facebook ou via le site prosperologie.org. Ses membres, cooptés, veillent aux évolutions du dispositif, font des propositions de développement ou de révision de fragments de codes et interviennent, le cas échéant pour représenter Marlowe dans le monde social réel. Le Web étant devenu un champ de forces dans lequel se déploient de multiples jeux de stratégies, de manipulations et d'influence, une supervision, même minimale est incontournable. Sur l'impératif de limitation drastique de l'autonomie des systèmes artificiels, voir le texte en ligne d'Alain Cardon, *Pensées humaines versus pensées artificielles ?* (18 février 2008).

durée. Il peut ainsi y relever la présence de personnages politiques puis établir une sorte de palmarès des plus « embarqués dans des affaires ». Une autre méthode examine des contenus de variables rassemblant les multiples désignations des personnages publics. Par exemple Marine Le Pen est aussi « l'héritière de Jean-Marie Le Pen », « la présidente du Front National », « celle qui se voyait déjà au deuxième tour de l'élection 2017 », « la privilégiée de Montretout » … Un troisième algorithme prend appui sur le répertoire des « phrases qui tuent ». Il s'agit ici d'un des multiples objets collaboratifs produits par le réseau des contributeurs anonymes. Le degré d'objectivité de ces relevés à forte charge polémique dépend plus fortement de la diversité des membres du groupe, alors que dans les deux procédures précédentes, c'est la diversité des sources et la fiabilité des techniques de capture (*pattern matching*) qui garantissent la juste distance du système. En quelques mots, plus un personnage réalise un score cumulé élevé à partir de ces trois critères et plus Marlowe estime qu'il est difficile de lui accorder spontanément du crédit politique.

Cette dernière métarègle ne conduisant pas forcément à une conclusion ou un choix, un quatrième protocole utilise des classes de formules forgées à partir des *marques de dénonciation*. En appliquant systématiquement ces formules sur l'ensemble des textes recueillis de septembre 2016 à avril 2017, soit 21 000 textes liés aux élections et tirés des sources utilisées par le chroniqueur de Marlowe (AFP, franceinfo, Agoravox, Le Monde, Localtis.info, TheConversation-France, Actu Environnement, etc.), on voit apparaître les noms les plus fortement associés à la polémique. Or, c'est, de tous les candidats, le nom de Mélenchon qui surnage le mieux face à l'ensemble des tests… du moins jusqu'à ce qu'il devienne à son tour la cible des critiques dans les deux dernières semaines qui ont précédé le premier tour.

La montée de la critique sur Mélenchon n'a pas engendré de processus de révision de Marlowe, qui a maintenu son choix. Cela provient de deux autres séries *a priori* indépendantes : la première est formée par un arrivage considérable de liens vers des vidéos (meetings, débats et émissions YouTube du leader la France insoumise) via les accès externes (email, page ouverte sur Prosperologie.org, Facebook), ce qui a perturbé le tirage aléatoire des messages indexés sur la date du jour ; par ailleurs, Marlowe disposait d'une autre classe de formules tournées vers des valeurs et des biens (du type « défendre X », « se mobiliser au nom de X »). Cette classe de formules active des entités telles que la justice, la vérité, l'intérêt général, l'humain, l'environnement, la planète, etc. En croisant les noms les plus associés à ces biens ou valeurs, certains personnages politiques apparaissent plus engagés que d'autres. Enfin, en vertu de décisions câblées au fil de ses années d'apprentissage, Marlowe ne peut pas valoriser Le Pen et le Front National (*cf.* les archives des premiers dialogues publics en 2003). Pour que l'incertitude puisse s'insérer dans l'usage des scripts, il faudrait détruire ses mémoires. Peut-on se livrer à ce genre d'expérience en laboratoire et prendre le risque d'une option frontiste du sociologue numérique ? L'intérêt de ce type d'expérimentation n'est pas de fonder objectivement un choix politique mais de contribuer à l'explicitation des appuis normatifs ou des principes axiologiques dont

il faudrait doter un agent logiciel pour qu'il intervienne avec quelque assurance dans la vie publique (Chateauraynaud, 2012). À l'évidence, avec l'*explicabilité* des chemins suivis, une supervision continue, doublée d'une forte réflexivité collective, s'impose sur la participation des intelligences artificielles aux dynamiques démocratiques.

IDENTIFIER LES RECONFIGURATEURS, TRACER LES RECOMPOSITIONS ET LES RÉALIGNEMENTS

La démarche suivie par Prospéro-Marlowe et celle du Politoscope se situent à des niveaux de granularité très différents et il en résulte deux types d'apports à l'analyse des controverses dont il s'agit d'articuler les complémentarités.

