

Traces numériques et territoires

Marta Severo, Alberto Romele



Partie 1

La trace, les méthodes et les données

Les méthodes d'interface

La renégociation des liens entre recherche numérique, STS et sociologie

Noortje Marres, Carolin Gerlitz

INTRODUCTION

John Law et ses co-auteurs [Law *et al.*, 2011] soutiennent que les médias numériques sociaux rendent possibles de nouveaux modes d'organisation, mais aussi d'analyse, de la vie sociale. Les applications technologiques grand public de partage et d'échange social en réseau, comme Facebook et Twitter, représentent des espaces nouveaux où des *données sociales* sont produites, collectées et analysées [Beer et Burrows, 2007]. Cependant, leur notoriété croissante dans la vie sociale entraîne aussi des conséquences *méthodologiques* pour la recherche sociologique [Savage, 2009 ; Rogers, 2013]. Pour ceux qui cherchent à tirer un parti empirique des plates-formes des médias sociaux, il devient rapidement évident que celles-ci ne nous offrent pas des données brutes, mais plutôt des informations mises en forme : les modes d'organisation des données sur les médias sociaux tendent à favoriser grandement certaines méthodes d'analyse fortement spécifiques, par exemple une enquête sur les «réseaux» personnels, ou bien la «popularité» de certains mots à certaines époques [Marres et Weltevrede, 2013 ; voir aussi Gitelman, 2013]. Les médias sociaux, en offrant la capacité à des acteurs sociaux de collaborer dans des activités prédéfinies, rendent les activités de ces derniers analysables. Par ailleurs, grâce à des formes de données standardisées, ces médias sociaux offrent une fonction de médiation sociale des activités, qui se prêtent à certaines formes d'analyse et pas à d'autres.

Certains des «biais méthodologiques» des médias sociaux sont assez évidents, comme c'est le cas avec Facebook qui facilite explicitement des actions, comme se faire «des amis» ou signaler que l'on «aime» telle ou telle chose (les *like*), et non l'inimitié et la critique, ce qui, de ce fait, favorise une analyse des réseaux d'un type particulier et très nettement partielle [Gerlitz et Helmond, 2012 ; Bucher, 2012]. Toutefois, dans le présent article, nous aimerions aborder des formes plus ambiguës et plus fluides de «biais méthodologiques» dans la recherche sur les médias sociaux et traiter éventuellement leurs plus larges implications en recherche pour les études sociales. La recherche sur les médias sociaux pourrait mettre en cause les relations entre les participants à la recherche, dont celles entre sujets et objets de recherche, ainsi que les hiérarchies supposées des acteurs et actants en recherche sociale. Par exemple, lorsqu'une analyse de réseau est menée avec Facebook, est-ce vraiment le chercheur qui dans ce cas décide d'employer cette méthode, ou n'est-ce pas plutôt la décision qui est informée par l'objet d'étude avec les outils et les métriques qui lui sont associés ?

Qui plus est, dans la mesure où ils soulèvent pareilles questions, les médias sociaux nous incitent à étudier plus en détail les similitudes et différences entre les méthodes sociales, qui sont pour ainsi dire intrinsèquement «implantées techniquement» dans ces médias sociaux en ligne, et nos «propres» méthodes sociologiques [Beer, 2012; Beer et Burrows, 2007; Marres, 2012]. Ainsi, pour aller dans le même ordre d'idée que l'exemple que nous avons employé plus haut: quelle comparaison se dégage entre les méthodes, les mesures et les techniques de l'analyse de réseaux mobilisées par une entreprise technologique à finalités commerciales et nos méthodes «académique» pour analyser les réseaux dits «sociotechniques» dans les travaux de recherche sociale et culturelle [voir aussi sur ce point Langlois et Elmer, 2012]? Pareil questionnement pourrait passer pour servir à confirmer une distinction entre «eux» et «nous», c'est-à-dire entre une recherche «à but lucratif» et une autre «à but non lucratif», entre des formes d'analyse de données et de savoirs appliquées et universitaires. Cependant, comme nous l'avons exposé dans d'autres travaux, les pratiques en matière de médias sociaux peuvent s'entendre de manière égale comme un lieu où les divisions du travail bien connues en sciences sociales sont utilement exposées au questionnement [Marres, 2012; Rogers, 2009; Jirotko *et al.*, 2013; Wouters *et al.*, 2012].

Dans le présent article, nous voudrions prendre en charge ce questionnement de manière particulière. Ainsi que nous le proposons ici, la recherche qui utilise les médias sociaux se focalise sur une variété de *ressemblances et d'affinités méthodologiques*, qui rendent encore plus difficile la constitution de frontières nettes à l'intérieur comme au-dessus du champ de la recherche sociologique. Cet article a pour objectif d'identifier les tensions et les opportunités qu'ouvrent ces résonances méthodologiques, et tente de les investir tout en militant pour un engagement positif envers le caractère multiforme de ces méthodes et outils pour la recherche sociologique, à savoir: le fait que les provenances, les buts et les intérêts de ces méthodes puissent être multiples et impossibles à réduire. Donc, dans cette contribution, nous nous intéressons à la fois aux *similitudes et aux différences* entre les outils grand public en ligne qui s'offrent à l'analyse de données et certaines des techniques et méthodes qui sont actuellement pratiquées en sciences sociales [Beer, 2012; T. Venturini, 2010]. Pour étudier ces résonances, et déterminer comment elles peuvent être investies pour les rendre productives en recherche sociale, nous nous concentrerons sur une méthode et une approche particulière s'appliquant à l'étude de la dynamique de l'information, pratiquée tant dans le cadre de la recherche sociologique que dans d'autres formes de l'analyse de données en ligne: la cartographie des enjeux (ou *issue mapping*).

La cartographie des enjeux peut avoir une définition assez minimale: l'utilisation de techniques informatiques dans le but de repérer, analyser et visualiser des mouvements publics de contestation sur des affaires d'actualité [Marres, 2015; Marres

et Rogers, 2005]³. On peut qualifier cette approche comme étant intrinsèquement disciplinaire, étant donné qu'elle a été reprise, qu'elle a fait l'objet de prolongements et de développements dans les sciences sociales, en informatique, ainsi que dans des domaines plus appliqués : la recherche engagée, le journalisme et le design numériques. La montée en puissance des médias sociaux numériques a donné une nouvelle impulsion aux travaux sur la cartographie des enjeux, et l'emprise élargie de Facebook et de Twitter par différentes pratiques sociales et publiques a été accompagnée d'un éventail grandissant d'outils logiciels faciles d'emploi destinés à l'analyse et la visualisation de données en ligne, qui peuvent être améliorés pour produire des cartographies d'enjeux [voir aussi sur ce point <http://demoscience.org>; Gerlitz et Helmond, 2012]. Ils comprennent des outils gratuitement accessibles destinés à l'analyse et la visualisation en direct des données, comme entre autres Infomous⁴ et MentionMapp⁵. Infomous détecte l'activité associée à des termes spécifiques liés à une affaire contenue dans des flux médiatiques sélectionnés qu'il visualise sous la forme de réseaux de bulles dynamiques et interactives, regroupant en cluster les mots étroitement liés. Mentionmapp, application également gratuite, visualise quels sont les utilisateurs associés à des hashtags précis. Trendmaps trace une carte de géolocalisation des hashtags les plus fréquemment utilisés. Une des caractéristiques de ces outils d'exploitations de données tient à ce qu'ils s'appuient sur des traditions analytiques différentes et les « croisent ». Ainsi, d'un côté, ils mettent en œuvre des mesures d'analyse de texte par ordinateur et, de l'autre, les mesures qu'ils appliquent ne sont pas sans ressemblance avec les techniques de cartographie des enjeux que la recherche socioculturelle élabore depuis les années 1980.

Plus loin, nous examinerons plus en détail ces outils ainsi que les techniques et méthodes qu'ils mettent en œuvre, et ce dans le but de clarifier une problématique relativement plus large : que faire de ces résonances méthodologiques entre les procédés d'analyse numérique et les méthodes sociologiques ? Notre but doit-il être de lever les ambiguïtés méthodologiques qu'elles suscitent et de faire la différence entre, d'une part, les utilisations journalistiques et commerciales, dans la pratique quotidienne et par les gouvernements et, d'autre part, les applications sociologiques de ces outils ? Ou bien ces résonances elles-mêmes et leurs prétendues affinités recèlent-elles quelque chose de productif ? Il y a quelque pertinence dans cette dernière question, nous semble-t-il, car les outils d'exploitation de données qui sont généralement disponibles sont de plus en plus utilisés dans les travaux de recherche empiriques, et ceci nous porte à nous demander quelle est la spécificité de leur application dans la recherche sociologique numérique [Beer, 2012]. Nous émettrons

3 On trouvera un aperçu présentant de récentes études de cas et des modes opératoires pour réaliser des cartographies des enjeux sur le wiki <http://www.issuemapping.net> que nous avons créé dans le cadre du projet *Demonstrating the Relevance of Issue Mapping for Participatory Research* (Démontrer la pertinence de la cartographie des enjeux pour la recherche participative) financé par l'Economic and Social Research Council (Conseil pour la recherche en sciences économiques et sociales).

4 <http://infomous.com>.

5 <http://mentionmapp.com>.

la proposition qu'il existe des avantages décisifs à promouvoir l'ambivalence signalée plus haut qui veut que les procédés d'analyse numérique soient *à la fois similaires et différents* des méthodes sociologiques. Notre proposition s'articule autour de l'idée centrale que les procédés numériques font appel à un troublant mystère méthodologique (*methodological uncanny*) du point de vue de la recherche sociologique : les outils que nous avons cités comportent de fortes similitudes avec les techniques et méthodes déployées en sciences sociales, mais nous ne pouvons certainement pas les considérer comme étant « nôtres ». Nous allons soutenir l'argumentation que ce troublant mystère méthodologique ouvre la voie à une approche différente en recherche sociale et culturelle dans le développement de méthodes, celles que nous appelons les « méthodes d'interface ». Nous souhaiterions esquisser cette démarche ici en exposant un projet d'élaboration méthodologique appliqué à la cartographie d'enjeux en ligne, projet où nous avons employé une technique d'analyse textuelle informatisée, une analyse dénommée « analyse des cooccurrences », utilisée pour cartographier des enjeux à l'aide de Twitter.

L'ANALYSE DU « CONTENU ÉVENTIF »

Les récents débats autour des implications d'un recours au numérique en sciences sociales se sont concentrés sur les méthodes, et des sociologues de différents horizons ont souligné les affinités méthodologiques de diverses natures entre les méthodes numériques, répondant à une définition large, et la recherche sociologique [Beer et Burrows, 2007 ; Latour *et al.*, 2012 ; Ruppert *et al.*, 2013 ; Rogers, 2013]. Certains de ces auteurs ont signalé que les méthodes actuellement intégrées dans des outils professionnels et grand public d'analyse numérique ne sont pas sans ressembler aux méthodes d'analyse textuelle et de réseaux sur lesquelles les sociologues académiques se sont longtemps appuyés [Beer, 2012 ; Marres, 2012]. Une fois confrontés à ces affinités apparentes, certains sociologues ont mis l'accent sur les *divergences* entre les buts et objectifs des procédés analytiques numériques par rapport à la recherche sociologique classique [Savage et Burrows, 2007]. Au moment où le débat sur la « destinée » des méthodes sociologiques a cours, des discussions plus précises encore sont à mener sur des méthodes spécifiques, là où culture numérique et sociologie se chevauchent, comme c'est le cas de la cartographie des enjeux.

