



HAL
open science

Les transformations des comportements politiques au prisme de l'e-pétitionnement

Jean-Gabriel Contamin, Thomas Léonard, Thomas Soubiran

► **To cite this version:**

Jean-Gabriel Contamin, Thomas Léonard, Thomas Soubiran. Les transformations des comportements politiques au prisme de l'e-pétitionnement : Potentialités et limites d'un dispositif d'étude pluridisciplinaire. Réseaux : communication, technologie, société, 2017, 2017/4 (204), pp.97-131. 10.3917/res.204.0097 . halshs-01577546

HAL Id: halshs-01577546

<https://shs.hal.science/halshs-01577546>

Submitted on 8 Oct 2018

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

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

Les transformations des comportements politiques au prisme de l'e-pétitionnement : potentialités et limites d'un dispositif d'étude pluridisciplinaire

Jean-Gabriel CONTAMIN (CERAPS, Université de Lille),

jean-gabriel.contamin@univ-lille2.fr

Thomas LEONARD (CERAPS, Université de Lille), tl.leonard@laposte.net

Thomas SOUBIRAN (CERAPS, Université de Lille), thomas.soubiran@univ-lille2.fr

En 1997, Nonna Mayer, Pierre Favre et Olivier Fillieule (Favre et al., 1997) proposaient de recourir aux sondages en manifestation afin de résoudre une question que les méthodes alors dominantes ne permettaient pas de traiter : savoir « selon quelles modalités une disposition à (disposition à manifester, disposition à la violence, disposition à l'engagement) se traduit ou non par une action effective ». Il s'agissait de prendre les manifestants « sur le fait », en action, afin de définir leurs traits socio-démographiques ainsi que les « itinéraires » qui les ont conduits à manifester. Cet article méthodologique se terminait en annonçant un second article destiné à montrer que cette technique nouvelle pouvait permettre « d'obtenir des résultats significatifs qui enrichissent notre connaissance des fondements et des mécanismes de l'action collective ». Article qui ne fut jamais publié.

Il est bien sûr difficile de savoir pourquoi ce second article n'a pas alors été publié. Mais, peut-être peut-on rattacher cela à certaines des difficultés que pose la méthode proposée, qui s'est institutionnalisée sous le nom d'INSURA (INDividual SURveys during RAllies) : difficulté à obtenir un échantillon représentatif des manifestants présents lors de l'étude; difficulté, aussi et surtout, à interroger la significativité même des manifestations étudiées. Dans la mesure où le dispositif méthodologique est lourd à mettre en place, on ne peut le démultiplier si bien qu'on peut apprendre beaucoup de ou des manifestations analysées mais qu'il peut être délicat d'en induire des résultats sur « les » manifestants en général. Du reste, si cette méthode a par la suite été réutilisée pour étudier différents événements protestataires, notamment des forums altermondialistes, c'est plus pour connaître le public spécifique de ces événements ou types d'événements que pour connaître l'« être manifestant » dans sa généralité (Agrikoliansky *et al.*, 2005, Della Porta, 2009, Siméant, 2015).

Le développement des nouvelles technologies et de leurs usages par les militants et organisations militantes semble ouvrir des possibilités nouvelles qui permettraient de passer outre certaines de ces difficultés. En l'occurrence, cette contribution vise à présenter un dispositif méthodologique inédit fondé sur une collaboration entre des sociologues, des politistes, des juristes, des spécialistes du langage et des informaticiens, français et belges¹, à qui le propriétaire d'un des principaux sites francophones de e-pétitionnement a donné accès, sous forme anonymisée, à l'ensemble de sa base. Ces données concernent plus de 15 000 pétitions mises en ligne entre le 31 octobre 2006, date de la première pétition du site, et le 12 février 2015, date de l'extraction de la base, lesquelles ont récolté au total près de 3,8 millions de signatures, ce qui correspond à un peu plus de 700 000 adresses courriel encore valides.

L'accès à une telle base offre en effet un ensemble d'opportunités nouvelles : travail sur des engagements réels et non potentiels; informations précises sur l'action pétitionnaire et ses acteurs, etc. On sait par exemple exactement à quels moments les signatures ont eu lieu et on peut replacer ces engagements dans la dynamique générale de mobilisation. Nous disposons aussi à l'occasion de

¹ ANR APPEL « Analyse Pluridisciplinaire du Pétitionnement En Ligne ». Le dispositif d'étude, on y reviendra, combine des analyses fondées sur la base de données à un questionnaire en ligne, puis à une batterie d'entretiens auprès de promoteurs et de signataires de pétitions. S'y ajoutent une réflexion juridique et le développement d'outils informatiques d'analyse des pétitions.

commentaires qui donnent un sens aux engagements, voire permettent de saisir la pluralité des significations attachées à l'engagement dans une même mobilisation (Contamin, 2001). Enfin, nous disposons d'un ensemble d'éléments sur celles et ceux qui s'engagent, en l'occurrence, un sexe, une profession et les autres engagements que la personne a pu réaliser sur le même site. Qui plus est, on n'est pas, comme dans les travaux sus-cités, limité à l'étude d'un petit nombre d'actions de mobilisation puisque la base offre l'opportunité de comparer les processus de mobilisation liés à un très grand nombre d'initiatives, de tailles, de portées et d'objets très différents.

Invoquer certaines potentialités du dispositif, c'est aussi implicitement en souligner certaines limites. Quant aux informations dont on ne dispose pas via la base. Et surtout quant à la significativité même de ce qu'on trouve sur le site. De fait, si nous disposons de données sur un ensemble de e-pétitions, rien ne nous assure que les résultats obtenus à partir de ce site peuvent être extrapolés à l'ensemble des e-pétitions a minima françaises et belges, notamment du fait de « l'effet de plateforme » ou du rôle des algorithmes (Cardon, 2015).

C'est précisément l'enjeu de cette contribution, en s'inspirant des réflexions développées autour des méthodes d'analyse du web (Barats, 2013) tout en les nourrissant à partir de ce cas spécifique, que de présenter les potentialités que peut offrir l'accès à une base de plus de 3 millions de signatures, les difficultés d'ordre multiple auxquelles se heurte une telle étude, mais aussi certaines des voies de résolution de ces difficultés, ainsi que quelques premiers résultats liminaires.

En questionnant les conditions de possibilité d'un usage scientifique de ces données, nous nous inscrivons dans une voie moyenne entre ceux qui critiquent les éventuelles dérives de l'émergence du « big data », en mettant en exergue les usages commerciaux qui en sont faits (Bourdoncle, 2014) et ceux qui y voient, à l'inverse, dans un excès d'enthousiasme, l'émergence d'une « Physique sociale » permettant une « théorie causale de la structure sociale » (« causal theory of social structure ») basée sur des explications mathématisées (Pentland, 2014).

On soulignera d'abord les potentialités d'un dispositif inédit si on le compare aux méthodes utilisées jusque là pour étudier les mobilisations en général, et le pétitionnement en ligne en particulier. On rappellera dans un second temps que ce dispositif se heurte toutefois à une difficulté à laquelle sont confrontés nombre des chercheurs qui souhaitent travailler sur des Big Data issus de l'internet : ces données proviennent d'une collecte non supervisée, ce qui induit un ensemble de difficultés de traitement. On présentera, enfin, dans un dernier temps, quelques stratégies méthodologiques, pour l'essentiel non spécifiques aux Big Data, qui permettent de « faire parler » nos données, ou tout au moins de mieux comprendre ce qu'elles peuvent dire, notamment en combinant les méthodes inédites que permet l'accès à de telles données à des méthodes plus traditionnelles qui répondent pour partie aux limites de ces méthodes innovantes.