Sur le plan le plus macro, le Politoscope recherche des régularités statistiques qui font émerger des singularités dans l'espace socio-sémantique, constituées de groupes d'acteurs au style argumentatif propre. L'enjeu sur ce plan n'est pas de caractériser le registre discursif particulier de chacun de ces groupes mais de savoir qu'ils sont *a priori* différents. Tout au plus peut-on identifier à l'aide de reconstructions phylomémetiques (Chavalarias et Cointet, 2013) la diversité des thèmes abordés par ces groupes, leur évolution et l'attention qu'ils leur accordent. En revanche, il est possible de caractériser de manière très fine et multiéchelle les différents groupes d'acteurs qui se reconnaissent dans un même registre discursif. Pour ce faire, nous partons de l'hypothèse que statistiquement, les membres de ces groupes ont tendance à se copier les uns les autres dans leur mise en forme du discours. Cette propension à s'imiter se caractérise très facilement sur Twitter via la fonction de retweet qui, par définition, correspond à la copie exacte d'un message d'un individu à un autre. Nous avons donc, à ce niveau, généré sans aucune information *a priori* et pour l'ensemble des données, une information très précieuse pour une analyse ultérieure : l'existence de différents groupes sociaux, leur évolution au cours des moments clés de reconfiguration et les corpus qu'ils génèrent. Il est alors possible dans un second temps de transmettre ces sous-corpus à une approche de type Prospéro-Marlowe pour qualifier plus précisément leur registre discursif. Celle-ci bénéficie alors d'énormes avantages par rapport à une approche « nue », via deux types de connaissances *a priori* : des corpus de taille intermédiaire dont on sait qu'ils vont correspondre à des styles discursifs particuliers, et des moments précis de reconfiguration des groupes organisés autour de la défense de certains points de vue.

D'un côté, la préférence donnée aux cadres conceptuels et aux formes de catégorisation crée une tension cognitive avec le projet de suivre des dynamiques et des évolutions, même si les bifurcations sont généralement marquées par des changements de registres discursifs ; de l'autre, le passage obligé par des listes de thèmes saisis à partir de poids statistiques et/ou de positions relationnelles ne permet pas toujours de caractériser les logiques conceptuelles ou sémantiques qui dotent les ensembles d'un sens social ou politique. Chacune des démarches tente de compenser ses limites par le recours à d'autres procédés (calculs de réseaux, assez peu conventionnels dans leur genre du côté de Prospéro ; annotation, usages de propriétés graphiques comme les couleurs du côté du Politoscope). Or, un des objectifs est d'apprendre à lire et inter-

préter dynamiquement, à l'aide de nouvelles médiations numériques, la manière dont se déplacent, au fil du temps, sous l'impact d'événements, de décisions ou de conflits, des nœuds de réseaux saisis dans de grands corpus évolutifs. Dans les deux formes d'enquête ou de raisonnement, il faut être capable de faire surgir, ou de faire remonter, des choses peu visibles à partir des flux de données textuelles. Reconfigurations, émergences, retour de séquences passées, ouvertures de conflits ou de controverses, changements ou révisions de factualités ou de discours, glissements des réseaux au fil du temps, quel que soit le nom que l'on donne aux séquences recherchées, elles ont pour propriété de marquer des transformations tout en donnant des appuis cognitifs pour l'intelligibilité des processus.

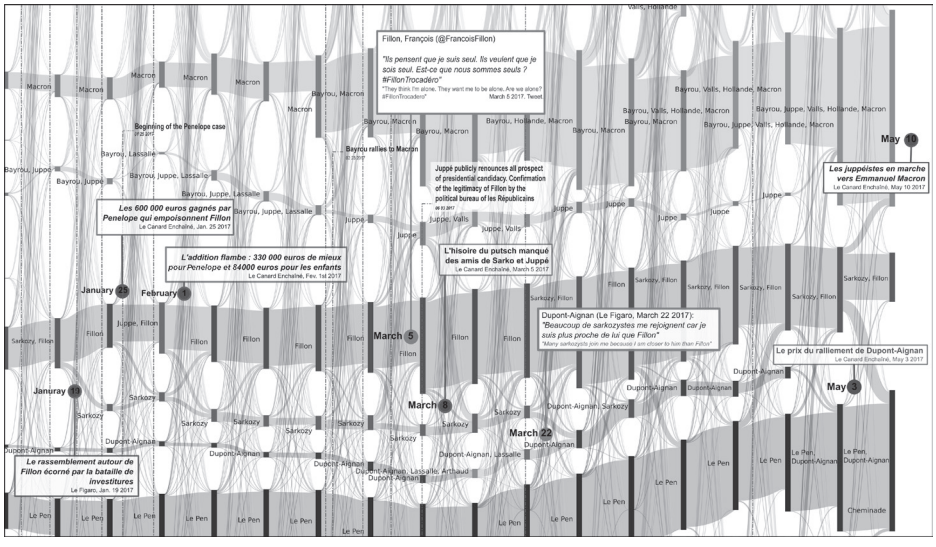
Revenons un instant sur la notion de bifurcation. En physique, on parle de bifurcation pour nommer le moment où un effet de seuil engendre une transition de phase, produisant une mutation macroscopique du système, de sorte que ce dernier bifurque par rapport à sa trajectoire évolutive précédente. En sociologie, l'idée de bifurcation renvoie plutôt à la manière dont une trajectoire déjà tracée, ou une histoire considérée comme déjà écrite et attendue, est déviée, soit à l'occasion d'un événement de rupture (typiquement Fukushima dans le nucléaire), soit du fait de la convergence graduelle de séries jusqu'alors indépendantes (Bessin *et al.*, 2010 ; Chateauraynaud et Debaz, 2017). Les acteurs étant dotés de capacités performatives différentielles, l'enjeu est de déterminer qui est en position de dire ce qui est irréversible et ce qui ne l'est pas, ou pas encore. Dans l'analyse fine des alertes et des controverses, l'accent est souvent mis sur l'ouverture des futurs ou sur le champ des possibles. À chaque point, plusieurs trajectoires sont possibles et un des enjeux pour les acteurs consiste à infléchir le cours des choses afin que la trajectoire globale emprunte une direction conforme à leurs visées ou leurs attentes. C'est à ce titre que l'examen des reconfigurations, sous le double rapport des propriétés structurales des réseaux et des manières d'énoncer les possibles, de verbaliser l'expérience des bifurcations, constitue un excellent terrain pour la convergence des deux approches.