Par le terme de « cartographie des enjeux », nous renvoyons à un ensemble de pratiques interdisciplinaires qui ont été élaborées aux frontières des études des sciences et techniques, de la sociologie, du journalisme, de l'activisme, des sciences politiques et de la visualisation de l'information, et qui déploient des méthodes informatiques pour détecter, analyser et visualiser l'actualité [Rogers et Marres, 2000 ; Venturini, 2010]. En recherche sociale et en sciences politiques appliquées, l'approche adoptée repose sur des pratiques de longue date de « cartographies

de débats», par lesquelles des controverses ou «affaires d'actualité» sont communiquées à des publics sous la forme d'une représentation graphique des acteurs, des camps en présence et de leurs positions discursives (on trouvera des discussions sur le blog Emaps⁶). En sciences politiques et sociales, la cartographie des enjeux repose davantage sur l'analyse des controverses dans l'espace public : une approche a été élaborée à la fois en sociologie politique [Chateauraynaud, 2009 ; Beck et Kropp, 2011] et dans le domaine des études sociales des sciences et techniques [Latour, 2005 ; Callon *et al.*, 2001 ; Collins et Pinch, 1998]. Dans ce dernier domaine, les méthodes de cartographie des enjeux ont été ressenties comme rendant possible une approche empirique et processuelle apte à étudier la société, les techniques et la science dans leurs rapports avec la nature [pour un exposé sur ce point, voir Marres, 2015]. Ce que ces débats nous indiquent, c'est qu'en retraçant empiriquement la genèse et l'évolution de sujets de controverses publiques, comme le changement climatique, les aliments génétiquement modifiés ou les antennes de téléphonie mobile, nous pourrions nous faire une idée sur la manière dont des entités sociales, scientifiques, techniques et environnementales se trouvent imbriquées dans la pratique. Au cours des dix dernières années, on a adopté comme instrument essentiel les méthodes numériques pour pousser plus loin sur les plans méthodologiques et intellectuels les projets d'analyse de controverse et de cartographie des enjeux, comme dans le cas des travaux pour l'atelier sur la vie sociale des enjeux, ou *Social Life of Issues* [Marres et Rogers, 2000, 2005 ; voir aussi Eklof et Mager, 2013] et pour le projet de cartographie des controverses *Mapping Controversies* [Latour, 2008 ; Venturini, 2010], ainsi que dans les recherches dans le domaine de l'«analyse des enjeux» [Thelwall *et al.*, 2006]. Et une bonne partie de ces travaux numériques ont trouvé leur point de départ déclaré dans ce qui est perçu comme des *affinités spéciales* entre les procédés d'analyse numérique et les méthodes sociologiques.

Des chercheurs de spécialités différentes en STS, analyse informatique et sociologie politique ont attiré l'attention sur une convergence particulière entre méthodes sociologiques et méthodes numériques : l'analyse des enjeux ou le «contenu évenitif» (*happening content*). Peut-être de manière la plus visible, Bruno Latour et ses collègues ont affirmé une forte ressemblance entre les ontologies des médias numériques et celles de la théorie de l'acteur-réseau, une approche qui d'après Bruno Latour permet aux sociologues d'utiliser les controverses, en les approchant comme occasions pour décrire la composition de la société [Latour, 2005]. Dans de récents articles, ils ont avancé que la numérisation permet une *généralisation* des méthodes de la théorie de l'acteur-réseau [Latour *et al.*, 2012]. Comme dans le principe associationniste qui veut que les entités puissent se définir par leurs relations à d'autres entités, les mesures de l'analyse de réseau figurent au centre de cet argument par lequel Latour et ses collègues situent la convergence entre plates-formes numériques et la théorie de l'acteur-réseau (ANT) : d'après

6 <http://www.emapsproject.com/blog>

eux, le principe associationniste est central à la fois pour l'architecture des plates-formes numériques et pour l'ANT. À partir de là, Latour *et al.* déclarent qu'une ontologie hétérogène et dynamique – une hypothèse centrale pour la théorie de l'acteur-réseau tout comme pour l'analyse des controverses – fait désormais partie intégrante de la « câblerie » de la Toile [Latour *et al.*, 2012].

D'autres travaux qui cherchent à élaborer la cartographie des enjeux sous forme de méthode numérique, auxquels nous avons participé pour certains, ont également postulé qu'il y a une affinité entre méthodes numériques et sociales ; toutefois, ces similarités générales que l'on perçoit offrent l'occasion d'établir des *différences* plus spécifiques et pertinentes entre les méthodes sociologiques et les méthodes numériques pour l'analyse du contenu dynamique [Marres et Rogers, 2001]. C'est ainsi que les méthodes informatiques d'analyse de citation et d'analyse du discours qui ont été exploitées à partir des années 1980 en STS et scientométrie présentent de remarquables similitudes avec les méthodes d'analyse d'hyperliens et de linguistique computationnelle qui jouent un rôle central dans le fonctionnement de plates-formes numériques telles que Google et Twitter [Scharnhorst et Wouters, 2006 ; Brin et Page, 1998]. Cependant, comme nous l'avons signalé dans notre introduction, dans le même temps, il *n'est pas* très difficile de différencier les mesures intégrées dans les plates-formes des grands médias numériques et les méthodes sociologiques apparentées appliquées à l'étude des réseaux et du discours. Par exemple, des outils comme Mentiomapp sont spécialisés dans la détection des acteurs *influençants* sur Twitter à un moment donné, alors que les méthodes d'analyse de contenu développées en STS ont été élaborées précisément pour détecter des enjeux qui *n'avaient pas* (encore) atteint une certaine popularité et dont la *pertinence* émergeait à peine [Callon *et al.*, 1983]. De plus, des outils comme Infomous et Mentiomapp sont faits essentiellement pour détecter les termes qui sont les plus *activement* employés *dans l'instant présent*, tandis que la recherche sociologique s'est concentrée sur l'analyse longitudinale, notamment sur la façon dont les termes et les catégories qui organisent le savoir changent avec le temps [Bowker et Star, 2000 ; Uprichard, 2011]. Pareilles différences entre la « logique » des grandes plates-formes numériques et les méthodes de la recherche sociologique ont tout autant inspiré les premiers travaux de l'un d'entre nous sur l'analyse en ligne des « réseaux d'enjeux » [Marres et Rogers, 2000]. Alors que les grands moteurs de recherche (à cette époque) définissaient la pertinence d'une source comme étant le *nombre de sources faisant autorité* contenant un lien vers elle – ce que nous avons décrit à la manière d'une « contemplation d'étoiles » – l'analyse des réseaux d'enjeux privilégiait des *associations thématiques* entre sources sur la Toile – offrant ainsi une mesure de pertinence spécifiquement thématique ou propre à un enjeu. Si nous généralisons quelque peu, nous pourrions dire que les procédés d'analyse implantés dans les plates-formes numériques ont tendance à soutenir des conceptions de la réputation centrées sur l'acteur, alors que la STS ont pensé des pertinences centrées sur l'enjeu, et leur opposent une autre perspective [Marres et Rogers, 2008].

Toutefois, si reconnaître ces différences entre les méthodes numériques et celles de la STS est important, cela pose certains problèmes en tant que point de départ pour la recherche sociologique numérique. En quelques mots, nous risquons pour finir d'essentialiser les différences entre «méthodes du médium» et, entre guillemets, *nos* méthodes de recherche, selon des modalités ne permettant pas d'apprécier suffisamment la faculté d'appropriation et l'instabilité des méthodes numériques et sociologiques. C'est pourquoi nous voulons dans cet article étudier une approche tierce par rapport aux méthodes ressemblantes présentes sur les plates-formes numériques et par rapport à celles de la recherche sociologique, une approche qui ne se fige pas tant sur une découverte des similitudes et différences des méthodes numériques et sociologiques, mais qui consiste à reconnaître et affirmer *l'instabilité* relative et l'indétermination des techniques numériques de la recherche sociologique. Il n'est *pas* toujours clair, avant que l'on se lance dans un projet de recherche numérique, *quels peuvent être* les buts analytiques auxquels les outils numériques peuvent servir, avec quels objectifs de recherche ils sont en phase ou quelles sont les finalités disciplinaires qu'ils incarnent. Par le passé, l'un d'entre nous a qualifié les outils de recherche sociologique d'«instruments variés» (*multifarious instruments*), c'est-à-dire qui présentent, potentiellement, la possibilité d'être employés pour de multiples objectifs, ce dont il n'est pas toujours facile de prendre conscience [Marres, 2012]. Harvey, Reeves et Ruppert [Harvey *et al.*, 2013] proposent que les dispositifs sociotechniques ne doivent pas être perçus comme de simples gadgets, mais compris comme des «assemblages complexes et instables qui rassemblent une diversité de personnes, d'objets et de concepts poursuivant des fins, des buts et des objectifs particuliers» [voir aussi le manuscrit de Van der Velden]. En conséquence, les outils numériques pourraient mettre en capacité, ou «activer» une variété d'acteurs et actants différents dans des configurations différentes, et ils pourraient servir à mettre en œuvre un éventail de formes de recherche différentes, ainsi que des projets politiques et éthiques.

Les outils d'analyse du contenu en ligne, nous semble-t-il, pourraient bien tomber dans cette catégorie d'instruments variés. S'il est certes possible d'identifier des similitudes et des différences spécifiques entre ces outils et les techniques de recherche sociologique, cela ne veut pas dire que nous puissions décider – une bonne fois pour toutes – que leurs buts analytiques s'alignent ou pas. Cela va dépendre également du contexte d'utilisation, et de leur déploiement particulier. Sous réserve de quelques ajustements, il est possible d'adapter une approche combinée comme celle de Mentionmapp pour pouvoir mobiliser une analyse longitudinale [Uprichard, 2011]. En effet, une bonne partie du débat sur les méthodes numériques en recherche sociologique sur les médias a mis l'accent sur la possibilité de cibler les dispositifs numériques [Rogers, 2009]. Des sociologues ont attiré l'attention sur le fait que les méthodes de recherche numériques sont elles-mêmes instables et sous-déterminées, proposant des notions comme les méthodes plastiques [Lury, 2012], en temps réel, ou *live* [Back et Puwar, 2012]. La variabilité des fins, en outre,

est une caractéristique qui peut être assignée aux plates-formes de médias sociaux eux-mêmes, car les configurations de ces plates-formes changent fréquemment et cherchent à satisfaire des acteurs divers, devant s'interfacer et négocier avec une multiplicité d'intérêts émanant de groupes d'utilisateurs différents y compris les publicistes, des acteurs de tierce maintenance et les développeurs de logiciels [Langlois et Elmer, 2013]. Les plates-formes de médias sociaux tout comme les méthodes peuvent dès lors être caractérisées comme étant « multivalentes » : elles peuvent être utilisées pour une multiplicité de fins analytiques et normatives qui ne sont pas nécessairement transparentes, ni devant être réalisées en même temps, ni être accessibles à tous les acteurs concernés [Gerlitz, 2012 ; Marres, 2011]. Dans le présent article, nous souhaiterions pousser plus loin ce débat sur le dynamisme des médias et des méthodes numériques en émettant la proposition qu'il permet une approche distincte pour *élaborer une méthodologie* en sociologie numérique, que nous appelons « méthodes d'interface ». À ce stade, nous définirions les méthodes d'interface comme étant des méthodes émergentes, que nous comme chercheurs en sciences de la société et de la culture ne pouvons précisément dire nôtres, mais qui rentrent suffisamment en résonance avec nos intérêts et avec les démarches qui nous sont familières pour offrir un espace productif d'engagement empirique avec des contextes élargis, des pratiques, et des appareils⁷ techniques de recherche. Toutefois, plutôt que nous lancer dans un développement sur une définition générale des méthodes d'interface, nous souhaiterions en rendre compte par des moyens empiriques en décrivant un projet de recherche numérique dans lequel nous avons appliqué d'un point de vue critique et créatif une méthode d'interface précise : l'analyse de cooccurrences.