LES POTENTIALITÉS D'UN DISPOSITIF INÉDIT : FAIRE DU NEUF AVEC DU NEUF

Forme dite non conventionnelle de participation politique, le pétitionnement est une des pratiques les plus anciennes et communes des démocraties libérales. Cependant, peu de travaux spécifiques lui ont été consacrés (toutefois : Neiman, Gottdiener, 1982 ; Contamin, 2001, 2013), par rapport aux formes « conventionnelles » comme la participation électorale et partisane.

Le développement des technologies de l'information et de la communication (TICs) a toutefois contribué à lui donner une nouvelle actualité. Ces évolutions semblent en effet d'abord avoir facilité le recours à cette pratique comme en témoignent la prolifération de sites web consacrés au pétitionnement ainsi que le fait que l'e-pétitionnement apparaît comme la forme de participation civique la plus utilisée par les internautes. Elles semblent aussi lui avoir donné une nouvelle efficacité à la fois sociologique et institutionnelle (Alathur, 2007; Morva, 2016). L'e-pétitionnement a ainsi été admis par plusieurs « pays » (Écosse, Pays Basque, Corée du Sud, Allemagne, Grande-Bretagne, France, Queensland,...), ainsi qu'au niveau de l'Union européenne, comme un

nouvel instrument institutionnel au service d'une démocratie supposée plus participative (Macintosh *et al.*, 2002; Miller, 2009; Bohle et Riehm, 2013). Le pétitionnement, de pratique dépassée, semble pour partie devenue l'un des sites privilégiés d'observation de l'émergence potentielle d'un nouveau rapport au politique et de nouvelles formes de participation politique.

Ce retour dans l'actualité s'est aussi traduit au niveau scientifique par la multiplication des travaux qui visent à prendre en compte ce nouveau comportement et à en faire l'occasion d'une réflexion sur les transformations des formes d'engagement liées au développement des nouvelles technologies. Sans pouvoir être exhaustif, on peut distinguer les études qui, pour l'essentiel, recourent aux méthodes traditionnelles d'analyse des données pour tenter de cerner ce phénomène et son éventuelle nouveauté, et celles qui, à l'inverse, voient dans la nouveauté de ces pratiques et des données qui en sont issues l'occasion, sur le modèle de ce que proposait H. Margetts (2009), de développer de nouvelles approches méthodologiques et théoriques, de contribuer à l'émergence de sciences sociales computationnelles (*Computational Social Sciences*) (Jungherr et Jürgens, 2010).

Étudier de nouvelles pratiques avec d'anciennes méthodes

Dans la première veine, on trouve des travaux qui portent sur des pétitions électroniques particulières ou des campagnes de pétitions électroniques particulières, interrogées à partir d'analyses de contenu et d'analyse documentaire (Boure et Bousquet, 2010; Boure et Bousquet, 2011; Morva, 2016), d'entretiens auprès de pétitionnaires, de journalistes et de récepteurs de pétitions (Alathur, 2007), voire d'une combinaison de ces techniques avec le recours à un questionnaire (Mosca et Santucci, 2009). D'autres études portent sur les e-pétitionnaires plutôt que sur les e-pétitions, en recourant aux techniques classiques de questionnaire représentatif sur l'ensemble d'une population (Schlozman *et al.*, 2010), ou de manière moins classique à un échantillon d'utilisateurs d'une plateforme pétitionnaire (PPC, 2006; Jonsson, 2014) ou à des échantillons comparés de e-pétitionnaires et de pétitionnaires « classiques » (Escher et Riehm, 2016), voire en menant des campagnes d'entretiens avec des citoyens (Macintosh *et al.*, 2002) ou avec des promoteurs de pétitions (Scott, 2015).

Outre le caractère très chronophage de certaines de ces techniques (l'étude d'une seule pétition exigeant un travail de dépouillement très important sans permettre ensuite facilement de monter en généralité), elles se heurtent très directement à la difficulté de travailler sur ce que les enquêtés disent de ce qu'ils font ou de ce qu'ils ont fait, plutôt que sur ce qu'ils ou elles ont effectivement fait, alors même qu'on sait que les pratiques d'engagement sont particulièrement sensibles aux discours rétrospectifs de légitimation ainsi qu'aux phénomènes d'oubli, plus ou moins sélectif (Bernard, 1984).

Le recours au Big Data au risque de la désociologisation

À l'inverse, d'autres dispositifs d'étude cherchent à tirer partie du potentiel que peuvent constituer ces bases de données e-pétitionnaires qui renseignent sur des populations entières (non échantillonnées, au moins pour ce qui concerne le site pris en compte) et sur des pratiques en temps réel. Là, encore, on peut noter un degré de sophistication et d'inventivité très différent dans l'étude de ces traces de dynamiques pétitionnaires présentes sur le Net. Certains se contentent de rapporter le nombre de signatures obtenues par chaque pétition à d'autres faits sociaux : les textes eux-mêmes pour s'interroger sur l'efficacité différentielle des contenus linguistiques (Hagen *et al.*, 2016), les mobilisations via les réseaux sociaux en comparant appartenances à des groupes Facebook et signatures de pétitions (Panagiotopoulos *et al.*, 2011). D'autres prennent en compte de manière plus précise les dynamiques des campagnes pétitionnaires en vue de mettre en avant la variété des formes de mobilisation (Yasseri *et al.*, 2013) et d'en rendre compte à partir d'expérimentations psycho-sociales (Margetts *et al.*, 2015). Enfin, certains partent des pétitionnaires eux-mêmes pour

mettre en évidence la variété de leurs pratiques (signature d'une ou de plusieurs pétitions, sur quelles thématiques, à quel moment d'une mobilisation) et l'interrelation entre les pétitions (quelles pétitions sont signées par les mêmes personnes ou communautés) en vue de faire des typologies d'utilisateurs (Jungherr et Jürgens, 2010) ou de mettre en évidence des communautés d'utilisateurs (Dumas *et al.*, 2015), éventuellement en combinant le travail sur la base avec l'envoi de questionnaires à des échantillons d'utilisateurs plus ou moins bien cadrés (Schmidt et Johnsen, 2013).

Reste que, pour des raisons pratiques (travail uniquement sur la partie publique des sites de pétitionnement, travail sur des sites institutionnels de pétitionnement,...), ces travaux manquent de sociologisation – on distingue des types de pétitionnaires, mais on ne sait guère qui ils sont – et se limitent à une trace réduite de comportements réels (la signature) là où d'autres sites permettent de prendre en compte pour partie la signification donnée sur le moment à la signature, via l'ajout de commentaires.

Les potentialités de l'accès à une base de données inédite

C'est là précisément que se trouve tout l'intérêt d'avoir eu accès à une base de données telle que celles du site *lapetition.be*. Celle-ci permet en effet de dépasser plusieurs des limites liées aux formes traditionnelles d'étude de la participation politique, mais aussi, d'aller plus loin que les méthodes plus innovantes utilisées jusque là pour travailler à partir d'un site de e-pétitionnement.

En effet, ce biais permet d'abord de passer de l'étude des déclarations de comportement à l'étude des comportements réels. On ne travaille plus ni sur des dispositions à agir comme le font les études qui portent sur le potentiel protestataire, ni sur des déclarations *a posteriori* qui se heurtent à des effets de mémoire sélective (Pierce et Lovrich, 1982). On travaille directement sur les traces laissées par le comportement des enquêtés sur un site au moment même où ils l'ont fait. En l'occurrence, on dispose « potentiellement » pour chaque signature : de données sur la pétition (texte, moment de lancement, initiateur de la pétition avec des données sur celui-ci, catégorie dans laquelle il a classé la pétition², dynamique de rassemblement des pétitions), de données sur le signataire (son prénom, son lieu de résidence, sa profession, les autres pétitions qu'il a signées et/ou initiées) et de données sur le moment où il/elle a signé et sur les commentaires qu'il/elle a éventuellement ajoutés à sa signature.