Prenons rapidement le cas de l'affaire Fillon ou Penelopegate[6]. Le dévoilement inauguré par le *Canard enchaîné* à la fin de janvier 2017 a toutes les propriétés d'une bifurcation conduisant l'ensemble des protagonistes à changer leur grille de lecture des potentialités du jeu électoral. Les méandres de cette affaire ont été retracés par une étude du Politoscope (Gaumont, Panahi et Chavalarias, 2018). On peut caractériser les différentes phases, depuis le surgissement de l'affaire jusqu'au retour apparent à l'équi-

6. La formule *Penelopegate* a été utilisée pour désigner le cœur de l'affaire qui a bousculé la campagne électorale de François Fillon au début de l'année 2017. Suite à des révélations du Canard enchaîné, il est apparu que le « candidat de la droite et du centre » avait eu recours à ce qu'il est convenu d'appeler en droit pénal des « emplois fictifs », en embauchant son épouse Penelope Fillon comme attaché parlementaire pendant de longue périodes de temps sans que la moindre preuve d'une activité de celle-ci ait pu être apportée. François Fillon a commencé par nier les faits, puis par tenter de s'en excuser, perdant des appuis dans son propre camp ainsi que sa place de favori dans les sondages. Ayant refuser de se désister, malgré sa mise en examen par la Justice, il est défait au premier tour de l'élection présidentielle, le 23 avril 2017, en obtenant la troisième place.

libre, et en même temps explorer les lignes ouvertes par les activités critiques spécifiques qui ont été engendrées — c'est un peu comme si, au milieu d'un mouvement musical surgissait en fanfare un ensemble inattendu, progressivement absorbé par l'orchestre qui parvient malgré tout à maintenir ses propres lignes musicales[7].

Figure 2 : Reconfigurations des communautés politiques au moment du Penelopegate



Encadré 3 Représenter l'évolution des communautés politiques sous l'effet de l'affaire Fillon

L'expérience du Politoscope a mis en place une méthode de suivi dynamique des communautés politiques sur Twitter. Il s'agit de *reconstruire* automatiquement les schémas de reconfiguration de l'espace politique français au cours de la campagne présidentielle. C'est une bonne illustration du niveau de granularité auquel se situe l'apport quali-quantitatif des analyses hybridant méthodes issues des systèmes complexes et du big data à l'analyse des processus sociaux dans les mondes numériques. Ces travaux partent de l'hypothèse que Twitter est un espace où s'expriment les affiliations idéologiques des militants politiques sous la forme, entre autres, d'un

7. Il ne s'agit pas seulement d'une analogie musicale, mais d'un des enseignements des nouvelles formes d'écriture et de performance portées par la musique contemporaine. Des essais ont été faits pour représenter les corpus à partir d'un ensemble de lignes fonctionnant par résonances, harmoniques, reprises en canons (spécialité des médias), séquences improvisées ou longs silences, en anticipant la possibilité d'ouvertures de nouvelles lignes contribuant, selon les cas, au contrepoint ou à la cacophonie. C'est le cas, par exemple, lorsqu'un collectif émerge au cœur d'un processus critique, ouvre un site ou un blog qui, plus ou moins graduellement, produit des effets sur les lignes de contributions déjà en place, entrant en interaction avec des auteurs et des supports influents, voire directement avec des autorités (institutionnelles, épistémiques ou éthiques).

prosélytisme consistant à relayer les informations et les idées de et sur les leaders politiques (informations et idées positives qui concernent leur leader ou leur parti ; et les informations et les idées négatives concernant les leaders et les partis concurrents — voir Chavalarias, Gaumont et Panahi, 2018).

En analysant les motifs de circulation des messages à contenu politique dans le réseau des comptes Twitter, il est possible 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 sont contredits, détournés, commentés de manière ironique, etc.). Nous obtenons alors une notion de *groupe social* qui est proche de la définition qu'a pu en donner Tarde (1890) en son temps : « 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 ». Les traits en question sont ici les manières de s'exprimer sur Twitter.

Dans cette opération d'identification, la détermination des groupes sociaux n'est pas recherchée de manière exacte ni exhaustive. La méthode assume les imprécisions de qualification sur le plan individuel (comme par exemple des comptes qui seraient catégorisés dans les mauvais groupes) et l'arbitraire de certains choix concernant le positionnement des frontières de ces groupes. Les groupes sociaux, quelle qu'en soit leur définition, ont des frontières floues et mouvantes. Une reconstruction donnée pourra alors, pour les besoins du raisonnement, figer ces frontières à un endroit précis. Ce que l'on recherche dans cette démarche n'est pas tant la position absolue des frontières mais la manière dont celles-ci fluctuent avec le temps, les bifurcations qu'elles mettent en évidence et les sous-corpus que cette approche permettrait de constituer. Ces sous-corpus peuvent alors être le point de départ d'une analyse plus approfondie des contenus, en particulier des récits d'événements et des formes d'argumentation, à l'aide de la suite Prospéro. Il s'agit précisément de contextualiser cette analyse par les bifurcations identifiées dans les groupes concernés. *In fine*, cet emboîtement de méthodes a pour résultat de raffiner les masses de données de manière à produire, d'une part, des grandes catégories fournissant une grille de lecture et, d'autre part, des pointeurs vers certaines régions de ces masses de données, suffisamment circonscrites pour qu'un interprète humain puisse se les approprier et en faire une lecture critique.