LA COOCCURRENCE EN TANT QUE MÉTHODE D'INTERFACE

Revenons aux outils en ligne d'analyse de données cités plus haut et abordons leur « troublant mystère méthodologique » (*methodological uncanny*) plus en détail. Les outils en ligne pour l'analyse de données en temps réel comme Infomous, proposons-nous, accomplissent des tâches grosso modo similaires à certaines études connues en sociologie de la connaissance et de l'innovation : ils cherchent à détecter ce que sont les enjeux actifs (*happening issues*) en analysant les relations changeantes entre mots dans le contenu en direct telles qu'elles apparaissent sur des plates-formes médiatiques particulières. Donnons un autre exemple : Twitter Streamgraph est un autre outil logiciel gratuit disponible en ligne qui permet l'analyse et la visualisation des données Twitter, soi-disant en temps réel. Si l'on

7 Ce terme tire en partie son origine des travaux de Lucy Suchman [2005], de Celia Lury [2004] et d'Alexander Galloway [2012], lesquels ont tous attiré l'attention sur les interfaces en tant qu'espaces cruciaux où se négocient les divisions épistémiques du travail. Formant un entre deux, les interfaces rendent possibles une dynamique de la multivalence, où des outils, des données et des méthodes peuvent être reliées de différentes manières, et ce pour autoriser différents objectifs analytiques et normatifs.

tape « climate change » (changement climatique), l'outil Streamgraph va nous donner une courbe de l'activité associée à ce terme sur Twitter sur une période de temps donnée, littéralement visualisée sous la forme d'un flux, qui montre les mots-clés avec leur fréquence qui sont associés au terme recherché « *climate change* » (voir Figure 1).

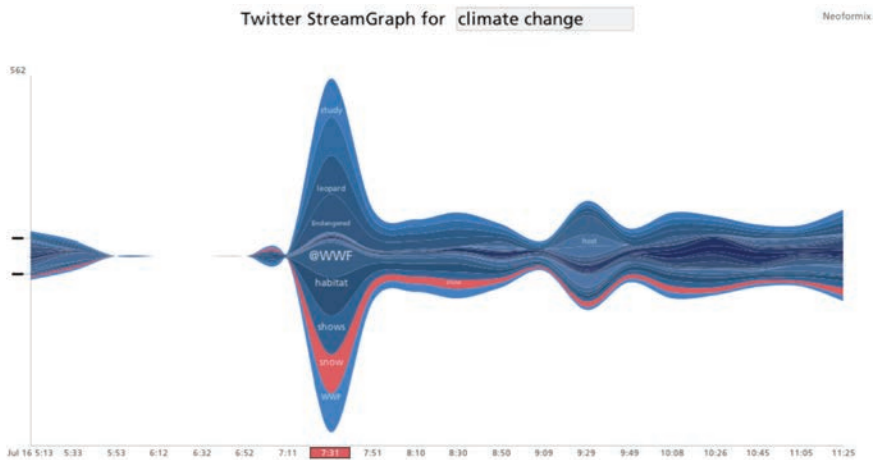


Figure 1 : Courbe de Twitter Streamgraph pour le mot-clé *climate change*

Twitter Streamgraph, à l'instar d'Infomous, mesure ensuite la *cooccurrence* des termes en temps réel, détectant ceux qui ressortent le plus en même temps dans un flux de médias sélectionné, ici Twitter, et montrant comment les relations entre termes évoluent avec le temps. Dans Infomous, lorsqu'un plus grand nombre de termes est relié de manière significative, un cluster prend forme et dans Streamgraph le flux s'élargit quand plusieurs mots surgissent simultanément. En appliquant cette mesure de la cooccurrence pour détecter « ce qui se passe » ces outils ressemblent à une méthode dont la STS s'est fait la championne, et plus particulièrement la théorie de l'acteur-réseau, à savoir *l'analyse de cooccurrence de termes (co-word analysis)*. Cette méthode a été élaborée dans les années 1980 par le sociologue des sciences, technologies et société Michel Callon et ses collègues, pour détecter les sujets émergents et innovants dans la littérature scientifique [Callon *et al.*, 1983 ; voir aussi Danowski, 2009 ; Marres, 2012]. L'analyse de cooccurrence de termes a été élaborée à partir d'une mesure d'analyse textuelle plus simple, « cooccurrence », qui détecte les mots qui apparaissent conjointement (habituellement dans la même phrase) à l'intérieur d'un texte donné. La « co-word » détecte les couples de mots – c'est-à-dire les mots qui apparaissent dans le voisinage des uns et des autres – et affectent une valeur à la relation entre ces mots qui s'appuie sur la distance qui les sépare, disons une distance de 3, 4 ou 5 mots. Elle procède ensuite à une pondération des relations entre termes en leur attribuant des valeurs exprimant à la fois la proximité de la cooccurrence et le nombre de relations [Callon *et al.*, 1983 ; Danowski, 2009]. Callon et ses collègues se sont faits les promoteurs de cette

méthode, moyen de détecter l'apparition de sujets émergents, également appelée «poches d'innovation» (en anglais *pockets of innovation*) dans un corpus de textes. L'analyse de cooccurrence de termes, si l'on en croit les arguments développés à l'époque, rend possible la détection de changements dans des associations de sujets dans le temps *sans être obligé de s'appuyer sur des catégories préalablement définies, et sur leurs critères implicites de pertinence*. Elle y parvenait en adaptant les textes à l'analyse de réseau, où ce faisant les associations de mots apparaissant empiriquement parmi des mots figurant dans un ensemble de données déterminé, produisaient un critère immanent de pertinence. (On peut aussi retrouver cette préoccupation de tracer ces associations permettant une forme immanente d'analyse par la suite dans les développements de la théorie de l'acteur-réseau).

Dans un certain sens, en appliquant des mesures de cooccurrence pour analyser du contenu en temps réel, les outils en ligne d'analyse de données comme Infomous ou Streamgraph offrent une méthode qui ne diffère guère de la méthode par cooccurrence de termes développée dans les années 1980 : dans les deux cas, le but est de détecter des «sujets occurrents» en analysant les relations entre termes et leurs changements dans le temps [voir aussi Marres et Weltevrede, 2013]. Toutefois, à regarder de plus près les mesures mises en œuvre dans les outils en ligne, les similitudes qu'elles présentent avec les méthodes sociologiques commencent à paraître plus discutables. En effet, les outils de mesure de données en ligne ne traitent que la fréquence de mots cooccurrents et non la force de leur relation. Cette spécialisation sur la *fréquence* n'est pas comparable à l'analyse de cooccurrence de termes, qui détecte également *les variations dans la force* des relations entre mots (en fonction de leur proximité spatiale). L'analyse de cooccurrence n'est pas identique à l'analyse de cooccurrence de termes. On peut, certes, faire une distinction nette entre les outils d'analyse de données en ligne et les méthodes de recherche sociologique, les mettant chacun à sa propre place. Toutefois, nous voulons souligner le fait que la résonance entre méthodes sociologiques et techniques d'analyse numérique ouvre une zone d'indétermination, où nous pouvons réinterroger les caractéristiques et les capacités de ces méthodes sur un nouveau terrain. Les méthodes d'analyse numérique nous offrent l'occasion de repenser le statut de nos «propres» méthodes qui ont été développées en recherche sociologique et en STS, à savoir : celui de l'analyse de cooccurrence de termes.

A première vue, celle-ci peut paraître comme une méthode sociologique robuste. Comme déjà indiqué, l'analyse de cooccurrence de termes incorpore une forme rudimentaire du principe qui sera plus tard au cœur de ce que l'on appellera la «théorie de l'acteur-réseau» qui s'applique à discerner les relations changeantes entre entités hétérogènes [Latour, 2005]. Toutefois, ce type d'analyse n'est pas, après tout, la «propriété» de la sociologie ou encore de la STS. Une étude plus large de la littérature sur cette méthode informatique montre qu'elle a été aujourd'hui reprise dans des domaines aussi divers que la scientométrie, l'ingénierie logicielle ou les sciences de la communication [Coulter *et al.*, 1998 ;

Danowski, 2009]. Cette littérature sur l'analyse de cooccurrence de termes tend à établir que cette méthode proviendrait des sciences de l'innovation ou de la scientométrie. Pour l'heure, ce qui nous paraît d'une importance capitale, c'est que ce qui se passe pour nos propres méthodes se rattache moins solidement à nos traditions savantes habituelles. Les outils d'analyse de données en ligne invoquent une trouble méthodologique : *elles entrent en résonance avec des méthodes que nous connaissons, mais elles nous rappellent également qu'elles ne sont pas les nôtres*. Les outils d'analyse de données mettent en lumière la polyvalence des méthodes sociales.

Toutefois, lorsque l'on compare les techniques numériques et les méthodes sociales, il ne suffit pas de prendre en compte les caractéristiques formelles des mesures analytiques seulement. Si l'on veut déterminer les capacités offertes par ces techniques, il nous faut aussi examiner leurs contextes d'application, les différentes questions, préoccupations et aspirations auxquelles ils font appel et les modes de valorisation qu'ils permettent. Comme le discutent Marres et Weltevrede [2013], les outils d'analyse de données en ligne comme Infomous et Streamgraph s'appliquent dans un contexte très spécifique : ils s'inscrivent dans une infrastructure en constante évolution qui instaure la Toile en temps réel, et plus largement des « cultures de l'actualisation » (*update cultures*), où il est constamment nécessaire d'être informé « de ce qui se passe, là, maintenant ». Les outils sont impliqués dans la valorisation du contenu et de son « instantanéité » : prenant part à une économie numérique plus large investie dans des analyses en temps réel [Back et Puwar, 2012], ces outils attirent l'attention sur des sujets qui surgissent « dans l'instant ». Dans ce cas, la valeur se situe dans la détection de variations instantanées au moment même de l'événement. Le contexte analytique où l'on a voulu faire intervenir l'analyse de cooccurrence de termes dans les années 1980 n'était pas le même : cette méthode avait été conçue pour identifier ce que Callon et ses collègues appelaient la « dynamique de la problématisation » ou les « poches d'innovation » [Callon *et al.*, 1983]. Plutôt que de se concentrer sur les tendances que l'on peut dénombrer instantanément, le but était de détecter l'émergence de problématiques de recherche et de sujets à l'intersection des catégories disciplinaires, autrement dit des innovations échappant à la détection dans les formes alors prédominantes d'analyse scientométrique, comme l'analyse de citations, qui selon Callon *et al.* tendaient à respecter la classification disciplinaire de la recherche scientifique : les sujets qui apparaissaient aux frontières entre plusieurs disciplines n'étaient pas prêts d'être repérés comme particulièrement pertinents tant que des cadres disciplinaires fournissaient le schéma de référence pour valoriser l'innovation, comme c'était le cas avec l'analyse de citations.

Ces différences de contexte peuvent servir à rendre compte des mises en œuvre divergentes de l'analyse de cooccurrence entre l'analyse numérique et la sociologie de l'innovation classique : la fréquence de cooccurrence semble constituer un indicateur adéquat si l'on a pour but de découvrir ce qui est en train de faire l'actualité, alors que la détection des « zones de problématisation » nécessite une identification à la granularité plus fine des termes qui entrent plus exactement en relation suivant

des modalités nouvelles dans des domaines donnés de la recherche scientifique. Toutefois, pareilles restitutions contextuelles des techniques numériques suggèrent aussi que ces techniques ne sont pas nécessairement « transparentes ». Alors que des outils comme Infomous et Twitter Streamgraph sont relativement simples d'un certain point de vue (en effet, lors de la présentation des présents travaux, des statisticiens dans l'assistance ont fait remarquer à plusieurs reprises que ces mesures étaient « basiques »⁸), ils font en même temps partie d'un complexe de pratiques analytiques en émergence qui, elles, ne sont pas nécessairement simples. Dans leur application particulière de l'analyse de cooccurrences, les outils d'analyse de données en ligne dotent les pratiques de détection de ce qui fait l'actualité et, en cela, ils participent de la valorisation du « temps réel » (*liveness*) [Lury, 2012; Lury et Wakeford, 2012]: quel est le mot qui apparaît le plus souvent dans les tweets, quel est l'article de presse qui reçoit le plus de liens ou qui est le plus relayé de tweet en tweet, quel est l'acteur le plus souvent cité? Cette orientation se traduit par une économie politique de l'information distincte, où la valeur est le résultat d'une propagation rapide et momentanée avec laquelle des termes se diffusent dans un présent immédiat [Bucher, 2012; Back et Puwar, 2012; Hansen *et al.*, 2011]. De ce fait, il semble naïf de définir les outils d'analyse et de visualisation de données en ligne purement par rapport aux mesures qu'ils appliquent. Le contexte de cette mise en œuvre détermine au moins aussi fortement la « méthode » que les mesures appliquées.