Encadré n°1 : Présentation sommaire de la base

Les données transmises par le propriétaire du site *lapetition.be* comportaient cinq fichiers distincts.

Quatre fichiers relatifs aux pétitions :

- Deux fichiers sur les promoteurs de pétitions de 23 301 lignes (compte utilisateur) et 15 209 lignes (renseignements) respectivement
- Un fichier contenant le texte des pétitions de 15 015 lignes
- Une table donnant le libellé des catégories de pétition (9 lignes)

Et un fichier sur les signatures de 3 795 565 lignes donnant des éléments sur les signataires, le moment de la signature et les commentaires éventuellement ajoutés. Il n'est pas nécessaire d'avoir un compte utilisateur sur le site pour signer.

Il était possible de faire le lien entre ces différents fichiers via des informations identifiantes hashées par le propriétaire.

On reviendra plus avant sur les données manquantes et éventuelles incohérences entre les tables.

² Neuf catégories sont disponibles sur le site, dont une catégorie « autres ».

À partir de la seule base, différentes stratégies de recherche sont alors possibles. On peut choisir de travailler au niveau des signataires, en comparant non pas seulement des types de pétitionnaires, mais, plus précisément, par exemple les caractéristiques des signataires de pétitions et des promoteurs de pétitions, celles des premiers signataires et des suivants, celles des monosignataires et des multisignataires (Bright *et al.*, 2016). On peut aussi, à la manière de H. Magretts et son équipe, analyser la variété des dynamiques pétitionnaires, mais aussi et surtout « sociologiser » ces typologies, en les rapportant aux thèmes et textes des pétitions, ou aux caractéristiques des pétitionnaires aux différents moments de la mobilisation.

Les particularités de cette base permettent aussi de dépasser les formes traditionnelles d'étude des actions de mobilisation par un autre biais. Là où celles-ci ne portent le plus souvent que sur un fait (untel s'est mobilisé de telle manière à tel moment) sans pouvoir interroger la signification de cet engagement autrement qu'après coup, les commentaires laissés sur le site permettent d'analyser la tonalité que les signataires donnent à leur engagement au moment même où ils le font : 23 % des signatures sont en effet assorties de commentaires. Il ne s'agit pas de supposer que ces commentaires nous donnent la vérité de chaque engagement. Chacun est en effet une forme de présentation de soi qui a sa part d'artificialité comme peut l'avoir un propos dans un entretien rétrospectif. Mais, ces commentaires révèlent *a minima* la pluralité des significations données à une même signature et des formes d'endossement d'un même texte-pivot. En ce sens, une telle base de données est un instrument particulièrement pertinent pour travailler sur l'appropriation des pétitions par les signataires eux-mêmes et pour montrer en quoi « une pétition [en l'occurrence électronique, mais cela est vrai de toute mobilisation] n'est pas un texte revendicatif intangible que chacun des signataires devrait endosser dans sa totalité, mais plutôt un texte-pivot auquel un ensemble de personnes qui n'ont pas participé à sa rédaction acceptent, pour des raisons hétérogènes, d'être rattachées » (Contamin, 2001). Autrement dit, via les commentaires, presque chacun signe « sa » pétition et le chercheur dispose d'un moyen privilégié de comprendre quelle pétition il/elle a signée.

Ce dispositif méthodologique présente une seconde vertu par rapport aux dispositifs jusque là utilisés. Certains travaux ont déjà porté sur des comportements d'engagement réels, tels que les sondages en manifestation, mais ils étaient contraints de se limiter à quelques événements choisis, empêchant les possibilités de montée en généralité. L'usage de cette technique lors des Forums sociaux mondiaux est significative puisqu'il s'agit alors moins de travailler sur les dynamiques de mobilisation en général que sur les spécificités et transformations des dynamiques de mobilisation altermondialistes. À l'inverse, l'usage de cette base de données permet de comparer une grande variété de dynamiques protestataires, de questionner leurs formes de diffusion, leurs ampleurs, et les facteurs de ces réussites et non-réussites. Nous pouvons ainsi interroger les succès différentiels des actions pétitionnaires en fonction des thématiques des textes, en induisant des hypothèses sur les communautés portant ou non tel ou tel type de pétitions. Nous pouvons surtout interroger le succès différentiel d'actions de mobilisation lancées autour d'une thématique commune selon le moment du lancement, le contenu exact du texte pétitionnaire, mais aussi les groupes initiateurs et les lieux et milieux de diffusion.

On voit ainsi toute la fécondité potentielle d'une telle base de données. Toutefois, souligner que ces données permettent d'émettre des hypothèses, c'est implicitement se demander dans quelle mesure elles peuvent suffire à induire des conclusions. Insister sur l'ampleur des données, parfois inédites, dont on dispose, c'est aussi sous-entendre les difficultés auxquelles on peut être confronté face à ces données, certes massives, mais souvent fragiles, lacunaires et ambiguës. Pour ne prendre qu'un exemple, nous sommes ainsi supposés disposer de la localisation des signataires. Reste que cette donnée n'est pas toujours présente (cf tableau n°1). Surtout, même quand elle est présente, il reste à pouvoir la traiter. Dans la colonne « pays », on a par exemple pu dénombrer 493 mots différents pour désigner la Belgique. Quand on passe au niveau plus fin des communes, le recodage

devient encore plus compliqué, au risque de mettre en cause la possibilité même d'une analyse. C'est ce qu'on va aborder dans une deuxième partie.

LES ÉCUEILS D'UNE COLLECTE NON-SUPERVISÉE

Différents auteurs ont souligné les difficultés rencontrées lors du travail sur des traces laissées sur le web ou autre part en les imputant au volume d'informations à traiter. Cependant, ce type de démarche butte aussi sur de nombreux autres écueils (Ollion et Boelaert, 2015) à commencer par leur mode de collecte. En effet, les données n'y sont le plus souvent pas recueillies mais littéralement *données*. Il n'est, ce faisant, pas possible de concevoir la collecte dans la perspective du traitement. Ce mode de collecte non-supervisé contraint fortement à la fois la transformation en données, les analyses envisageables et l'interprétation des résultats. Car le mode de collecte n'impacte pas seulement l'extraction d'information en l'absence de contrôle du processus de réponse. Il rend difficile l'identification de la population analysée et des individus qui la composent. Il en résulte de plus un volume conséquent de valeurs manquantes qui est loin d'être sans conséquences sur les analyses, et ce, malgré la « masse » des données.

Ces écueils seront illustrés par deux aspects de l'exploitation de la base. D'une part, le codage des informations extraites de la base permettra de prendre la mesure concrète des effets de l'absence de supervision sur les renseignements disponibles et les nombreux flous qu'il induit. D'autre part, l'analyse du graphe biparti des signatures aux pétitions donnera à voir, au-delà des difficultés pratiques de l'analyse, les difficultés d'interprétation de traces d'activités laissées sur le web.

Une collecte non-supervisée

De nombreuses difficultés du travail sur des données comme celles issues du site *lapetition.be* proviennent de leur mode de collecte. Les données mobilisées ici résultent en effet d'une extraction de la base sql servant de *backend* au site. Il s'agit donc du produit d'une collecte passive de données dont la finalité et donc les propriétés sont indépendantes des finalités de la recherche. La collecte des données n'est donc pas conçue ni réalisée dans la perspective du traitement et ce découplage impacte fortement les différentes étapes de l'analyse des données ainsi que le sens qui peut en être donné.