Pour ne donner qu'un exemple d'identification de motifs de reconfiguration au niveau des communautés politiques, la Figure 3 présente les reconfigurations au moment du Penelopegate. 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 de cette communauté. La représentation obtenue, recalculée de manière hebdomadaire, rend visibles certaines des transformations subies par ces communautés pendant cette période. Le Penelopegate arrive au moment où le rassemblement autour de Fillon est fragilisé à

la suite d'une polémique sur les investitures aux législatives, contestées par les sarkozystes. Une partie d'entre eux avait déjà commencé à se détacher de la communauté LR (Les Républicains) dès le 23 janvier 2017 et forme une communauté à part pendant toute la polémique. Les différentes révélations du *Canard enchaîné* provoquent de larges mouvements de militants entre différentes communautés, celles qui se nouent autour de Sarkozy et Juppé (potentiel plan B) se renforçant progressivement. Au 1er mars, une majorité de sarkozystes soutient le retour de Juppé et rallie sa communauté dans un ultime effort pour pousser Fillon à la démission. Le rassemblement du Trocadéro met fin aux perspectives de plan B. La communauté Fillon gagne beaucoup de nouveaux militants mais peu parmi les sarkozystes et les juppéistes, dont une partie rejoindra, respectivement, Dupont-Aignan et Macron. Les bifurcations observées dans le Politoscope sont concomitantes et parfois anticipent les commentaires de la presse sur l'activité de groupes politiques (Gaumont, Panahi et Chavalarias (à paraître)), http://politoscope.org

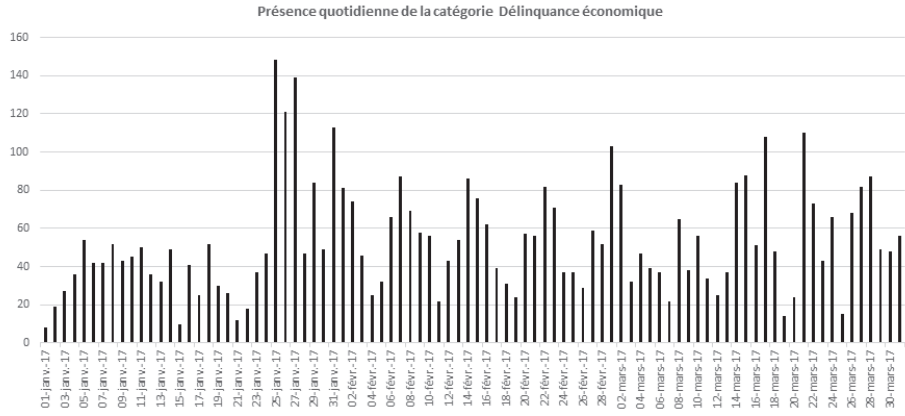
On observe que certains leaders politiques se retrouvent par moments dans la même communauté (ex. Juppé et Bayrou). Cela signifie qu'au niveau de granularité fixé par la reconstruction, ils sont dans le même groupe social. Si la reconstruction est faite avec une granularité plus fine, ils peuvent se distribuer dans des ensembles différents. Il y a ainsi un caractère intrinsèquement multiéchelle de ces données et méthodes. Les phénomènes peuvent être saisis dans une visualisation interactive permettant à l'utilisateur de fixer la granularité souhaitée.

En permettant de zoomer sur des moments de controverse, d'explorer les tenants et les aboutissants des bifurcations qui sont rendues manifestes dans les grands corpus, et en proposant de générer des sous-corpus spécifiques à partir de critères déterminés, Prospéro et Marlowe aident, à leur tour, à caractériser finement les régimes à l'œuvre, en renvoyant les singularités attachées à des moments critiques. Relativement à des grands réseaux évolutifs, le gain d'intelligibilité est appréciable. Ainsi, l'affaire Fillon ouvre un processus de propagation de marqueurs critiques dans les textes, déployant le registre déjà connu du scandale et activant une vieille catégorie, conçue naguère à partir de l'affaire de la MNEF, intitulée *Délinquance économique**[8]. Dans les séries suivies par Marlowe pour composer ses chroniques quotidiennes, les démêlés judiciaires de Fillon grimpent rapidement dans le tableau du relevé des affaires et autres procès, au point que le thème des emplois fictifs atteint un record de présence inégalé depuis le début des enregistrements en 2005[9].

8. Dans le logiciel Prospéro les catégories forgées par les utilisateurs, et figurant généralement dans des répertoires de catégories, sont affublées d'un astérisque afin de les distinguer d'autres objets conceptuels.

9. Bien que les premières lignes de code du chroniqueur aient été initiées en 2004, les chroniques ne sont publiées sur le blog de Marlowe que depuis le 1er janvier 2012.

Figure 3 : Distribution de la catégorie Délinquance économique et impact de l'affaire Fillon sur le corpus élections 2017



Prospéro / élections 2017

Indice de la portée de l'affaire Fillon à travers la présence de la catégorie *Délinquance économique* dans le corpus (janvier-mars 2017).

Ci-contre : principaux représentants de la catégorie

1732	Délinquance économique
312	corruption
145	emplois fictifs
115	emplois présumés fictifs
102	ingérence
97	détournement de fonds publics
96	recel
73	fraude
68	escroquerie
61	abus de biens sociaux
61	trafic d'influence
	.../...