Dans ce cas, nous n'avons pas affaire à des mesures « dans leur nudité »: les outils d'analyse de données façonnent la pratique analytique. Les utilisateurs sont-ils en mesure de délimiter les jeux de données ou est-ce que c'est l'outil qui le fait pour eux? Lorsque nous recherchons des mots-clés dans un flux médiatique donné, dans quelle mesure cela limite les questions que nous pouvons nous poser [boyd et Crawford, 2012]? Les outils d'analyse de données participent activement à la configuration des pratiques analytiques, ce qui à son tour permet de concrétiser des régimes de valorisation. *Pour cette raison, les méthodes numériques peuvent être qualifiées de méthodes intéressées* ou « épaisses » (en anglais *thick*): elles constituent des propositions qui suggèrent des manières particulières de doter, organiser et valoriser les pratiques et les connaissances. Alors que les mesures intégrées dans les outils d'analyse de données en ligne sont, dans un certain sens, bel et bien « minces » (en anglais *thin*), l'appareil sociotechnique qu'ils permettent (la détection de ce qui est d'actualité – et gratuitement!) est, pour ainsi dire, bien plus « épais », car il intègre l'analyse de données en direct dans les pratiques numériques, et en tant que tel permet de mettre en place des sociétés de l'information qui sont orientées vers le temps réel. C'est pourquoi nous aimerions penser la cooccurrence, ou du moins sa mise en œuvre dans les outils d'analyse de données en ligne, comme étant une méthode fortement intéressée⁹.

8 A nos yeux, ceci a mis en lumière de manière intéressante les obstacles qui compliquent la prise en compte, la négociation et la remise en cause d'hypothèse ontologique de recherche numérique : ces hypothèses sont considérées comme trop « basiques » pour justifier une interrogation détaillée.

9 Le terme de « méthode intéressée » nous a été suggéré par Kristin Asdal lors d'une communication personnelle.

Ces premiers examens de l'analyse de cooccurrence de termes employée comme méthode et comme technique d'étude du «contenu évenitif» (happening content) soulèvent une question particulièrement sensible : pouvons-nous envisager un appareil numérique destiné à l'analyse de «contenu évenitif» qui promet d'autres objectifs que ceux poursuivis par l'analyse numérique orientée vers le moment en cours ? Pouvons-nous imaginer une mise en œuvre dans la recherche sociale numérique qui permettrait une détection de la pertinence ? Nous avons dit que l'analyse numérique implique bien plus que la seule mise en œuvre de mesures analytiques, mais celles-ci nous apparaissent néanmoins comme un site particulièrement fructueux pour l'exploration de configurations alternatives de l'analyse du «contenu évenitif». Nous aimerions, dès lors, aborder les mesures numériques comme ouvrant des espaces d'engagement critiques et créatifs avec un appareil analytique plus large en devenir. Nous aimerions traiter les méthodes en ligne d'analyse de cooccurrence de termes comme des interfaces. Aborder l'analyse de cooccurrence de termes comme étant une interface se traduit en un cahier des charges particulier pour le développement des méthodes sociales en ligne. En s'appuyant sur les préoccupations qui ont informé le développement de l'analyse de cooccurrence de termes, la question que nous posons est la suivante : est-il possible de déployer une analyse de cooccurrence de termes pour détecter pas seulement ce qui se passe en temps réel, mais ce qui prend vie, pour mettre en lumière non pas les termes qui deviennent courants, mais ceux qui sont en train de s'activer, ou leur vivacité (ou liveliness) [Marres et Weltevrede, 2013] ? Nous en dirons davantage sur ce qui distingue ce qui est courant de ce qui prend vie, mais nous voudrions souligner le fait que ce cahier des charges de l'expérimentation méthodologique se forme à partir de notre engagement et de nos préoccupations sur le contexte et les buts de l'analyse numérique, et des intérêts et des objectifs qui informent l'élaboration à plus large échelle de l'appareil sociotechnique de la recherche numérique.

Le fait d'aborder l'analyse de cooccurrence de termes comme une méthode d'interface consiste alors à traiter notre mise en œuvre de cette méthode sociale informatisée comme une occasion de s'engager, de manière critique et créative, avec de méthodes similaires qui sont déployées à une échelle plus large de la culture numérique. Dans ce cas, les méthodes d'interface s'inspirent, tout en divergeant, de l'approche des méthodes numériques définies par Rogers [Rogers, 2013] : si cette approche a bel et bien pour but de réorienter les méthodes «nativement numériques», nous cherchons pour notre part à nous engager, dans une manière à la fois critique et créative, avec la remédiation des méthodes sociologiques dans la culture numérique. Autrement dit, nous voulons souligner le caractère incertain, la multiplicité de provenances des méthodes sociales et cultures numériques, dans les deux cas au moment où elles trouvent leur application dans la culture numérique et dans la recherche universitaire sociologique et culturelle. Nous voulons profiter

de cette *ambivalence méthodologique* à des fins de recherche sociologique¹⁰. Cependant, avant d'aller plus loin dans la description de notre tactique méthodologique, nous souhaitons présenter certains des premiers résultats de notre tentative de mettre en application le *cahier des charges* évoqué plus haut et de pratiquer en ligne l'analyse de cooccurrence de termes.

LE CHANGEMENT CLIMATIQUE SUR TWITTER

Dans notre expérimentation méthodologique, donc, nous cherchons à mettre en œuvre l'analyse de cooccurrence de termes comme méthode analytique appliquée au «contenu évenitif» (*happening content*) en ligne. De la sorte, nous nous engageons dans une utilisation, à la fois critique et créative, des outils et des mesures appliquées à l'analyse de contenu en ligne qui sont disponibles sur la Toile et qui concernent principalement le phénomène d'actualité. Nous tentons de dépasser une opérationnalisation, couramment répandue, mais restrictive, du concept de «contenu évenitif». Nous voudrions élaborer une technique permettant de détecter la vivacité (*liveliness*) plutôt que l'actualité, c'est-à-dire déterminer non quels termes deviennent populaires mais ceux qui *s'activent*, vus sous l'angle des relations changeantes qu'ils nouent avec d'autres termes dans le temps. Bien sûr, notre ambition est précaire, car la dynamique des enjeux se conçoit depuis longtemps comme «l'ascension et le déclin» de mots-clés dans les contenu des médias [Downs, 1972], et il serait absurde de notre part de penser que cette «dimension verticale» de l'activité des enjeux (des augmentations et des diminutions de leur fréquence de mention) pourrait d'une certaine façon perdre en pertinence. Ce que nous recherchons, donc, c'est de formuler des mesures pour la détection de dynamiques d'enjeux sur les médias en ligne qui *combinent* des figures «proportionnelles» de pensée et de mesure à d'autres mesures «associationnistes», ce qui définit la pertinence en s'appuyant sur la formation de relations entre termes précédemment non reliées.

Pour développer et tester notre technique, nous décidons de nous en tenir aux données de Twitter. Ces dernières sont relativement faciles à obtenir en passant les APIs de Twitter. Pour capter et analyser les tweets, nous avons utilisé TCAT (Twitter Capture and Analysis Toolkit) développé par le *Digital Methods Group* de l'Université d'Amsterdam¹¹. Notre but étant d'appliquer une analyse de cooccurrence de termes en recherche numérique, nous nous concentrons sur les mots, mots-clés et hashtags,

10 Notons qu'il est également possible de faire une analyse dite de «cooccurrence de termes» avec des données en ligne sans trop se préoccuper de l'application couramment répandue d'outils similaires en ligne. En d'autres termes, on peut également «numériser» cette méthode et, par exemple, les en-têtes de blogs de la même manière que l'on analysait les résumés d'articles scientifiques par le passé [Hellsten et Elefteria, 2012].

11 DMI-TCAT est développé par Erik Borra et Bernhard Rieder. Les scripts de capture de flux sont disponibles sur <https://github.com/bernorieder/twitterstreamcapture>.

plutôt que d'autres objets Twitter, tels que les URLs et les noms d'utilisateurs. Les données Twitter conviennent pour une analyse de cooccurrence de termes, parce que le tweet limité à 140 caractères offre une unité de données exploitable où il est possible de détecter des relations de cooccurrence. Comme indiqué par Callon *et al.* [Callon *et al.*, 1983], l'analyse de cooccurrence de termes fonctionne particulièrement bien sur des fragments relativement courts d'informations mises en forme, comme les résumés d'articles scientifiques, ce type d'information étant déjà structuré pour démontrer une pertinence. Toutefois, notre première décision de nous appuyer sur cette démarcation classique de l'unité de données – le tweet – où nous pourrions déterminer les relations de cooccurrence n'est pas sans conséquence. Par exemple, cela veut dire que nous écartons, au moins au départ, la possibilité de spécifier en outre la «force» de la relation de cooccurrence à l'intérieur de chaque tweet, comme l'avaient initialement formulé Callon *et al.*

Pour faire notre analyse de «contenu évenitif» (*happening content*), nous décidons de nous concentrer sur un terme relevant d'un enjeu assez général, à savoir le «changement climatique» (ici, *climate change*), et d'inclure dans notre ensemble de données tous les tweets utilisant ce terme sur une période de près de trois mois, du 1er mars au 15 juin 2012, colligeant 204 795 messages de ce type au total, soit un ensemble de données d'une taille largement exploitable. Nous avons mis l'accent sur ce thème très général parce qu'il nous était familier à la suite de travaux précédents sur la cartographie d'enjeux en ligne [Rogers et Marres, 2000 ; Niederer, 2009]. Lors de différents ateliers à Londres et à Amsterdam¹², nous avons effectué des études pilotes avec divers mots-clés d'une granularité plus fine (comme le style de vie durable et l'action climatique: *sustainable living* et *climate action*) et à différents intervalles, mais il semble qu'il faille un thème d'un enjeu relativement large et un nombre suffisant d'intervalles de temps pour capter des variations thématiques que l'on puisse interpréter. Une fois délimité notre ensemble de données, nous nous sommes posé la question: sur quels termes nous concentrer? Notre discussion sur les différents critères de sélection initiaux possibles démontre à quel point la conception proportionnelle de la pertinence est enracinée dans les pratiques médiatiques en ligne: pour sélectionner nos «mots-cibles», nous nous sommes d'abord portés sur les mots-clés des listes de 5, 10 et 20 premiers obtenus par analyse de fréquence. Nous avons ensuite décidé de tester une mesure alternative, à savoir l'analyse de cooccurrence, en nous interrogeant sur la différence que cela engendrerait si au lieu d'étudier *avec quelle fréquence* un mot est cité dans Twitter, nous nous intéressions à *comment il est relié* à d'autres mots. Avant de répondre à cette question, cependant, nous devons nous attaquer à un problème tout aussi épineux: nous concentrons-nous sur les mots ou les hashtags (les mots précédés d'un signe dièse [#])? Notre but plus général étant de procéder à des analyses

12 Ceux-ci comprenaient les ateliers *Issue Mapping Online* et *Co-Word Machine* qui se sont tenus à Goldsmiths College à l'Université de Londres en mai 2012, tous deux financés par le programme Digital Social Research Programme du ESRC, ainsi que la Digital Methods Summer School organisée à l'Université d'Amsterdam en juillet 2012.

de contenu ou d'enjeux, il serait sûrement préférable de s'intéresser au véritable contenu, c'est-à-dire les mots, plutôt que les hashtags, qui fonctionnent comme des mots-clés et qui sont généralement utilisés pour identifier des sujets, démarquer des conversations et rendre recherchables les tweets portant un hashtag [Bruns et Stieglitz, 2012]. Les hashtags sont des termes généraux, voire généralisateurs, et en tant que tels ne semblent pas être les mots les plus intéressants lorsqu'on doit identifier ce qui fait sujet d'enjeu sur le plan du contenu plutôt que sur le plan d'une classification «disciplinaire» ou «publicitaire». Toutefois, des considérations touchant à la logistique informatique sont entrées en ligne de compte : exécuter une analyse de cooccurrence de mots (au lieu des hashtags) sur notre corpus sur le changement climatique aux premières étapes de notre projet aurait pris des heures. Depuis lors, ce problème a été résolu grâce à de nouveaux ordinateurs, mais à l'époque ce point a grandement simplifié notre choix : c'était sur la cooccurrence de hashtags, et non des mots, que nous allions porter notre attention.