L'absence de supervision pèse en premier lieu sur la qualité des données et contraint très fortement les possibilités de traitements réalisables. Elle ne laisse en effet aucune liberté quant au choix des informations qui sont recueillies, à la façon dont elles le sont et auprès de qui elles le sont. Elle implique en conséquence une longue étape préalable d'extraction d'informations à partir des traces laissées par les utilisateurs dans leurs interactions avec le site.

Les difficultés tiennent d'abord aux inconsistances de l'instrument de collecte proprement dit (le site) qui récupère des informations variables dans le temps, par exemple en fonction de ses modifications ou des options proposées à ses utilisateurs. De plus, il n'est, par construction, pas possible d'encadrer les individus dans leurs interactions avec l'instrument de collecte. On peut ainsi noter la part très importante prise par les valeurs manquantes dans la quasi-intégralité des champs remplis par les utilisateurs (tableau 1). À cela s'ajoute l'imprécision sur les informations recueillies, en fonction de l'interprétation que les signataires font des intitulés des champs des formulaires. La collecte passive ne permet par définition pas d'organiser, par exemple, un questionnaire de façon à limiter les ambivalences dans la compréhension des questions en fonction du mode d'administration du questionnaire (Weisberg, 2005)³.

³ On retrouve ce même travers avec d'autres types de méthodes comme les enquêtes par questionnaires auto-administrés sans plan de sondage, notamment pour ce qui est de l'échantillonnage et des non-réponses. Néanmoins, à la différence d'un questionnaire, la récupération de traces ne laisse aucun contrôle sur le recueil des informations recueillies.

Tableau 1: Pourcentages de valeurs manquantes des colonnes des tables pétitions et signatures de la base de données *lapetition.be*

Table	Description de la colonne	Pourcentage de valeurs manquantes
Pétition	Prénom de l'auteur de la pétition	20.1
Pétition	Nom de la localité de l'auteur de la pétition	44.5
Pétition	Code postal de la localité de l'auteur de la pétition	47.1
Pétition	Pays de l'auteur de la pétition	38.6
Signature	Prénom du signataire	3
Signature	Nom de la localité du signataire	8.5
Signature	Code postal de la localité du signataire	76
Signature	Nationalité du signataire	23.2
Signature	Pays du signataire	12.8
Signature	<i>Hash</i> de l'adresse électronique du signataire	0
Signature	Profession du signataire	57.25
Signature	Commentaires	77.3

L'absence d'accompagnement pèse ici d'autant plus que la plupart des informations recueillies sont de nature textuelle et que les éléments collectés sur une base déclarative constituent l'essentiel des caractéristiques des utilisateurs du site. Or, l'hétérogénéité des réponses rend parfois difficile la traduction des traces en variables exploitables numériquement. L'absence de contrôle a aussi des conséquences sur la structuration même des données car l'identification des éléments fondamentaux de la compréhension du phénomène observé que sont la population et les individus qui la composent est entachée d'un certain flou.

Des traces aux données

Le titre des pétitions ainsi que le texte qui les accompagne permettent de donner une première illustration de ces différents problèmes. En ce qui concerne ces deux champs, on ne compte certes pas de valeurs manquantes mais il est apparu que certains textes ont pu être modifiés après publication. Ainsi, le texte original de la pétition 408 portant sur l'unité de la Belgique, a été remplacé par un texte de remerciements suite à son succès⁴. Mais, surtout, les propriétés morpho-syntaxiques (correction orthographique, grammaticale, syntaxique) de ces textes rendent la lemmatisation de certains textes parfois difficile (Barats *et al.*, 2016). Ceci a notamment pour effet

⁴ Il s'agit de la pétition ayant obtenu le plus de signatures sur ce site.

de perturber l'analyse textuelle et pourrait biaiser les résultats en faveur de ceux qui disposent d'une expression plus conforme aux standards d'écritures.

Le champ « profession » de l'application illustre les problèmes posés par l'absence de cadrage sémantique dans la collecte. Les 43 % de « professions » renseignées correspondent à 129 279 occurrences uniques. Leur examen a d'abord montré que le niveau de détail varie fortement d'une signature à l'autre. Certains signataires ne donnent ainsi qu'une catégorie générique (« ouvrier », « employé », etc.) là où d'autres donnent force détails. Il apparaît ainsi que les professions les plus qualifiées sont celles qui sont les plus promptes à détailler (enseignants, médecins, ...). Certaines réponses manifestent plutôt la situation de l'individu (« étudiant », « chômeur », « retraité », etc.) en la combinant parfois à la profession exercée (« retraité de l'enseignement »). D'autres semblent plus marquer le rapport du signataire à leur situation actuelle (« merde professionnelle »). D'autres réponses renvoient, elles, plutôt aux activités et aux responsabilités militantes, politiques ou associatives des signataires parfois en complément d'une indication de leur statut, peut-être pour légitimer ou donner plus de poids à leur signature. Les signataires déclarent ainsi être militant, conseiller municipal, membre, secrétaire ou président d'association en précisant parfois l'objet de leur engagement (« animalière », « anti-nucléaire ») ou le nom de l'association.

Il apparaît donc que, dans certains cas, les signataires ne répondent pas à la même question du fait de l'absence de tout cadrage de la collecte. L'hétérogénéité des réponses à ce champ rend donc difficile leur codage en utilisant une nomenclature comme la PCS-ESE. Car, au-delà du fait que les renseignements fournis par les signataires ne sont pas toujours d'ordre professionnel, le codage des professions est une opération complexe qui ne peut se baser sur un simple libellé (Desrosières et Thévenot, 2002). Les enquêtes produites par la statistique publique utilisent ainsi une batterie de questions sur la situation, le statut d'emploi ou encore le secteur qui sont absentes du site.

Les problèmes posés par une collecte non-supervisée affectent aussi des renseignements *a priori* mieux cadrés comme la résidence. 91,5 % des signataires ont ainsi renseigné leur localité, 87 % leur pays de résidence et 24 % leur code postal. Toutefois, au problème de graphie des toponymes, s'ajoute le problème des homonymes notamment induit par l'utilisation d'une version raccourcie du nom des localités. Ces faux positifs peuvent aussi provenir d'homonymies entre localités de pays différents (Lille en France ou en Belgique ?), ces deux sources d'erreur pouvant se combiner (Fribourg). Ce problème a son pendant pour les codes postaux. Différents pays comme la France, la Belgique ou l'Allemagne utilisent un code exclusivement numérique à cinq ou six chiffres faisant qu'un même code peut correspondre à des localités situées dans différents pays.

Le champ « pays » peut aider à lever ce type d'ambiguïtés même si l'on rencontre ici les mêmes écueils que ceux mentionnés précédemment, les signataires pouvant parfois exprimer plus ou autre chose que ce qui leur est demandé. Certains déclarent ainsi habiter au « Belgikistan » ou encore dans le « Désopiland », soit, dans ces deux cas, en Belgique. La dispersion géographique des signataires (près de trente pays) complique aussi la tâche même si l'essentiel des signataires de pétitions se concentrent en France et en Belgique. Au final, 3,3 millions de signatures (soit 88% du total) ont été géoréférencées. Les autres localités encore non identifiées se situent principalement en dehors de la Belgique et de la France ce qui rend leur identification plus complexe ne serait-ce que du point de vue de la quantité de données auxiliaires à compiler.