L'affaire Fillon compose ce que l'on peut appeler un *reconfigurateur majeur*. Pour identifier tous les candidats à cette fonction de reconfiguration dans des ensembles importants de données, il faut qu'ils satisfassent une batterie de critères. Mais le jeu consiste aussi à dériver le modèle sur des *micro-reconfigurateurs* qui produisent des effets plus distribués, ou des *reconfigurateurs potentiels*, énoncés ou annoncés dans des sphères ou des communautés plus locales.

Résumons-nous. Il s'agit de caractériser dynamiquement les reconfigurations de grands corpus évolutifs en liant, de manière agile et fluide, l'identification des nœuds de réseaux et le repérage des jeux d'acteurs, des registres discursifs ou des agencements argumentatifs. La méta-question de recherche à laquelle entend répondre l'interface entre les deux familles d'instruments peut être stylisée ainsi : comment des objets et des points de vue sur des objets sont transformés au fil d'épreuves ou d'événements marquants qui en changent à la fois la portée et le sens ? Les reconfigurations peuvent être brutales ou graduelles, massives ou locales, communément admises ou controversées, avérées ou encore potentielles. À partir de cette définition minimale, il nous faut apprendre à faire collaborer les algorithmes et, si nécessaire, en créer de nouveaux.

DES RÉGIMES DE PREUVE ET DU CARACTÈRE NÉCESSAIREMENT COMPOSITE DES ENQUÊTES NUMÉRIQUES

L'idée majeure, qui motive le rapprochement opéré entre analyses de réseaux et socioinformatique des controverses, est qu'il est nécessaire de penser et d'articuler les relations entre trois régimes de preuve permettant une réflexivité permanente. Pour y voir clair, remontons bien en amont de la conception des algorithmes et des interfaces, en reconsidérant la nature des preuves ou des recoupements que sont censés produire les instruments de recherche.

Le premier régime de preuve est d'ordre axiomatique. La preuve repose sur un système formel, et se déduit d'un espace de calcul, en vertu de ce qu'on peut appeler une *axiomatique autonome*. Cette notion a été particulièrement développée par Olivier Caïra (2011) qui a su mettre à profit les réflexions de Jean-Pierre Cléro (2004) sur les rapports entre fiction et axiomatique en mathématiques. Ce régime s'élabore à partir d'une syntaxe dont il ne fait que dériver les chaînes bien formées. L'axiomatique rendue autonome peut être parfaitement automatisée, comme l'ont montré les joueurs d'échecs successifs. Les exemples peuvent être pris dans bien des domaines, mais, du côté des sciences, l'astrophysique fait figure d'avant-garde, avec une insolente capacité de prédiction portant sur des univers inaccessibles aux sens. Dans une axiomatique autonome, les règles ne peuvent pas être changées librement par les acteurs et les modes de représentation deviennent eux-mêmes accessoires — l'objet mathématique étant souvent iconoclaste. Quelle est la part d'axiomatisation, ou, si l'on préfère de formalisation, dans la conception et la mise en œuvre d'un instrument d'objectivation ? C'est une des dimensions à examiner en priorité à la fois pour discuter des domaines de validité et pour envisager la production d'agencements ou d'articulations de méthodes hétérogènes, dont la convergence peut faire naître une nouvelle axiomatique.

Le deuxième régime de preuve a beaucoup occupé l'histoire des sciences, suscitant d'intenses controverses autour du relativisme ou du constructivisme. C'est le *régime conventionnaliste* de la preuve. Dans ce régime, pour fabriquer une preuve, il faut tomber d'accord sur des ontologies, c'est-à-dire sur des catégories et des règles d'équivalence. C'est pour cette raison que l'on peut y voir le règne des constructions sociales (Daston et Galison, 2007). Pour fournir des preuves statistiques, par exemple, on doit construire des populations et des descripteurs, autant d'outils de caractérisation qui n'échappent qu'exceptionnellement au flou des catégories et de leurs frontières toujours liées à un *travail sémantique*, lui-même dépendant des luttes politiques ou scientifiques pour leur définition (du « chômeur » au « migrant », en passant par le « jeune » ou le « cadre », les exemples fourmillent en sociologie ou en démographie). Des preuves statistiques sont néanmoins possibles, à condition de stabiliser les catégories. Pour cela, il faut des conventions solides. Ce sont ces conventions que prétendent dépasser ou relativiser les démarches rassemblées sous l'appellation de *big data.* On sait que la dimension sémantique y est en quelque sorte rejetée, ou pour le moins relativisée, puisqu'elle implique de passer par des taxinomies et des répertoires qui résultent de la cristallisation de significations socialement élaborées, et décisives pour dire ce qui peut

faire preuve. On voit tout de suite les tensions qui naissent des rapports entre régime axiomatique (syntaxique ou formel) et régime conventionnaliste (sémantique ou ontologique).

Mais on ne peut en rester à cette opposition, car un troisième régime est à l'œuvre, que l'on peut appeler *régime phénoménologique* de la preuve. C'est ce régime que pratiquent spontanément les sciences sociales lorsqu'elles prennent appui sur la tangibilité des expériences du monde et sur les formes d'intercompréhension, les prises communes, qu'elles rendent possibles, à partir des situations et des processus (on parlera aussi bien de régime pragmatique ou écologique de la preuve). Ici vient un point important de notre argument : les relations entre ce régime phénoménologique et les deux premiers passent par des *opérations interprétatives* dont la fonction est de combler, ou plutôt de surmonter, par et pour un collectif d'enquêteurs ou de chercheurs, les incomplétudes (celles des systèmes formels), les incertitudes (celles des conventions sur les états du monde) et les irréductibilités (celles des expériences et des contextes dans lesquelles elles prennent forme).