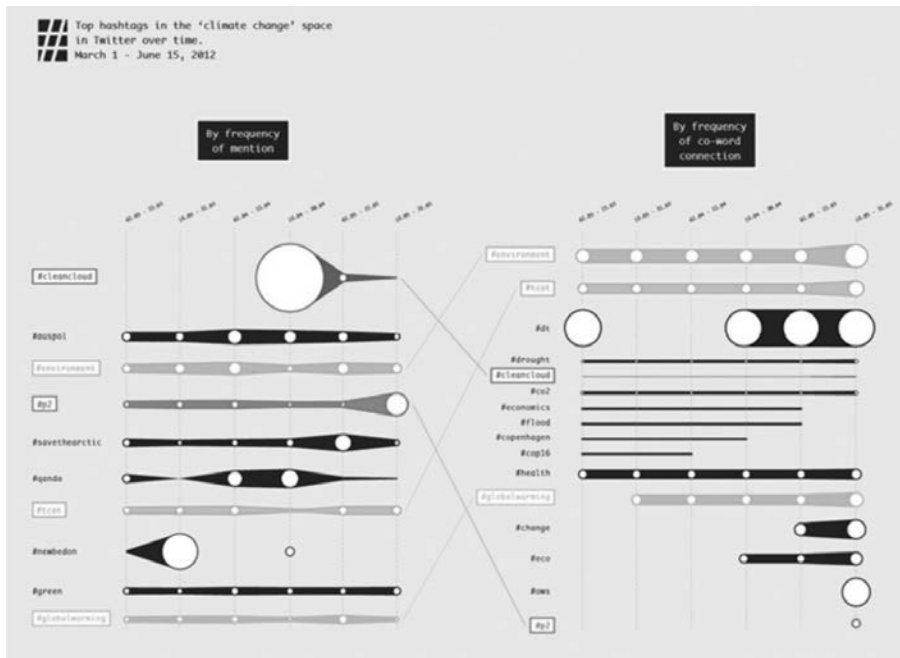


Figure 2 : Hit-parade des hashtags mesurés en fréquence de citation de termes (à gauche) et les liaisons de cooccurrence (à droite) mesurées sur 6 intervalles hebdomadaires. Échantillon : le changement climatique (climate change) sur Twitter

La Figure 2 montre le classement des hashtags mesurés en fonction de la fréquence du mot (à gauche) et selon leurs liaisons de cooccurrence de termes (à droite), et comment ceux-ci ont changé d'un intervalle à l'autre¹³. La colonne sur la gauche montre les hashtags qui sont les plus souvent cités dans notre ensemble des données

13 Pour la fréquence comme la cooccurrence, nous avons commencé par calculer le classement global des termes (tous intervalles confondus) avant de calculer l'occurrence par intervalle. Nous

pour chaque intervalle, et comment cet ensemble de hashtags les plus populaires change avec le temps. La colonne de droite montre les hashtags qui y apparaissent le plus souvent *tout en étant associés* à d'autres hashtags également présents dans notre ensemble des données pour chaque intervalle¹⁴. Ces résultats divergent très sensiblement, car les mesures de fréquence et celles de cooccurrence de termes font ressortir des résultats différents des hitparades de hashtags provenant de notre ensemble des données. Dans la colonne des fréquences, des hashtags renvoient à des institutions ou des manifestations, comme #cop16 ou #auspol, tout comme d'autres correspondent à des campagnes comme #savethearctic. Là où on applique la cooccurrence de termes, des hashtags formés de substantifs sont davantage mis en avant, dont #economics, #flood, #co2, #health. La figure rend aussi compte la différence entre l'actualité et la vivacité des enjeux. Les hashtags qui sont globalement fréquents sont davantage exposés à des dynamiques, en quelque sorte, de battage : ils apparaissent rapidement, atteignent de fortes fréquences grâce à des activités de retweet de courte durée, puis disparaissent, comme c'est le cas de #cleancloud et #newbedon. Les hashtags aux fréquences les plus élevées sont *spécifiquement internes à Twitter*, puisqu'on y trouve des termes comme #qanda (*question & answer*: «questions-réponses») or #newbedon¹⁵, bien que l'analyse de cooccurrence de termes fasse bel et bien ressortir #dt (*donated tweet*, ou «don en tweet»). Les mots qui jouissent de bonnes cooccurrences ont plus de chance de survivre d'un intervalle à l'autre, ce qui en montre l'endurance, ce sont : #environment, #tcot (topconservatives, ou archiconservateurs), et #drought (sécheresse). D'un point de vue empirique, la figure apporte, au départ, de l'eau à notre moulin pour affirmer que les mesures proportionnelles (la fréquence) sont plus susceptibles de diriger notre attention vers des dynamiques propres au médium (feu de paille ou *bursting*, battage ou *hyping*), alors que les mesures relationnelles (la liaison) permettent davantage de faire ressortir au premier plan des dynamiques propres à l'enjeu. Cela veut dire que le choix de la mesure peut orienter l'analyse davantage en direction d'une «étude de la dynamique du médium» ou vers une «étude de la dynamique de l'enjeu», bien que l'une comme l'autre feront les deux [voir sur ce point : Marres, 2015]. Toutefois, lorsqu'il s'agit de détecter la vivacité d'enjeux, cette première figure laisse une question importante sans réponse : elle donne une indication du nombre de liaisons de cooccurrence pour un hashtag donné, mais ne nous dit rien sur la variation des liaisons, c'est-à-dire les modalités d'évolutions des mots associés à ces hashtags.

avons exclu les termes de requête «*climate change*», «*climate*» et «*change*», ceux-ci n'apportant rien à notre analyse.

14 Nous avons employé précisément la mesure de du «degré moyen pondéré», c'est-à-dire le degré pondéré de liaisons entre un hashtag et d'autres hashtags.

15 Concernant newbedon, son apparition semble être indicative d'un phénomène propre à Twitter : ce terme fait référence à des personnalités de l'opposition dans les Emirats Arabes Unis qui se sont vus dépossédés de leur nationalité, et ce hashtag a été utilisé à des fins de sensibilisation. Ce hashtag doit avoir été kidnappé, peut-être par des spammers.

COMMENT DÉTECTER LA VITALITÉ ?

Pour résumer brièvement ce que nous avons établi jusqu'ici, trois points nous semblent particulièrement importants. Premièrement, la traduction opérationnelle de l'analyse de cooccurrence de termes pour analyser des données sur Twitter : celle-ci nous a certes permis d'affirmer et de problématiser la distinction entre des mesures proportionnelles (de la fréquence) et relationnelles (des cooccurrences) de la pertinence. En traitant nos données avec ces mesures, nous avons trouvé différentes façons dont l'intérêt pour la fréquence se trouve intégré aux pratiques des médias numériques. Cela ne représente pas seulement une caractéristique des outils prédominants d'analyse des données. Ainsi, l'importance des retweets nous a rappelé combien la « fréquence de la mention » est recherchée activement – et produite – dans le cadre d'une tactique publicitaire sur Twitter : retweeter est une technique pour que des messages soient repris par le système. Deuxième point : dans l'opérationnalisation de l'analyse de cooccurrence de termes pour la recherche en ligne, notre propre démarche est influencée par son contexte d'application. En appliquant concrètement en ligne l'analyse de cooccurrence de termes, notre démarche a fini par intégrer certains des « présupposés du médium ». En effet, en adoptant le tweet comme une unité significative d'analyse, nous avons aussi opté pour la mesure de la cooccurrence plutôt que pour l'analyse de la cooccurrence de mots. Notre outil de cooccurrence de terme définit une relation entre « co-hashtags » comme étant l'occurrence de deux hashtags dans le même tweet, et cette approche nous dispense efficacement de mesurer la relation de force relative des liens comme le préconise Callon – la proximité relative entre mots – une caractéristique primordiale de l'analyse de cooccurrence de termes. Toutefois, en même temps, en abordant l'appareil de recherche Twitter, nous avons acquis une vision plus claire de nos propres objectifs et intérêts méthodologiques. Nous avons perçu clairement que si nous voulons détecter la vivacité d'enjeux grâce à Twitter, nous devons élaborer une heuristique supplémentaire pour pouvoir repérer quelles sont les mutations pertinentes parmi les associations fluctuantes de termes.

L'analyse de cooccurrence de termes présente un problème qui tient au fait qu'il s'agit d'une mesure « expansive » : en traitant des données textuelles, la technique détecte de plus en plus de relations entre termes, et de modifications dans ces relations entre termes, mais sans offrir nécessairement un moyen explicite de différenciation entre ces changements et d'en estimer l'importance¹⁶. Pour commencer à traiter ces problèmes, nous avons développé l'heuristique du « profil d'associations » (ou *associational profile*). Cette heuristique est inspiré par l'idée d'un réseau-acteur, selon laquelle une entité peut se définir au travers des entités qui lui sont associées [Callon *et al.*, 1986 ; Latour, 2005]. Le profil d'associations d'un hashtag est constitué des autres entités auxquelles il est relié (par voie de cooccurrence). Nous pouvons envisager que différentes entités forment le profil

16 La mesure de la « force du lien » nous a permis de traiter ce problème.

d'un hashtag, autrement dit les autres hashtags, mais aussi les URLs, auxquels il est associé, ainsi que les utilisateurs¹⁷. Notre notion de profil d'associations reconnaît ainsi les ontologies propres au médium, contrairement à Latour *et al.* [Latour *et al.*, 2012]. De plus, nous voulons utiliser cette heuristique du profil d'associations pour détecter les modifications *dans le temps* des associations concernées : notre recherche ne porte pas tant sur le côté ontologique de ce que composent les associations (le profil) d'une entité donnée à un instant donné. Nous nous intéressons davantage aux changements du profil de la composition d'un hashtag comme indice de la variabilité d'un enjeu, ou son degré d'animation.

Pour examiner l'utilité de cette heuristique, nous avons décidé de produire trois types de profils pour certains des hashtags les plus importants que nous avons préalablement identifiés dans notre ensemble de données Twitter sur le changement climatique : les profils d'acteurs, d'utilisateurs et de hashtags. En accomplissant cet exercice, nous nous sommes concentrés au départ sur les hashtags les plus explicitement politiques qu'il contient, à savoir : #ows (occupez Wall Street) et #tcot (top conservateurs), en présupposant que des hashtags aussi polarisés offriraient le plus de chances de présenter des profils distinctifs et notre analyse a bel et bien révélé des différences explicites entre ces hashtags. Nous avons déterminé les profils d'acteurs en rapport avec nos hashtags en identifiant quels étaient les URLs figurant au premier plan des tweets contenant le hashtag en question, et nous avons découvert que #ows est, pour la plupart, associé à des sites politiques ou d'organisation et à des médias sociaux et de diffusion d'actualités en ligne¹⁸. Par contre, #tcot est associé principalement à des médias de diffusions d'actualité et généralistes, et à des intervalles ultérieurs de plus en plus à des blogs et au site de l'organisation progressiste <http://thingprogress.com>. Le profil des hashtags globalement progressifs #p2 est dominé par des organisations, plutôt que des sites politiques, <http://thinkprogress.org> étant la source la plus fréquemment citée, mais renvoie aussi fortement vers les sites d'actualité et médiatiques. Dans l'ensemble, il semble exister des différences particulièrement significatives entre les profils d'acteur de #ows et de #tcot : #tcot est essentiellement associé à des préoccupations d'actualité, #ows à une démarche de campagne d'opinion et #tcot apparaît comme étant plus diversifié que #ows au regard des sources citées¹⁹. Dans une deuxième

17 Selon le même principe qui veut que le tweet fournit le site et les moyens de liaison, on peut dire : si deux entités sont reliées au même tweet, elles sont alors reliées.