Le traitement des prénoms a lui aussi présenté des difficultés similaires (fautes de frappe, diminutifs, pseudonymes) lors de l'attribution d'un genre aux signataires par leur appariement avec la base des prénoms de l'état civil de l'Insee⁵. L'examen des prénoms a de plus permis de montrer que certains *hashs* md5⁶ d'adresses mail pouvaient correspondre à plusieurs personnes comme des couples ou les utilisateurs d'une adresse générique. Au total, près de 230 000 *hashs* d'adresses sur 1 928 744 présentent des prénoms différents après standardisation de ces derniers pour limiter les

⁵ Données obtenues auprès de l'ADISP.

⁶ MD5 (*Message Digest 5*) désigne un algorithme de hashage cryptographique.

effets de graphie sur les comptes. À l'inverse, on peut aussi envisager qu'une même personne utilise plusieurs adresses différentes ce qui est dans le cas présent impossible à détecter.

Incertitudes sur l'identification des individus et de la population analysée

Le mode non supervisé de collecte n'a donc pas seulement un impact sur les renseignements recueillis mais aussi sur l'identification même des unités de l'analyse. En effet, les signataires n'ont pas besoin de s'enregistrer pour signer une pétition. Le seul moyen de détecter les signataires multiples est donc le *hash* de leur adresse mail. Or, les méthodes et les théories du traitement de données supposent que les unités puissent être indexées de façon non équivoque ce qui n'est pas le cas ici⁷.

De surcroît, le mode de collecte n'implique pas seulement un flou sur les unités de la population mais aussi sur la population dont elles font partie. Le processus de signature d'une pétition peut être décomposé en au moins deux étapes : la connaissance de l'existence de la pétition et la signature proprement dite. La population dont sont issus les signataires correspond donc aux individus ayant une probabilité non nulle de signer les pétitions du site (les individus ayant été informés des pétitions ou susceptibles de l'être⁸). Or, cette population est impossible à circonscrire au regard des renseignements disponibles. Les premières exploitations semblent ainsi indiquer que l'information peut être transmise par différents moyens comme, par exemple, les réseaux interpersonnels, associatifs, syndicaux ou politiques, les médias, la lettre d'information du site ou même sa fréquentation plus ou moins assidue. Mais, en aucun cas, le moyen de transmission de l'information n'est connu⁹. En l'occurrence, « $N \neq all$ » contrairement à ce qui est parfois affirmé pour légitimer l'analyse de ce type de traces. On se retrouve plutôt en présence de la réalisation d'un échantillon dont la population serait inconnue et, *a fortiori*, la probabilité de sélection de ses individus. Le caractère inobservé des contours de la population dont sont issus les signatures fait que des variables telles que le nombre de signatures sont tronquées. Cette dernière n'est en effet observée que conditionnellement à un seuil : avoir signé au moins une pétition. Les analyses risquent donc d'être entachées par construction d'un biais de sélection difficile à quantifier en l'absence d'information sur l'intégralité de la population¹⁰.

La question du biais de sélection rejoint celle du volume des valeurs manquantes dont les niveaux dépassent ici ceux généralement observés dans les enquêtes par questionnaires¹¹. Le volume des données sert parfois à justifier un postulat d'innocuité qui paraît cependant contestable. De fait, le croisement des variables contenant des valeurs manquantes peut conduire à supprimer une partie conséquente des observations et donc à exclure des pans entiers des observations. Surtout, la question est ici plus celle de l'indépendance entre les processus de non-réponse¹² et les variables d'intérêt, que celle des taux de valeurs manquantes proprement dit¹³.

Pour comprendre les effets de la non-réponse sur l'analyse, Rubin et Little (Rubin et Little, 2002) distinguent trois cas : MCAR (*Missing Completely at Random*), MAR (*Missing at Random*) et MNAR (*Missing Not at Random*). Dans le premier cas, les valeurs manquantes ne sont liées à aucune variable. Il s'agit du cas le plus bénin et on considère parfois que la non-réponse peut être

⁷ Lorsqu'on dispose de suffisamment d'informations fixes dans le temps pour chacun des enregistrements, l'absence d'identifiant unique peut être palliée au moyen des méthodes de *record linkage* (Fellegi et Sunter, 1969).

⁸ Il semble en effet difficile de considérer cette population comme fixe du fait des aléas de la communication.

⁹ Le recoupement d'informations comme l'horodatage, la géoréférencement ou la « profession » déclarée peut néanmoins laisser transparaître le processus de diffusion dans certains cas comme on le verra plus loin au sujet de l'analyse du graphe des pétitions.

¹⁰ Ces remarques font écho aux vifs débats suscités par l'enquête *Great British Class Survey* (Ruppert *et al.*, 2013).

¹¹ Les enquêtes par questionnaire ne sont bien sûr pas immunes contre les problèmes de non-réponse (de Leeuw et de Heer, 2002). Néanmoins, la mobilisation de données auxiliaires permet d'évaluer les biais et, pour partie, peut aider à y remédier.

¹² Non participation au site (« non réponse » des unités) et non réponse au champs proposés. À ces biais de non-réponse peuvent s'ajouter de plus les biais liés aux réponses comme les effets de l'(in)correction morpho-syntaxique sur la lemmatisation ou encore les effets des caractéristiques des individus sur les réponses au champ « profession ».

¹³ Différents travaux empiriques et théoriques montrent d'ailleurs que taux de non-réponse à une enquête et biais sont des processus liés mais distincts (Groves et Peytcheva, 2008).

ignorée même si elle peut notamment conduire à une diminution de la puissance des tests. Dans le cas des MAR et MNAR, la situation est moins favorable car la non-réponse est liée aux variables. Le cas MNAR est le plus critique car la non-réponse est liée aux variables d'intérêt et porte la marque d'un biais systématique souvent difficile à corriger. En fonction des situations, considérer les valeurs manquantes comme MCAR peut donc se justifier mais certainement pas au seul regard du volume des données disponibles. La difficulté est évidemment de statuer sur les non-réponses puisque seule la disposition des informations manquantes permettrait d'expliquer leur absence...¹⁴

Comme on le peut le constater, l'extraction d'informations à partir des traces laissées par les utilisateurs du site comporte son lot de difficultés. Néanmoins, il ressort du travail de préparation que le volume des données ne constitue pas ici la principale source des problèmes rencontrés. Ceux-ci apparaissent plutôt inhérents au mode de collecte. Le volume agit alors plutôt comme un effet multiplicateur qui rend plus difficiles l'extraction des informations et leur transformation en données exploitables. De plus, la collecte non supervisée induit non seulement des risques de biaiser de manière difficilement quantifiable les estimateurs mais elle affecte aussi l'interprétation même des résultats comme ceux issus de l'analyse du graphe biparti des signatures aux pétitions.

Du graphe au réseau social

Les signatures enregistrées par le site lapetition.be permettent de tisser deux types de relations : les liens entre les signataires d'une part et les liens entre les pétitions d'autre part. Les signataires sont en effet reliés entre eux par les signatures aux pétitions qu'ils ont en commun de même que les signataires permettent de relier les pétitions entre elles. Dans la théorie des graphes, ce type de relations porte le nom de graphe biparti (cf. encadré n°2). Une des difficultés intrinsèques de l'analyse de ce type de structure est son caractère dual. En effet, ce type de graphe relie deux ensembles distincts dont on peut donc dériver non pas un mais deux types de relations (ie : les relations entre pétitions d'une part et entre signataires d'autre part)¹⁵.