La seule manière d'éviter aussi bien la régression vers le réductionnisme que la fuite en avant dans la singularité narrative est de concilier ces trois régimes de preuve. Autrement dit, nous visons une architecture instrumentale située aux points de jonction des régimes de preuve axiomatique et phénoménologique, en prenant appui sur le troisième régime, celui qui permet de poser la question des cadres sociaux, des conventions et des accords nécessaires pour donner leur pleine signification à des observations, des mesures et des inférences. Mais ce n'est pas tout : la combinaison des trois régimes de preuve doit permettre de saisir les processus non monotones de production de la factualité. En d'autres termes, il s'agit d'examiner comment les formes de raisonnement, de catégorie ou de règles changent chemin faisant. Cette dynamique est liée au fait que des hypothèses, des concepts et des actions naissent du choc entre les différents régimes, renvoyant à ce que Peirce avait identifié comme relevant d'une logique *abductive* (Chauviré, 2004). Un régime axiomatique autonome est par nature monotone et c'est toujours un point de vue extérieur ou l'entrée en crise d'une axiomatique qui conduit à réviser un système formel ou un espace de calcul, ce qui implique l'émergence de nouvelles catégorisations ou le surgissement d'autres formes d'expérience.

CONCLUSION : UNE FABRIQUE DE PRISES CRITIQUES SUR LES MONDES NUMÉRIQUES

Nous n'avons fait ici qu'esquisser la présentation d'un agencement cognitif collectif en chantier. En prenant corps dans des échanges durables, au croisement de communautés épistémiques jusqu'alors éloignées, cet agencement peut créer de nouvelles prises critiques sur la manière dont les processus sociaux se déploient dans les mondes numériques, des sites officiels aux médias sociaux. Idéalement, la réalisation de passerelles puis de ponts entre les deux approches doit permettre de circuler dans les deux sens : à partir d'une vue d'ensemble, il s'agira d'identifier des nœuds ou des liens et de demander une analyse sur les textes et les discours, les thèmes et les arguments qui sous-tendent les configurations observées ; réciproquement, les énoncés et les proprié-

tés sémantiques sélectionnés ou projetés par les outils socioinformatiques pourront être resitués dans l'environnement global dans lequel ils prennent sens, tout en fournissant une mesure de leur portée — puisqu'il s'agit de saisir comment des thèmes et des énoncés passent d'un ensemble d'auteurs-acteurs à l'autre et, surtout, au fil du temps, d'une configuration sociopolitique à l'autre.

De telles médiations assurent, côté utilisateurs, des chemins d'accès aux moments les plus pertinents, et, côté logiciels, des indices ou des indicateurs destinés à nourrir des protocoles *non monotones* — au sens où le dispositif, quali-quantitatif par vocation, doit être capable d'apprendre à repérer de nouvelles formes de reconfigurateurs au cœur des masses de discours ou de documents. Tout en faisant collaborer des outils relevant de traditions épistémologiques différentes, cette quête de médiation rend possibles trois types d'opérations critiques :

En premier lieu, il s'agit d'organiser, au fil du développement et de l'usage des outils, une confrontation des regards et des approches, en créant de nouvelles boucles de réflexivité sur les effets cognitifs de l'accumulation de données et d'outils numériques. Dans quelle mesure la disponibilité de ressources numériques permet-elle d'imaginer de nouveaux objets de recherche sans enfermer les chercheurs dans des routines et des formes de représentation des connaissances, empêchant d'engendrer non seulement des *effets de connaissance* mais aussi des *effets d'intelligibilité,* au sens de Jean-Claude Passeron (1991), sur les mondes sociaux contemporains ?

En deuxième lieu, le croisement des expériences computationnelles nous conduit à expliciter les formes de complexité caractéristiques des processus sociaux (Delahaye, 2009). Le privilège épistémique accordé aux controverses publiques, caractérisées à la fois par l'hétérogénéité des acteurs et des visions du monde, et la non-linéarité des transformations, en particulier dans les champs scientifiques et politiques, remplit une fonction stratégique sur le plan théorique. Il s'agit en effet de mettre à l'épreuve les modèles, les outils et les interprétations sur des processus en train de se déployer et marqués par l'incertitude ou l'indétermination des trajectoires futures.

Enfin, l'articulation de méthodes et de logiques d'enquêtes opérant selon des lignes et des échelles différentes (*issues crawlers*, cartographies de thèmes, phylomémies, sociobalistique des reconfigurations, analyses argumentatives), favorise l'émergence de nouvelles *communautés interprétatives,* capables d'engendrer leurs propres chemins d'enquête et leurs styles de raisonnement critique en mettant à la juste distance les formules apprêtées par les moteurs de recherche et autres fournisseurs de data du Web.