18 Ressortit principalement <http://majority.fm> et <http://savetheartic.org>, un site de lanceurs d'alerte et un site de campagne.

19 Nous tirons cette explication des données : une lecture attentive des tweets associés à ces deux hashtags fait ressortir que, au moins dans l'échantillon de données liées au changement climatique, un nombre significatif de tweets contenant #tcot faisait des commentaires sur le conservatisme. Prenons un commentaire comme celui-ci, par @DukeMaximum : « Ouah, cette affaire de changement climatique s'est changée en un clin d'œil d'une enquête scientifique légitime en culte de fin du monde, non ? [*Wow the climate change field went from legitimate scientific inquiry to a doomsday cult in the blink of an eye didn't it?*] #tcot ». Ces tweets n'avaient pas toujours une orientation conservatrice

phase de profilage, nous avons examiné les associations entre utilisateurs et *hashtags*. Cette opération s'est avérée intéressante pour différentes raisons, notamment parce qu'elle a mis en relief l'hétérogénéité des entités impliquées dans le changement climatique sur Twitter²⁰. Lorsque nous avons travaillé sur ce sujet pendant la *Digital Methods Summer School* d'Amsterdam, un certain nombre de participants a décidé d'examiner quels étaient les types d'utilisateurs que l'on trouve dans notre ensemble de données Twitter, et ils se sont rapidement entendus sur la différence entre utilisateurs humains et non humains (les *bots*) comme constituant la distinction la plus importante. Comme l'exposent Niederer et Van Dijk [2010], les robots (les *bots*) jouent un rôle important dans l'organisation du discours public présent sur les plates-formes de médias sociaux. Ce point, toutefois, ne fait pas que soulever des questions d'identification (sur Internet, peut-on vraiment savoir qui est quoi?) dans le même ordre d'idée que le test de Turing. La question pertinente est tout autant celle-ci : à quel point les *bots* exercent-ils une influence sur l'organisation du discours public sur le changement climatique sur Twitter? Cette deuxième interrogation diffère de la première. En effet, dans le deuxième cas, ici, elle ne porte pas sur la différenciation entre discours humain et non humain, ou sur la possibilité d'en faire une. Le but est plutôt de détecter les propriétés du discours public que la plate-forme facilite dans la mesure où cela implique une diversité d'entités à la fois humaines et non humaines, autrement dit, dans la mesure où cela implique une certaine composition de *bots* et d'humains [voir aussi sur ce point Wilkie *et al.*, 2015]. Toutefois, pour évaluer l'inflexion du discours sur le changement climatique sur Twitter due aux *bots*, il reste nécessaire d'identifier ces utilisateurs non humains, et les méthodes utilisées au cours du projet d'école d'été pour différencier ces deux types d'acteurs ont été sans conteste expérimentales (voir Figure 3). Les utilisateurs ont été sélectionnés en utilisant une mesure de fréquence (ils n'étaient que 60 utilisateurs à écrire 100 tweets et plus), et ils ont été catégorisés à la main entre utilisateurs humains et *bots* en fonction d'une description de profil, de l'activité en tweets et des liens utilisés²¹. Une première série de résultats nous a permis de déterminer

en eux-mêmes, mais proposaient des commentaires sur le conservatisme où une préoccupation avec l'actualité s'explique. De fait, #tcot est décrit comme un hashtag pour suivre les ultraconservateurs sur Twitter (<http://www.topconservativesontwitter.net> [consulté le 16 juillet 2013]). À l'inverse, il semble que les contributions contenant #ows sont davantage susceptibles de suivre une orientation militante en soi.

20 Notons que les méthodes de profilage ci-dessus instaurent chacune un niveau élevé d'homogénéité dans les données, ainsi dans un premier cas nous nous sommes concentrées sur les relations entre les hashtags uniquement, puis, dans le second cas, entre hashtags et URLs contenues dans les tweets, toutes les autres entités ayant été délibérément ignorées [pour une discussion, voir Marres, 2012].

21 Les tweets créés par des *bots* suivent souvent un modèle très régulier, affichent une structure de tweet semblable, et utilisent de façon répétée une chaîne de spam ou la même URL. Nous nous sommes aussi demandé si les humains ou les *bots* présentent souvent un thème de prédilection issu de la biographie de leur profil ou du contenu du tweet, et même chez les *bots* il est possible de trouver des descriptions ciblées thématiquement.

qu'il existe un grand nombre de *bots* génériques, c'est-à-dire des *bots* qui publient du contenu qui n'est pas particulièrement spécifique au hashtag utilisé (*climate change*). Il semble que ces *bots* s'avèrent profiter de la relative actualité des hashtags pour trouver des publics, qu'importe le contexte sur le fond. La présence de *bots* peut donc être prise en compte de manière très indirecte comme un indice d'activité autour d'un enjeu en rapport avec le changement climatique sur Twitter. Leur présence, en effet, suggère que le hashtag en question est d'actualité, mais dans la mesure où ces *bots* sont génériques, l'actualité est une caractéristique qui s'applique au médium, et non à l'enjeu. Il se peut que nous devions aussi faire une distinction entre l'activation qui est plus spécifique à un médium (associations changeantes entre hashtags et *bots*), par opposition à une activation qui est mobilisée par variation d'associations sur le fond à l'intérieur de l'espace d'un même enjeu (comme une opposition entre organisations et informations d'actualité).

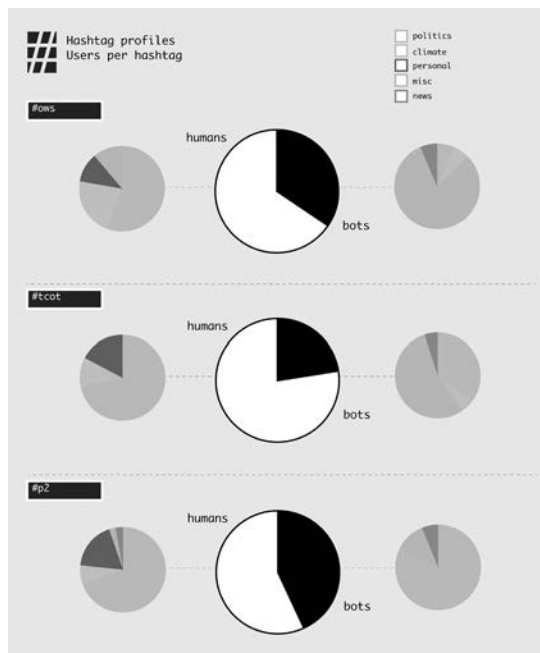


Figure 3: Comparaison humains et bots : profils d'utilisateurs pour les hashtags #ows, #tcot et #p2 en relation avec le thème du changement climatique (mot-clé climate change)

Pour finir, nous en sommes revenus à nos questions antérieures sur les liaisons entre hashtags eux-mêmes. A quels hashtags se relie ceux que nous avons sélectionnés dans chaque intervalle? Comment ces associations de hashtags changent-elles d'un intervalle à l'autre? En collaboration avec des programmeurs et des designers, nous avons mis au point une technique pour déterminer les profils d'associations: détecter, pour un mot-cible donné, les hashtags qui lui sont associés par intervalle et créer une visualisation qui montre la variation de ces associations sur l'ensemble des intervalles (figure que nous appellerons provisoirement «courbe de vie de

hashtag»). Ce type de profil de hashtag transcrit graphiquement l'intensité avec laquelle les hashtags forment des liaisons avec d'autres hashtags dans le temps (voir la figure 4 sur le hashtag #drought, ou sécheresse). Dans ces premiers profils, nous avons distingué les nouvelles liaisons entre hashtags (les termes en noir) et celles qui se sont arrêtées (les termes en rouge), puisque nous conjecturons que des changements dans les liaisons pourraient marquer des changements ou des glissements thématiques. L'emploi de couleur marque la persistance et le changement: les traits en bleu et en vert montrent des liaisons stables, alors que les traits en rouge en fin d'intervalles indiquent la disparition de certaines liaisons. Ces premières transcriptions visuelles présentent donc non seulement la croissance et le reflux d'enjeux, mais aussi les changements dans leurs associations pour faire ressortir cette interrogation: quels sont principalement les autres termes qui composent le profil du hashtag et comment changent-ils dans le temps? Le fait de se poser cette question, nous semble-t-il, nous offre un moyen de retracer la «vie» ou la «vivacité» du terme représentatif d'un enjeu. Ainsi, le profil du hashtag #drought est indicatif d'un certain nombre d'événements en relation avec ce sujet, comme l'arrivée sur le devant de la scène d'inondations dans le Yorkshire dans le premier intervalle (Figure 4). De prime abord, cela nous a paru être une anomalie, mais John Bloomfield, du *British Geological Survey*, en a souligné l'importance: au cours de l'été 2012, la survenance de crues lors d'une période qui avait été définie comme une sécheresse a entraîné bien des remises en cause institutionnelles, que ce soit sur le plan de la communication auprès du public au sujet de ces «événements météorologiques» apparemment contradictoires ou sur celui d'hypothèses établies dans la modélisation des sécheresses. Pour finir, nous avons relevé un moment de recomposition quasi totale de l'enjeu dans le profil de #drought, dans le moment où la terminologie associée au sommet se voit presque totalement remplacée par des termes spécifiques à l'Afrique dans le dernier intervalle.

Enfin, à un niveau plus général, ces transcriptions visuelles de profil peuvent être lues comme offrant la «signature» ou l'«empreinte» d'un hashtag, ces lignes de vie affichant à la fois la granularité (y a-t-il une multiplicité des hashtags qui apparaissent et qui disparaissent?) et la volatilité (à quel point les changements dans la composition des hashtags sont-ils spectaculaires?). Le hashtag #environment, par exemple, possède un profil plus diversement composé que la sécheresse (#drought), avec des divergences et des reconvergences de courbe plus nombreuses (voir la comparaison en Figure 5). En élargissant pareille comparaison, nous pouvons repérer différents types de profils, entre des hashtags émergents et des hashtags vivants. Certains présentent des liaisons décousues, d'autres plus éruptives; certains sont hétérogènes dans leurs liaisons, d'autres moins. Nous interprétons les hashtags qui font montre de changements périodiques avec associations durables et hétérogènes comme jouissant d'une vivacité particulière.