Encadré n°2 : les graphes bipartis

Un graphe biparti $G=(V_{\perp}, V_{\top}, E)$ est un graphe dont les sommets peuvent être séparés en deux ensembles V_{\perp} et V_{\top} distincts de façon à ce que les sommets de chaque ensemble ne soient reliés qu'aux sommets de l'autre ensemble (cf graphique 1-a). Autrement dit, les deux extrémités de chacune des arêtes de $E \subseteq V_{\perp} \times V_{\top}$ appartiennent nécessairement à deux groupes différents. Ainsi, dans le cas des pétitions en ligne, les pétitions ne sont pas directement reliées entre elles pas plus que les signataires. Ces deux ensembles de sommets ne sont reliés qu'indirectement par le biais des *signatures*¹⁶. Les graphes bipartis ont d'autre part une représentation matricielle appelée matrice d'affiliation (ou biadjacente) différente des matrices d'incidence ou d'adjacence utilisées pour représenter les graphes unipartis (cf. graphique 1-d).

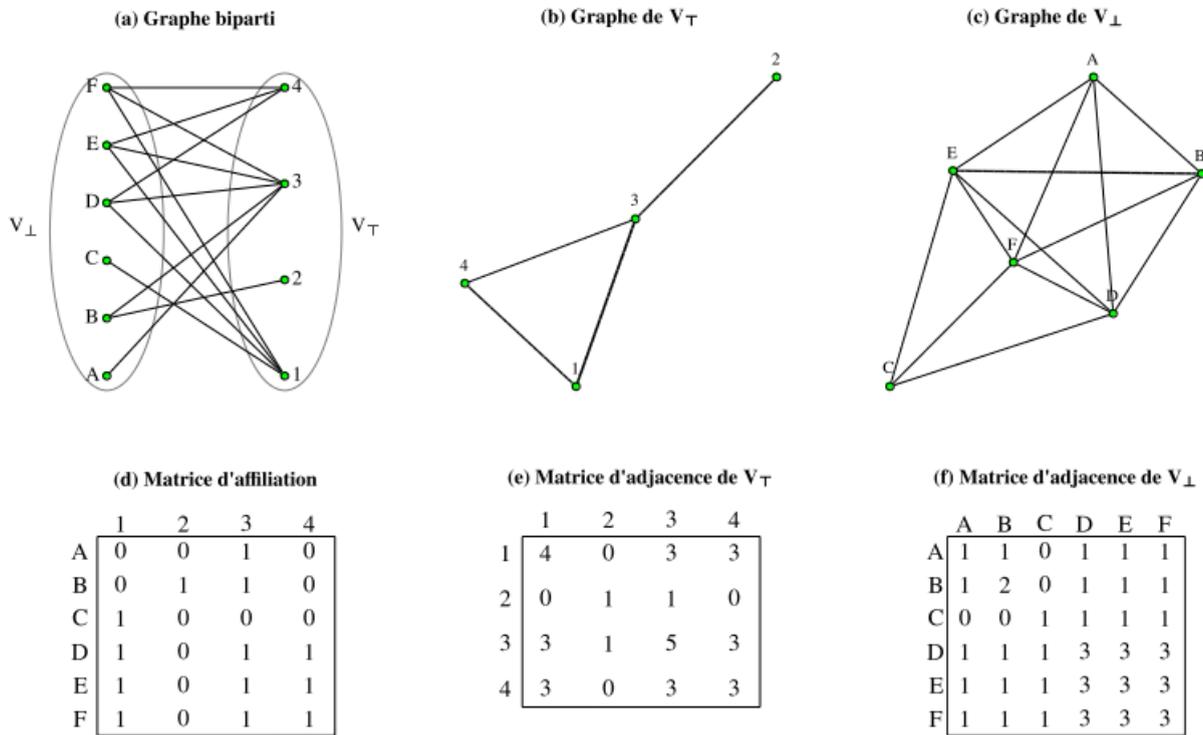
En matérialisant à la fois les relations entre deux ensembles comme les pétitions et les signataires, ce type de graphes présente donc une structure duale spécifique. Une des conséquences de cette propriété est que tout graphe biparti peut être transformé en deux graphes unipartis reliant les sommets au sein des deux ensembles (cf. graphiques 1-b et 1-c ainsi que 1-e et 1-f pour leur représentation matricielle). On parle alors de projection soit, par exemple pour \top , le graphe

¹⁴ Ce paradoxe peut cependant se résoudre en disposant de données auxiliaires.

¹⁵ Les liens observés peuvent donc être le produit de l'un ou l'autre des ensembles. Ainsi, la co-signature à deux pétitions différentes par deux individus différents peut être due au fait que les deux personnes sont connectées par leur réseau social ou (non exclusif) parce que les deux pétitions ont une thématique proche.

¹⁶ Il est à noter que les deux ensembles de sommets ne sont pas nécessairement de nature différente comme c'est le cas pour les signatures aux pétitions. Un graphe uniparti peut ainsi s'apparenter à un graphe biparti lorsqu'il s'organise autour de deux groupes où les membres de chaque groupe ne communiquent qu'avec les membres de l'autre groupe. Les forums d'entraide présentent souvent cette caractéristique. Les individus posant des questions ne répondent pas ou peu aux questions posées par les autres et, réciproquement, ceux y répondant ne posent de questions que très rarement. Un graphe uniparti peut donc aussi être biparti.

$G_{\tau}=(V_{\tau}, E_{\tau})$ dans lequel deux sommets de τ sont liés s'ils ont au moins un élément en commun dans \perp . Les entrées hors diagonale de la matrice d'adjacence de V_{τ} (resp. V_{\perp}) comptent ce faisant le nombre de sommets dans V_{\perp} (resp. V_{τ}) que les sommets de V_{τ} (resp. V_{\perp}) ont en commun¹⁷. Pour les pétitions, la projection τ désigne le graphe reliant les pétitions en fonction du nombre de signataires qu'elles partagent, soit 12 327 sommets (pétitions) et 5 334 872 arêtes pour un total de 32 688 661 signatures en commun. En pratique, les projections ont notamment pour utilité de permettre le recours aux méthodes développées pour les graphes unipartis dans le contexte de l'analyse de graphes bipartis. Elles conduisent néanmoins à une perte d'information par la non prise en compte de la dualité du graphe.



Graphique 1 : Exemple de graphe biparti (a) avec ses deux projections τ et \perp (b et c) ainsi que les représentations matricielles correspondantes (d, e et f).

Lecture : Les sommets D et E sont tous les deux liés aux sommets 1, 3, et 4. D et E ont donc trois éléments de V_{τ} en commun.

Cependant, cette difficulté structurelle se trouve ici renforcée par un effet de taille qui ne se limite pas au nombre de sommets et d'arêtes. En effet, le graphe des signatures présente toutes les propriétés des grands graphes bipartis (Latapy *et al.*, 2008). Tout d'abord, sa densité est faible, les distances sont courtes et la moyenne des degrés des sommets¹⁸ est faible relativement à leur nombre. Leur distribution est de plus fortement asymétrique ce qui marque la profonde hétérogénéité du réseau. De ce fait, même si l'espérance de la variable est basse, certains sommets

¹⁷ Les entrées de la diagonale comptent, elles, le nombre total d'éléments dans l'autre ensemble auquel le sommet correspondant est lié, soit le degré du sommet.

¹⁸ Le degré désigne le nombre total de liens qu'un sommet entretient avec son voisinage.

peuvent présenter un degré très élevé¹⁹. Au-delà de la taille *stricto sensu* du graphe, ces propriétés sont de nature à compliquer les traitements car elles sont la marque de liens le plus souvent très dispersés entre sommets (individus ou pétitions) qui rendent difficile l'objectivation de profils de signatures. Et les choses se compliquent d'autant plus lorsqu'il s'agit d'intégrer la dualité de ce type de graphes.