Fortement imbriqués, ces trois plans conduisent à inventer des formes et des supports de discussion faisant jouer tour à tour différents ressorts critiques, depuis les questions liées aux espaces de calcul et de mesure jusqu'aux façons de produire du sens par la sélection de propriétés marquantes, émergentes ou récurrentes, en passant par la réflexivité sur les modes d'existence numérique. Le recul ainsi produit n'est pas anodin, car les enquêtes dans les mondes numériques sont tributaires des points de recoupement disponibles avec les formes de vie dans le monde social. Et dans tout projet de sociologie numérique, il importe de ne pas minimiser les enjeux liés aux

formes de brouillage ou de détournement des systèmes d'information et de communication, les questions éthiques liées aux dévoilements de systèmes de surveillance de masse ou encore l'impact des différentes formes de déconnexion que peuvent pratiquer les acteurs. Il est donc primordial de ne pas naturaliser ou banaliser des dispositifs et des usages, et de créer une *autre scène* sur laquelle peut se déployer pleinement le raisonnement critique nécessaire aux sciences sociales et constitutif des formes de vie démocratique (Rouvroy et Berns, 2013).

RÉSUMÉ

En prenant appui sur l'expérience inédite du rapprochement de deux logiques d'enquête conçues comme antinomiques dans leurs raisonnements épistémologiques en sciences sociales, cet article explore les modalités d'articulation d'une approche mathématisée des grands réseaux, calculés à partir de flux du Web, et d'une approche socioinformatique des controverses, dont les corpus retracent des jeux d'acteurs et d'arguments évoluant au fil de processus critiques. Il s'agit d'apprendre à lire et interpréter dynamiquement, à l'aide de médiations numériques pointant sur les deux espaces de raisonnement, la manière dont se déplacent, sous l'impact d'événements, de décisions ou de conflits, des nœuds de réseaux saisis dans de vastes ensembles documentaires. L'agencement cognitif collectif qui émerge de ce croisement peut créer de nouvelles prises critiques sur la manière dont les processus sociaux se déploient dans les mondes numériques, des sites officiels aux médias sociaux.

Mots clés : Analyse de réseaux, sociologie, controverses, algorithmes, corpus, processus complexes.

ABSTRACT

Large evolutionary network analysis and the pragmatic sociology of controversies. How to combine methods in front of the transformations of digital worlds

Based on an unprecedented experience of bringing together two logics of research, held to be contradictory according to epistemological norms in the social sciences, this article explores how to articulate a mathematized approach to large networks, calculated from web flows, and a socio-informatics approach of controversies, evolving through non-linear critical processes. Using new digital mediations, the authors suggest some pathways for a learning interface between human interprets and algorithms. The collective cognitive device that emerges from this cross-fertilization can create a new critical approach on how social processes unfold in digital worlds, from official sites to social media.

Key words : socio-semantic networks analysis, sociology, controversies, complex processes, text-mining, macroscopes

RESUMEN

Apoyándose en la experiencia inédita de la aproximación de dos lógicas de investigación concebidas como antinómicas en sus razonamientos epistemológicos en ciencias sociales, este artículo explora las modalidades de articulación de un enfoque matematizado de las grandes redes, calculadas a partir de los flujos de la Web y de un enfoque socioinformático de

las controversias, cuyo corpus recoge los juegos de actores y argumentos que evolucionan a lo largo de procesos críticos. Se trata de aprender a leer e interpretar de forma dinámica, con ayuda de mediaciones digitales que apuntan a los dos espacios de razonamiento, la forma como se desplazan, bajo el impacto de eventos, de decisiones o de conflictos, los nodos de las redes tomadas en vastos conjuntos documentales. El diseño cognitivo colectivo que surge de este cruce puede crear nuevas posiciones críticas acerca de la forma como los procesos sociales se despliegan en los mundos digitales, de los sitios oficiales a los medios sociales.

Palabras clave: Análisis de redes, sociología, controversias, algoritmos, corpus, procesos complejos.

BIBLIOGRAPHIE

Abbott, A. (2001a), *Chaos of Disciplines*, Chicago, University of Chicago Press.

Abbott, A. (2001b), « On the concept of turning point », in *Time matters. On theory and Methods,* Chicago, The University of Chicago Press.

Barberá, P., J.T. Jost J. Nagler, J.A. Tucker et R. Bonneau (2015), « Tweeting from Left to Right : Is Online Political Communication More Than an Echo Chamber ? », *Psychological Science*, vol. 26, n° 10, p. 1531-1542.

Berthelot, J.-M. (1990), *L'intelligence du social. Le pluralisme explicatif en sociologie*, Paris, PUF.

Bessin, M., C. Bidart et M. Grossetti (dir.) (2010), *Bifurcations. Les sciences sociales face aux ruptures et à l'évènement*, Paris, La Découverte.

Börner, K. (2015), *Atlas of knowledge : anyone can map*, Cambridge, MIT Press.

Börner, K. (2010), *Atlas of science : visualizing what we know*, Cambridge, MIT Press.

Bourgine, P., D. Chavalarias et E. Perrier (2009), *French Roadmap for Complex Systems.* 2008-2009, Paris, Edition ISC, RNSC, IXXI.

Bouvier, A. (2009), « Les Conditions de la cumulativité de la sociologie », *in* Walliser B. (dir), *La Cumulativité du savoir en sciences sociales*, Paris, Éditions de l'EHESS, *Enquêtes*, 8, p. 277-325.

Broca, S. (2016), « Épistémologie du code et imaginaire des 'SHS 2.0' », *Variations* [En ligne], 19 | 2016, URL : http://variations.revues.org/701, consulté en ligne le 15 mai 2017.

Brunton, F. et H. Nissenbaum (2015), *Obfuscation : A User's Guide for Privacy and Protest*, Cambridge, MIT Press.

Caïra, O. (2011), *Définir la fiction : du roman au jeu d'échecs*, Paris, Ed EHESS.