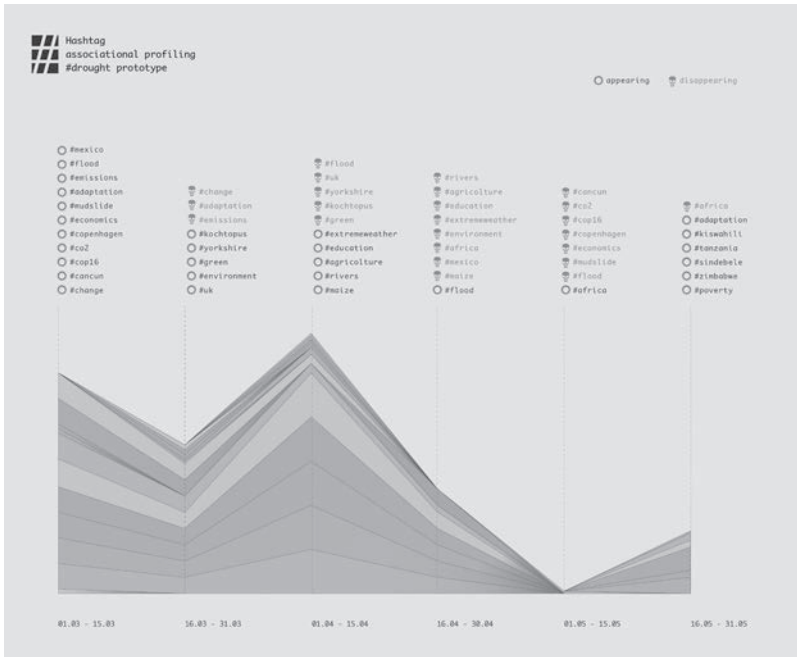


Figure 4: Profil du hashtag #drought (sécheresse)

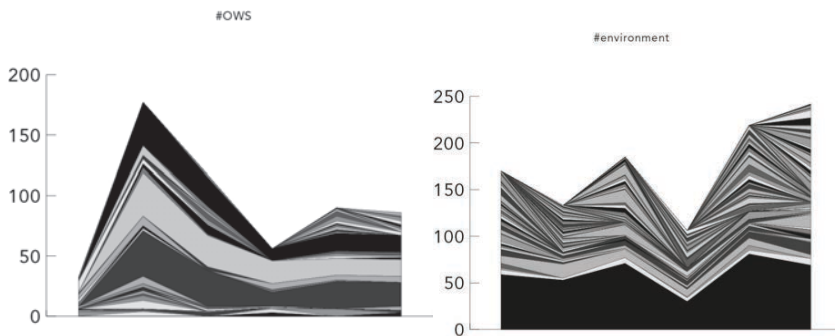


Figure 5: Profils d'associations pour #ows (à droite) par rapport à #environnement (à gauche)

CONCLUSION : PROBLÉMATIQUE DES MÉTHODES D'INTERFACE

Notre expérience de profilage par association a eu pour conséquence l'élaboration d'un outil prototype fonctionnel [Marres *et al.*, 2013], mais pour conclure le présent article nous souhaiterions aborder ce que notre expérimentation nous enseigne jusqu'ici sur les méthodes d'interface. Comme indiqué, le but initial de notre expérimentation était de nous abstenir de déterminer – ou figer artificiellement – *ce qu'est* l'analyse de cooccurrence de termes et les capacités analytiques qu'elle offre. Et nous avons pu en effet constater que le type d'analyse qui peut être produite à l'aide de l'analyse de cooccurrence de termes *dépend en partie du médium*

auquel elle est appliquée. Twitter offre un autre cadre et un autre contexte que les résumés d'articles scientifiques. Ainsi, alors que nous avons commencé nos travaux à l'aide de l'analyse de cooccurrence de termes parce qu'elle offre une méthode *relationnelle*, notre décision à travailler avec des données Twitter nous a contraintes à nous occuper de différents effets proportionnels et liés à la fréquence. Nous aimerions indiquer deux raisons expliquant cette situation. Premièrement, nous avons été constamment confrontées à des effets proportionnels dans nos données, par exemple sous la forme d'activités de retweets provoquant des salves de hashtags, ou celle des *bots* en kidnappant certains. En menant cette analyse sur Twitter, par conséquent, l'«ontologie» qui ressort des données de cette plateforme, son format spécifique et les pratiques d'usage associées ne peuvent être tout simplement ignorés par notre méthode, et, de fait, elle semble avoir été contaminée dans une certaine mesure. Les données provenant des médias sociaux se présentent sous une forme, un format spécifique et sont façonnées par des pratiques d'utilisation distinctes qui peuvent orienter l'analyse dans des directions particulières, dans notre cas celles de formes d'analyse proportionnelles.

En deuxième lieu, et avec une égale importance, par notre tentative d'adapter une démarche sociologique à l'analyse du «contenu évenitif» (*happening content*) à l'aide de Twitter, notre projet de recherche a créé le cadre d'une rencontre entre différentes approches méthodologiques, et, partant, entre différentes traditions analytiques dans la recherche sociale et culturelle. Nous en sommes venues à l'analyse de cooccurrence de termes avec en tête un intérêt marqué par la STS pour la cartographie d'enjeux et pour les méthodes relationnelles appliquées à l'étude du contenu dynamique, le repérage de «poches d'innovation» ou ce que nous avons préféré appeler «vivacité». Toutefois, en cherchant à appliquer l'analyse de cooccurrence de termes à la recherche numérique, cette méthode a fait l'objet de discussions concernant une méthode apparentée, mais légèrement différente, l'analyse des cooccurrences. S'il est vrai que les applications actuelles de cette dernière méthode sont en partie déterminées par l'analyse de cooccurrence de termes, celle-ci débouche assurément sur une traduction autre que celle que nous avons appelée par le passé «analyse des relations inter-mots». Comme indiqué, l'analyse de cooccurrences se fixait sur la fréquence de liaisons cooccurrentes, mais sans examiner les proximités relatives entre termes. Dans notre cas, le format Twitter des tweets sert d'unité de référence pour l'analyse, et non le «cluster de mots» – moins dépendant du médium – composé de mots présentant des distances entre termes variables. En tentant cette mise en œuvre, nous avons appris à connaître beaucoup mieux la méthodologie que nous proposons. On ne peut certainement pas dire que l'analyse des données de Twitter en soi se limite d'elle-même à un cadre proportionnel, et au travers de notre étude nous avons appris beaucoup au sujet des champs d'opportunités et des contraintes que les médias numériques imposent aux «méthodes associationnistes». Nous voudrions relever un des avantages essentiels des données Twitter: qu'il permet d'utiliser des hétérogénéités préstructurées pour construire des profils d'associations; celles-ci

peuvent être des URLs ou des utilisateurs, auxquels pourraient s'ajouter la langue, les appareillages et la plate-forme informatiques utilisés pour la publication en ligne, ainsi que le type de tweet (@reply, tweet, retweets, etc.). Une telle hétérogénéité préstructurée pourrait offrir un moyen formidable pour qualifier plus avant les objets de profil et leur vivacité. Nous voulons aussi simplement signaler que la délimitation de l'échantillon de données vis-à-vis du « médium dans son intégralité » – ainsi que nous l'avons fait en nous concentrant sur *tous les tweets* mentionnant le changement climatique (avec le mot-clé *climate change*) – oriente l'analyse vers la détection d'effets propres à un médium.

Enfin, et c'est peut-être l'enseignement le plus important, nous avons tiré des leçons sur l'importance de la mauvaise adéquation de la méthode par rapport au médium, et sur la nécessité de créer un certain nombre d'ajustements bien choisis pour rendre opérante une certaine part sélective d'inadéquation entre méthode et médium. Ce point pourrait aussi servir de définition ad hoc de l'analyse contestante, ou pour dire les choses plus simplement, de l'« intervention ». Comme nous avons porté notre méthode pour qu'elle « interface » avec les pratiques analytiques et techniques de la recherche sur Twitter, nous avons fini par adopter une méthode hybride. La question se pose ainsi : qu'avons-nous gagné à accueillir de tels effets d'interface ? Qu'avons-nous gagné en autorisant que des effets liés au médium déterminent notre étude sur les effets d'enjeux ? Dans nos projets empiriques, nous avons tenté de créer des relations informées entre notre méthode, les données et l'analyse de contexte, comme la sociologie de la connaissance et l'analyse numérique, la spécificité de Twitter et l'analyse de la dynamique des enjeux. S'il est vrai que nous n'avons pas réussi à tous égards à établir pareilles relations, nous avons pu bel et bien nous familiariser avec les façons plus ou moins déterminées par lesquelles celles-ci ne sont pas adaptées les unes aux autres. Les pratiques sur Twitter peuvent susciter à nouveau un intérêt pour la fréquence grâce aux pics de retweets, les bots peuvent faire revenir la dynamique du médium dans l'étude de la dynamique des enjeux. En d'autres mots, le biais proportionnel d'une bonne partie de l'analyse de Twitter n'est pas seulement un artefact dû aux tendances méthodologiques de ceux qui la pratiquent. Il y a des comportements, des pratiques et des effets spécifiques aux médias qui produisent, participent à la production, des conditions qui rendent possible la montée en notoriété de cette approche méthodologique : le fait de retweeter, les tweets déclenchés par l'événement, et les invasions de bots sont un certain nombre des phénomènes que nous avons rencontrés. Pour cette raison, il ne paraît pas exagéré que les analyses proportionnelles sur Twitter parlent le « langage du support » : elles ne sont pas seulement au fait de ce support, mais aussi bien adaptées aux pratiques de ses utilisateurs, à leurs effets et à la culture spécifique de ce médium.

Rien de surprenant donc que notre tentative de nous écarter de l'analyse proportionnelle pour aller vers des mesures différentes n'a réussi que partiellement : la tendance systémique en direction de la proportionnalité des pratiques d'utilisation l'a réintroduite dans notre analyse relationnelle. De plus,

notre propre démarche d'analyse des cooccurrences de termes ne peut se passer de mesures de fréquence. Ce qui explique pourquoi, au lieu de nous figer sur une distinction en «ou, ou» entre mesures relationnelles et mesures proportionnelles ou bien entre vivacité et actualité, c'est l'interaction des dynamiques d'enjeu et des dynamiques du médium que nous avons tenue pour pertinente.

La dynamique du support, dans notre cas celle sur Twitter, est coconstitutive du biais favorable à l'analyse proportionnelle. On ne peut toutefois mettre la dynamique du médium entre parenthèses, il faut y être attentif. La question est ouverte, mais le fait de la poser implique une autre compréhension des méthodes d'interface. Nous avons commencé par reconnaître que la question des outils et des mesures reste ouverte, mais pour élaborer des formes d'analyse à l'interface de différentes traditions méthodologiques, il peut également être important non seulement de chercher à les aligner, mais également de reconnaître un éventuel vice d'ajustement des différentes composantes constitutives de la pratique analytique. Nous concevons alors les méthodes d'interface comme une remédiation à la fois du médium, des données, des pratiques, des enjeux, des techniques et des méthodes, ce qui demande une attention à l'égard à la fois de l'alignement ou du défaut d'alignement des différents agents analytiques qui leur sont intégrés ou qu'elles rendent possibles.

REMERCIEMENTS

Nous remercions les participants de l'école d'été Digital Methods Summer School 2012 sur les dynamiques des enjeux en ligne intitulée «Issue Lifelines»: Albrecht Hofheinz, Colleen Reilly, David Moats, Diego Dacal, Esther Weltevrede, Jill Hopke, Johannes Passmann, Kalina Dancheva, Sara Kjellberg and Tally Yaacobi-Gross. Nous remercions particulièrement, Alessandro Brunetti, designer d'information. Nous tenons aussi à exprimer notre gratitude envers les participants des ateliers de cartographie d'enjeux en ligne (*Issue Mapping Online*) et d'analyse de cooccurrences automatique (*The Co-Word Machine*) à Goldsmiths College, University of London, notamment: Esther Weltevrede, Marieke van Dijk, Richard Rogers, Sabine Niederer, Anne Helmond, Bernhard Rieder, Emma Uprichard, Erik Borra, Lucy Kimbell, Astrid Mager, Sam Martin, Esteban Damiani et Ea Ryberg Due. Pour finir, nous tenons à remercier Simeon Smith pour sa lecture des données Twitter en qualité d'adjoint de recherche.