La taille et les propriétés du graphe contraignent donc fortement les possibilités d'exploitation des données. Cependant, les problèmes théoriques et pratiques posés par l'analyse d'un grand graphe biparti ne constituent pas la seule limite de l'exercice. En effet, ce graphe ne peut pas être proprement parler considéré comme un réseau social. Il semble plutôt la réalisation d'autres réseaux qui se manifestent de façon *latente* dans ce graphe²⁰. L'analyse exploratoire a permis de faire ressortir certains réseaux qui se dégagent assez clairement comme dans le cas des pétitions animalières. D'autres ont pu être dégagés par la combinaison de l'horodatage, du géoréférencement et de la « profession ». Ainsi, l'examen des informations disponibles pour la pétition 13820 intitulée « Pour une formation Universitaire et Professionnelle des enseignants en Aquitaine » a permis de dessiner le déroulement de sa diffusion. Cette pétition portait sur les conditions de la mise en place de l'École supérieure du professorat et de l'éducation (ESPE) d'Aquitaine. Signée au total par 846 personnes, la pétition a d'abord circulé parmi les enseignants du secondaire de l'agglomération de Bordeaux avant de se diffuser progressivement dans le Sud-Ouest puis sur tout le territoire tout en diversifiant son recrutement social (étudiants, enseignants du supérieur...).

Il est cependant envisageable que le graphe soit aussi le produit de réseaux dont l'intersection est réduite, voire même de réseaux complément déconnectés (ou, du moins, suffisamment éloignés pour être considérés comme tels). Le caractère plus ou moins disjoint de ces réseaux pourrait d'ailleurs expliquer la faiblesse des liens entre pétitions. Celle-ci peut aussi manifester le caractère artificiel de la projection du graphe dans l'espace qui relie des unités n'ayant sans doute rien en commun dans les faits au-delà de la co-incidence de la signature. En effet, la génération des liens peut résulter d'autres facteurs que la dynamique de réseaux (diffusion par les médias, navigation sur le site,...). Les liens du graphe n'attestent donc pas de l'existence de réseaux entre les individus. Ils n'en sont qu'un éventuel signe.

Enfin, cette représentation du processus générateur implique qu'au travers des traces laissées sur le site on observe plutôt le résultat du processus que le processus proprement dit. L'essentiel semble plutôt se passer ailleurs que sur le site, en amont de la connexion et de la signature. Dans ces conditions, l'analyse de la base ne saurait suffire pour appréhender les dynamiques des pétitions en ligne.

LES POTENTIALITÉS D'UN USAGE « CLASSIQUE » D'UN NOUVEL OUTIL

L'ensemble de ces remarques nous invitent à nous interroger sur ce qu'on peut faire – ou non – de cette base et sur les manières de la « faire parler » en nous demandant notamment dans quelle mesure l'accès à ce type de données nouvelles suppose de recourir à des méthodologies elles-mêmes nouvelles.

Il s'agit donc de trouver les méthodes adéquates pour traiter ces différentes données et pour les interpréter. Or, l'une des principales difficultés rencontrées dans le traitement d'un tel matériau tient à son importance et à la méconnaissance scientifique relative des méthodologies idoines à l'analyse.

Nous plaiderons ici pour deux méthodes d'analyse complémentaires de la base de données. Nous présenterons d'une part les vertus heuristiques du croisement des données internes à la base

¹⁹ L'asymétrie des données pose un problème de visualisation en soi qui ne concerne pas que les graphes. Près des trois quart (74 %) des signataires n'ont en effet signé qu'une seule fois. La masse de la distribution se trouve donc concentrée autour de ce point et « écrase » visuellement les autres valeurs.

²⁰ De ce point de vue, la multiplicité des utilisateurs d'un même adresse devient moins problématique en ce qu'elle manifeste un lien entre individus (si le compte n'est pas utilisé frauduleusement bien évidemment).

en elle-même, en montrant, à partir d'un sous-corpus spécifique, comment on peut induire un ensemble d'hypothèses fortes, voire de résultats, sur les dynamiques pétitionnaires en ligne. D'autre part, nous montrerons que pour tester de manière plus approfondie ces hypothèses il est nécessaire de mobiliser d'autres matériaux empiriques. Aux données collectées via Internet, il convient d'ajouter d'autres données issues d'autres modes d'investigation classiques (Boyadjian, 2016) : questionnaires ou entretiens semi-directifs, par exemple.

La base comme outil de construction d'hypothèses

En dépit des difficultés auxquelles on est confronté face à ce type de matériau, il autorise aussi un ensemble de traitements auxquels ne peuvent recourir ceux qui ne disposent que des données disponibles par la seule consultation du site en ligne, en offrant pour chaque signature un ensemble de données complémentaires et exhaustives.

Ce matériau assure d'abord une grande fiabilité pour objectiver les dynamiques de chaque pétition en croisant les données de l'horodatage, de la profession et de la géo-localisation. La base fournit ainsi pour chaque signataire des informations relatives à trois types d'espaces. D'abord, *via* le code postal déclaré, on peut déterminer l'espace géographique sur lequel se répartissent les signataires d'une pétition. Ensuite, *via* la profession déclarée, on peut déterminer approximativement les différents espaces sociaux dans lesquels navigue une pétition. Enfin, l'horodatage permet la mise en relation des données de l'ancrage géographique et social des signataires et des réseaux de signataires avec leur diffusion dans le temps.

Dès lors, à défaut de pouvoir définir les PCS des signataires de pétitions, on peut saisir la dynamique de diffusion d'une pétition à travers les récurrences constatées quant à la profession et au lieu d'habitation. Par exemple, ces informations permettent de comprendre dans quelle mesure la forme électronique des pétitions peut être ou non un outil d'émancipation des réseaux et structures habituels de mobilisation politique.

Ces données permettent aussi la construction d'hypothèses relatives aux différents relais publics, semi-publics ou privé qu'autorise internet (Boure *et al.*, 2012) ainsi qu'aux relais existants en dehors. Nous pouvons alors étudier la plate-forme internet comme un médiateur spécifique parallèle aux réseaux existants. Nous avons d'abord constaté que les signataires d'une pétition à un moment donné tendent à présenter des caractéristiques assez homogènes soit en termes de profession et de localisation, soit en termes de fréquentation du site. Ceci suggère que les signatures dépendent, dans le premier cas, de réseaux concrets, assimilables sur certains aspects à ceux qui jouent le rôle de déclencheur de signatures pour les pétitions sous forme papier et, dans le second cas, d'un réseau propre à la plate-forme *lapetition.be*.

Enfin, la connaissance du lieu d'habitation et de la profession permet d'observer l'existence de récurrences qui sont autant d'indices d'une proximité effective entre les différents signataires d'une même pétition. Par exemple, la pétition 7182, signée par 256 personnes (dont 127 déclarent une profession) et intitulée *Citoyens, Exigeons la Création d'une Région Métropolitaine du Brabant !* a été signée par 10 médecins soit une sur-représentation particulièrement forte par rapport aux autres pétitions.

L'analyse des signatures, *via* l'étude des réseaux objectivable, ouvre la possibilité d'étudier dans quelle mesure les pétitions en ligne permettent ou non de s'émanciper d'un quelconque ancrage territorial. La récurrence de l'homogénéité géographique des signatures alors même qu'il serait en théorie possible de signer instantanément toute pétition en tout lieu du globe laisse penser que nombre de signatures sont en réalité dépendantes de réseaux concrets qui sont largement déterminés par les contraintes géographiques. Le cas des pétitions en ligne peut ainsi permettre de montrer en quoi « les réseaux virtuels n'ont pas éliminé les réseaux réels » (Lasserre, 2000). La base de données autorise le dépassement du constat qu'internet n'abolit pas les réseaux sociaux structurés autour de territoires physiques pour étudier plus précisément comment réseaux physiques

et réseaux virtuels entrent en interaction pour produire un espace qui a ses propres contraintes spécifiques (Greffet, 2012).