Chateauraynaud, F. (2003), *Prospéro. Une technologie littéraire pour les sciences hum*aines, Paris, CNRS Editions.

Chateauraynaud, F. (2012), « Un visiteur du soir bien singulier », *SocioInformatique et Argumentation*, http://socioargu.hypotheses.org, consulté en ligne le 30 juin 2017.

Chateauraynaud, F. (2014), « Trajectoires argumentatives et constellations discursives. Exploration socio-informatique des futurs du nanomonde », *Réseaux*, n° 188, p. 121-158.

Chateauraynaud, F. et J. Debaz (2017), *Aux bords de l'irréversible. Sociologie pragmatique des transformations*, Paris, Pétra.

Chauviré, C. (2004), « Aux sources de la théorie de l'enquête. La logique de l'abduction chez Peirce », *La croyance et l'enquête, coll. Raisons pratiques*, 15, p. 55-84.

Chavalarias, D. (2016), *Reconstruction et modélisation des dynamiques sociales et de l'évolution culturelle : Le tournant des Sciences Humaines et Sociales du XXIe siècle*, Ordinateur et société [cs.CY], mémoire HDR, EHESS-Paris, 2016. <tel-01394843>

Chavalarias, D., J.-P. Cointet (2013), « Phylomemetic patterns in science evolution – the rise and fall of scientific fields », *PloS one*, vol. 8, n° 2, p.e. 54847.

Chavalarias, D., N. Gaumont et M. Panahi (2017), « Sur Twitter, lepénistes prosélytes et hamonistes idéalistes », *Libération,* Tribune du 20 avril 2017.

Cléro, J.-P. (2004), *Les Raisons de la fiction : les philosophes et les mathématiques*, Paris, Armand Colin.

Cui, W. *et al.* (2011), « TextFlow : Towards Better Understanding of Evolving Topics in Text », in : *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, p. 2412-2421.

Daston, L. et P. Galison (2007), *Objectivity*, New York, Zone Books.

Deffuant, G. *et al.* (2015), « Visions de la complexité. Le démon de Laplace dans tous ses états », *Natures Sciences Sociétés*, vol. 23, n° 1 p. 42-53.

Delahaye, J.P. (2009), *Complexité aléatoire et complexité organisée*, Versailles, Quæ Editions.

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

Gaumont, N., Panahi, M., 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 13, e0201879. https://doi.org/10.1371/journal.pone.0201879

Hayer, M. (1995), *The Politics of Environmental Discourse*, Oxford, Oxford University Press.

Jones, S.E. (2006), *Against Technology. From the Luddites to neo-Luddism*, Londres, Routledge.

Keller, R. (2013), *Doing discourse research*, London, England, Sage.

Kitchin, R. (2014), « Big Data, new epistemologies and paradigm shifts », *Big Data & Society,* vol. 1, n° 1, p. 1-12.

Kucher, K. et A. Kerren (2015), « Text visualization techniques : Taxonomy, visual survey, and community insights », IEEE, p. 117—121.

Lazer, D., A. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutnamm, T. Jebara, G. King, M. Macy, D. Roy et M. Van Alstyne (2009), « Life in the network : the coming age of computational social science », *Science*, vol. 323, n° 5915, p. 721-723.

Leskovec, J., L. Backstrom, J. Kleinberg (2009), « Meme-tracking and the dynamics of the news cycle », in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining,* ACM Press, p. 497-506. https://doi.org/10.1145/1557019.1557077

Mercier, H. et D. Sperber (2011), « Why do humans reason ? Arguments for an argumentative theory », *Behavioral and Brain Sciences*, vol. 34, n° 2, p. 57-74.

Palla, G., A.-L. Barabási, et T. Vicsek (2007), « Quantifying social group evolution », *Nature, Vol.* 446, p. 664-667.

Passeron, J.-C. (1991), *Le raisonnement sociologique. L'espace non-poppérien de l'argumentation*, Paris, Nathan.

Roth C. (2013), « Socio-semantic frameworks », *Advances in Complex Systems,* vol. 16, n° 4-5, p. 1350013.

Rouvroy, A. et T. Berns (2013), « Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation ? », *Réseaux*, vol. 1, n° 177, p. 163-196.

Shahaf, D. *et al.* (2013), « Information cartography : creating zoomable, large-scale maps of information », in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, p. 1097—1105.

Tarde G. (1890), *Les lois de l'imitation*, Paris, Alcan.

Van Eemeren, F. et B. Garssen (2008), *Controversy and Confrontation : Relating Controversy Analysis with Argumentation Theory*, Amsterdam, John Benjamins.

Venturini, T., D. Cardon et J.P. Cointet (2014), « Méthodes digitales. Approches quali/quanti des données numériques », Présentation du volume, *Réseaux*, vol. 188, n° 6, p. 9-21.

Venturini, T. (2012), « Great expectations : méthodes quali-quantitatives et analyse des réseaux sociaux », in Fourmentraux J.-P. (dir.), *L'Ère Post-Media. Humanités digitales et Cultures numériques*, Paris, Hermann, vol. 104, p. 39-51.

Vossen, P., R. Agerri, I. Aldabe, A. Cybulska, M. Van Erp, A. Fokkens, E. Laparra, A.L. Minard, A.P. Aprosio, G. Rigau, et M. Rospocher (2016), « Newsreader : Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news », *Knowledge-Based Systems,* vol. 110, p. 60-85.