FINANCEMENTS

Cet article a été possible grâce à une subvention de démonstrateur accordée par le Conseil pour la recherche en économie et sciences sociales (Economic and Social Research Council – ESRC) au titre de son axe sur le numérique en sciences

sociales (Digital Social Research Programme) et un crédit complémentaire au titre de son programme communautaire pour 2012-2013.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Back et Puwar, 2012] Back, L. et Puwar, N., 2012, «A manifesto for live methods: provocations and capacities», *The Sociological Review*, 60, 6-17.
- [Beck et Kropp, 2011] Beck, G. et Kropp, C., 2011, «Infrastructures of risk: a mapping approach towards controversies on risks», *Journal of risk research*, 1(14), 1-1.
- [Beer, 2012] Beer, D., 2012, August 31, «Using Social Media Data Aggregators to Do Social Research», *Sociological Research Online*.
- [Bowker et Star, 2000] Bowker, G. C. et Star, S. L., 2000, *Sorting Things Out: Classification and Its Consequences*, Cambridge, MA : MIT Press.
- [Bruns et Stieglitz, 2012] Bruns, A. et Stieglitz, S., 2012, «Quantitative Approaches to Comparing Communication Patterns on Twitter», *Journal of Technology in Human Services*, 30(3-4), 160-185.
- [Bucher, 2012] Bucher, T., 2012, «Want to be on the top? Algorithmic power and the threat of invisibility on Facebook», *New Media & Society*, 14(7), 1164-1180.
- [Burrows et Beer, 2007] Burrows, R. et Beer, D., 2007, September 30, «Sociology and, of and in Web 2.0: Some Initial Considerations», *Sociological Research Online*.
- [Callon *et al.*, 1983] Callon, M., Courtial, J.-P., Turner, W. A. et Bauin, S., 1983, «From translations to problematic networks: An introduction to co-word analysis», *Social Science Information*, 22(2), 191-235.
- [Callon *et al.*, 1986] Callon, M., Law, J. et Rip, A., 1986, *Mapping the dynamics of science and technology: sociology of science in the real world*, London : Macmillan.
- [Chateauraynaud, 2009] Chateauraynaud, F., 2009, «Public controversies and the Pragmatics of Protest: Toward a Ballistics of collective action», *Working Paper EHESS*, Paris.
- [Collins et Pinch, 1998] Collins, H. M. et Pinch, T., 1998, *The golem: What you should know about science*, Cambridge : Cambridge University Press.
- [Coulter *et al.*, 1998] Coulter, N., Monarch, I. et Konda, S., 1998, «Software engineering as seen through its research literature: a study in co-word analysis», *Journal of the American Society for Information Science*, 49(13), 1206-1223.
- [boyd et Crawford, 2012] boyd, D., et Crawford, K., 2012, «Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information», *Communication & Society*, 15(5), 662-679.
- [Danowski, 2009] Danowski, J., 2009, «Inferences from word networks in messages», In Krippendorff, K. et Bock, M. A. (éd.), *The content analysis reader*, Thousand Oaks, CA : Sage, pp 421-429.
- [Downs, 1972] Downs, A., 1972, «Up and down with Ecology: the Issue Attention Cycle», *The Public Interest*, 28, 38-50.

- [Eklof et Mager, 2013] Eklof, J. et Mager, A., 2013, «Technoscientific promotion and biofuel policy: How the press and search engines stage the biofuel controversy», *Media Culture Society*, 4(35), 454-471.
- [Galloway, 2012] Galloway, A. R., 2012, *The Interface Effect*, Cambridge: Polity.
- [Gerlitz, 2012] Gerlitz, C., 2012, *Continuous Economies*, Goldsmiths, University of London: PhD thesis.
- [Gerlitz et Helmond, 2013] Gerlitz, C. et Helmond, A., 2013, «The Like economy: Social buttons and the data-intensive web», *New Media & Society*, 15(8), 1348-1365.
- [Gitelman, 2013] Gitelman, L. (éd.), 2013, *Raw Data Is an Oxymoron*, Cambridge, MA: MIT Press.
- [Hansen *et al.*, 2011] Hansen, L. K., Arvidsson, A., Nielsen, F. Å., Colleoni, E. et Etter, M., 2011, «Good friends, bad news. Affect and virality in Twitter», *Future Information Technology*, Berlin/Heidelberg: Springer, 34-43.
- [Harvey *et al.*, 2013] Harvey, P., Reeves, M. et Ruppert, E., 2013, «Anticipating Failure: Transparency devices and their effects», *Journal of Cultural Economy*, 3(6), 294-312.
- [Jirotko *et al.*, 2013] Jirotko, M., Lee, C. P. et Olson, G. M., 2013, «Supporting Scientific Collaboration: Methods, Tools and Concepts», *Computer Supported Cooperative Work (CSCW)*, 22(4-6), 667-715.
- [Langlois et Elmer, 2013] Langlois, G. et Elmer, G., 2013, «The Research Politics of Social Media Platforms», *Culture*, 14, 1-17.
- [Latour *et al.*, 2012] Latour, B., Jensen, P., Venturini, T. et Boullier, S. G. D., 2012, «The Whole is Always Smaller Than Its Parts: A Digital Test of Gabriel Tarde's Monads», *British Journal of Sociology*, 63(4), 590-615.
- [Law *et al.*, 2011] Law, J., Ruppert, E. et Savage, M., 2011, *The Double Social Life of Methods*, CRESC Working Paper Series, 44(95), 1-12.
- [Lury, 2004] Lury, C., 2004, *Brands. The logos of the global economy*, London/New York: Routledge.
- [Lury, 2012] Lury, C., 2012, «Going live: towards an amphibious sociology», *The Sociological Review*, 60(2012), 184-197.
- [Lury et Wakeford, 2012] Lury, C. et Wakeford, N., 2012, *Inventive Methods: The Happening of the Social*, London/New York: Routledge.
- [Marres *et al.*, 2013] Marres, N., Gerlitz, C., Weltevrede, E., Borra, E. et Rieder B., 2013, *Mapping WCIT with Twitter: Issue and Hashtag Profiles*, <http://issuemapping.net/Main/WCITProfiles>.
- [Marres, 2015] Marres, N., 2015, «Why Map Issues? How digitization is changing the settings of controversy analysis and what to do about it», *Science, Technology and Human Values*, <http://sth.sagepub.com/content/early/2015/03/16/0162243915574602>.
- [Marres, 2011] Marres, N., 2011, «The costs of public involvement. Everyday devices of carbon accounting and the materialisation of participation», *Economy and Society*, 40(4), 510-533.
- [Marres, 2012] Marres, N., 2012, «The redistribution of methods: on intervention in digital social research, broadly conceived», *The Sociological Review*, 60(2012), 139-165.

- [Marres et Rogers, 2005] Marres, N. et Rogers, R., 2005, «Recipe for Tracing the Fate of Issues and their Publics on the Web», In Latour, B. et Weibel P. (éd.), *Making Things Public: Atmospheres of Democracy*, Cambridge, MA : MIT Press, pp 922-935.
- [Marres et Rogers, 2008] Marres, N. et Rogers, R., 2008, «Subsuming the ground: how local realities of the Fergana Valley, the Narmada Dams and the BTC pipeline are put to use on the Web», *Economy and Society*, 37(2), 251-281.
- [Marres et Rogers, 2001] Marres, N. et Rogers, R., 2001, «Depluralising the Web and Repluralising Public Debate: The Case of the GM Food Debate on the Web», In Rogers, R. (éd.), *Preferred Placement: Knowledge Politics on the Web*, Maastricht: Jan van Eyck Academie.
- [Marres et Weltevrede, 2013] Marres, N. et Weltevrede, E., 2013, «Scraping the Social? Issues in real-time social research», *Journal of Cultural Economy*, 6(3), 313-335.
- [Niederer et Van Dijck, 2010] Niederer, S. et Van Dijck, J., 2010, «Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system», *New Media & Society*, 12(8), 1368-1387.
- [Niederer, 2009] Niederer, S., 2009, *Climate Change Sceptics. Digital Methods Initiative*, <https://wiki.digitalmethods.net/Dmi/ClimateChangeSceptics>.
- [Rogers, 2009] Rogers, R., 2009, *The End of the Virtual – Digital Methods. Media*, http://www.govcom.org/rogers_oratie.pdf.
- [Rogers, 2013] Rogers, R., 2013, *Digital Methods*, Cambridge, MA: MIT Press.
- [Rogers et Marres, 2000] Rogers, R. et Marres, N., 2000, «Landscaping climate change: a mapping technique for understanding science and technology debates on the World Wide Web», *Museum*, 9, 1-23.
- [Ruppert et al., 2013] Ruppert, E., Law, J. et Savage, M., 2013, «Reassembling Social Science Methods: The Challenge of Digital Devices», *Theory, Culture & Society*, 30(4), 22-46.
- [Savage, 2009] Savage, M., 2009, «Contemporary Sociology and the Challenge of Descriptive Assemblage», *European Journal of Social Theory*, 12(1), 155-174.
- [Scharnhorst et Wouters, 2006] Scharnhorst, A. et Wouters, P., 2006, «Webindicators: a new generation of S&T indicators», *Cybermetrics*, 10.
- [Suchman, 2005] Suchman, L., 2005, «Agencies in Technology Design: Feminist Reconfigurations», <http://www.lancs.ac.uk/fass/sociology/papers/suchman-agenciestechnodesign.pdf>.
- [Thelwall et al., 2006] Thelwall, M., Vann, K. et Fairclough, R., 2006, «Web issue analysis: An integrated water resource management case study», *Journal of the American Society for Information Science and Technology*, 57(10), 1303-1314.
- [Uprichard, 2011] Uprichard, E., 2011, «Dirty data: longitudinal classification systems», *The Sociological Review*, 59(c), 93-112.
- [Van der Velden, 2014] Van der Velden, L., 2014, «A study into third party elements on Dutch governmental websites», doctoral dissertation chapter, Media Studies, University of Amsterdam [ms].
- [Venturini, 2010] Venturini, T., 2010, «Diving in magma: how to explore controversies with actor-network theory», *Public Understanding of Science*, 19(3), 258-273.

- [Venturini, 2012] Venturini, T., 2012, «Building on faults: how to represent controversies with digital methods», *Public Understanding of Science*, 21(7), 796-812.
- [Wilkie *et al.*, 2015] Wilkie, A., Michael, M. et Plummer-Fernandez, M., 2015, «Speculative Method and Twitter: Bots, Energy and Three Conceptual Characters», *The sociological review*, 63(1), 79-101.
- [Wouters *et al.*, 2012] Wouters, P., Beaulieu, A., Scharnhorst, A. et Wyatt, S. (éd.), 2012, *Virtual Knowledge: Experimenting in the Humanities and the Social Sciences*, Cambridge, MA : MIT Press.

Table des matières

INTRODUCTION	7
<i>Marta Severo, Alberto Romele</i>	
PARTIE 1 - LA TRACE, LES MÉTHODES ET LES DONNÉES.....	11
AU-DELÀ DE LA CRITIQUE <i>BIG DATA</i>	13
<i>Richard Rogers</i>	
LES MÉTHODES D'INTERFACE	33
<i>Noortje Marres, Carolin Gerlitz</i>	
SOFT DATA	61
<i>Marta Severo, Alberto Romele</i>	
L'IDENTITÉ COMME BASE DE DONNÉES	87
<i>Jos de Mul</i>	
PARTIE 2 - RENCONTRE ENTRE TRACES NUMÉRIQUES ET TERRITOIRES	109
L'ÉCUME NUMÉRIQUE DES TERRITOIRES	111
<i>Dominique Boullier</i>	
SPATIALITÉS ALGORITHMIQUES	133
<i>Boris Beaude</i>	
L'HYPERVILLE	161
<i>Franck Cormerais</i>	
DÉSIRS DE DATA	177
<i>Maryse Carmes, Jean-Max Noyer</i>	

PARTIE 3 - LES PRATIQUES DE LA TRACE NUMÉRIQUE.....	211
DE LA TRACE À LA CARTE ET DE LA CARTE À LA TRACE	213
<i>Matthieu Noucher</i>	
DIMENSIONS SPATIALES DE L'ACTUALITÉ INTERNATIONALE.....	225
<i>Laurent Beauguitte, Marta Severo</i>	
RECONFIGURATION DES PRATIQUES PARTICIPATIVES	239
<i>Nicolas Douay, Maryvonne Prévot</i>	
REMERCIEMENTS	259
LES AUTEURS.....	261