Elle fournit aussi un ensemble d'entrées pour interroger les spécificités de l'engagement en ligne. L'étude comparative de deux pétitions en ligne permet par exemple de montrer en quoi le pétitionnement électronique peut être lu comme un vecteur de « consumérisme politique » entendu comme une mobilisation politique produite par une « constellation d'engagements individuels épars » (Baygert, 2014), dans le prolongement de la thèse du développement d'un engagement « à la carte » (Ion, 1997) moins dépendant des structures de politisation traditionnelles et davantage des préoccupations individuelles.

La base de données comme outil de sélection des cas : les pétitions « La Belgique VIVRA ! » et « Unité de la Belgique »

Pour cette étude, deux pétitions appartenant à un sous-corpus de 147 pétitions portant sur la thématique de l'unité ou de la scission de la Belgique ont été extraites, en l'occurrence la 655 (« La Belgique VIVRA ! ») et la 845 (« Unité de la Belgique ») en raison de leurs fortes similitudes, qui permettent en retour de questionner leurs différences.

La base de données nous permet d'abord d'identifier l'une et l'autre des deux pétitions comme ayant eu « deux vies ». Chacune a été lancée dans la seconde partie de l'année 2007 (respectivement le 7 juillet et le 22 octobre), a connu une première vague de signatures (respectivement 55 et 341 en 2007), avant de cesser d'être signée puis d'être relancée plusieurs mois plus tard et, ce n'est pas anodin, le même jour (le 15 juillet 2008). Lors de ces « secondes vies », les pétitions 655 et 845 ont respectivement récolté 509 et 357 nouvelles signatures. En dépassant le nombre de 400 signatures, ces pétitions font toutes deux parties des dix les plus signées de notre sous-corpus de 147 pétitions.

Par ailleurs, L'une et l'autre des deux pétitions sont liées par de très nombreux signataires communs (166 en 2008) lors de leurs secondes vies alors qu'elles ne l'étaient pas lors de leurs premières (2 signataires en commun). En outre, les données étayent la thèse d'une action coordonnée entre signataires comme le montre le fait que beaucoup de ces signatures communes ont été apposées dans un laps de temps très réduit. Ainsi, le 15 juillet, 29 signatures communes aux deux pétitions sont apposées entre 12h52 et 15h29. Dans ce cas, c'est la possibilité de croisements entre les données de l'horodatage et l'identification des individus à travers les adresses e-mails haschées qui rend possibles ces déductions.

La possibilité de rapporter une signature à un individu *via* son e-mail permet aussi de savoir que beaucoup des signataires de ces pétitions de 2008 signent un nombre élevé d'autres pétitions sur la thématique de l'unité/scission de la Belgique sur le site, quand seule une petite proportion des signataires de 2007 sont dans ce cas. On parvient alors à déterminer que les signataires des secondes vies se caractérisent par un intérêt tout particulier pour l'action pétitionnaire sur le site *lapetition.be* sur cette problématique, ce qui ne se constate ni lors des premières vies de ces pétitions, ni sur la grande majorité des pétitions signées sur ce site.

Malgré ces similitudes entre les secondes vies de ces pétitions, elles se distinguent quant à leurs ampleurs respectives. Au cours des cinq heures qui suivent la première signature de la journée, le nombre de signataires ayant signé la 845 et non la 655 (74) est près de deux fois plus élevé que celui des signataires de la pétition 655 qui n'ont pas signé la 845 (39).

Cette différence en dépit de profondes similitudes entre les deux pétitions interroge. En développant l'analyse des pétitions relevant de cette thématique, on peut distinguer trois profils de signataires distincts entre ceux qui ont signé à la fois les pétitions 645 et 845 et ceux qui n'ont signé que l'une d'entre elles.

**Tableau 2 : Proportion de signatures sur les autres pétitions relatives
à l'unité/scission de la Belgique**

	408	2817	7009	7143	8517	autres	N=
655 et 845	81 %*	10 %	45 %	10 %	17 %	14 %	42
655 seule	85 %	8 %	26 %	10 %	23 %	8 %	39
845 seule	78 %	4 %	22 %	5 %	7 %	3 %	73

Champ : Signataires ayant signé les pétitions 655 et/ou 845 le 15 juillet 2008 avant 18h (N=154).

* lecture : 81 % de ceux qui ont signé à la fois la pétition 655 et la 845 ont également signé la pétition 408

La pétition 408 est signée dans d'importantes proportions par les individus des trois groupes. En revanche, ces derniers se distinguent quant à leur propension à signer d'autres pétitions. Ceux qui signent les deux pétitions signent beaucoup d'autres pétitions sur la thématique. À l'opposé, ceux qui signent la pétition 845 mais pas la 655 signent dans de moindres proportions d'autres pétitions sur la thématique. Entre les deux, ceux qui signent la pétition 655 mais pas la 845 signent d'autres pétitions sur la thématique dans d'importantes proportions, mais ils signent cependant bien moins fréquemment la pétition 7009 que ceux qui ont signé les deux pétitions.

« Consumérisme politique » Versus « Réseaux politiques »

Comprendre ces différences entre deux pétitions par ailleurs très proches permet d'interroger le poids respectif du « consumérisme politique » et des « réseaux politiques » dans les dynamiques pétitionnaires.

En effet, il est possible dans un premier temps d'interpréter les liens entre ces pétitions comme le produit de liens directs entre signataires ayant coordonné leurs actions pour diffuser ces pétitions. Ces liens peuvent également être indirects et dépendre d'un intérêt commun pour la thématique et de la fréquentation du site de pétitions en ligne. Parmi ces engagements pétitionnaires communs, il s'agit donc pour nous de distinguer en quoi ils sont le produit de réseaux pré-existants ou, à l'inverse, d'actions non-coordonnées de « consommateurs » de pétitions.

Avec les réseaux pré-existants, les individus sont liés entre eux si ce n'est par de fréquentes interactions de face à face, tout au moins par l'appartenance à un groupe structuré potentiellement lié à d'autres (par exemple des organisations ou des groupes politiques, présents exclusivement ou non sur les réseaux sociaux). Suivant cette explication, la communauté de signatures entre les deux pétitions s'explique pour partie par le fait que des acteurs sociaux ont fait en sorte de mobiliser leur réseau pour favoriser la réussite de ces deux pétitions.

Dans le cas d'une action conjointe de « consommateurs » politiques, la communauté de signatures trouve son origine dans la commune fréquentation du site associée à un intérêt commun pour la problématique. L'existence d'une communauté de signatures sur deux pétitions s'explique ici pour partie par un « effet de plateforme », entendu comme tout effet de la structuration du site contribuant à mettre en avant certaines pétitions davantage que d'autres. Dans cette optique, les signataires seraient « consommateurs » dans le sens où ils effectuent un choix sur une pétition dans une liste de pétitions et n'inscrivent pas leur action dans une mobilisation coordonnée. Cette consommation reste pourtant socialement déterminée, la communauté de signatures entre les deux

pétitions étant fondée sur des dispositions spécifiques à signer ce type de pétitions ainsi que sur la structuration du site orientant les pratiques pétitionnaires.

En se rendant sur le site *lapetition.be*, on constate que les pétitions sont mises en avant selon trois critères : « les dernières signatures », « les dernières pétitions » et « les pétitions les plus signées de la semaine ». Les pétitions mises en visibilité sont donc d'abord celles qui ont fait l'objet des signatures les plus récentes, celles qui ont été mises en ligne le plus récemment et celles qui ont fait l'objet du plus de signatures au cours de la semaine (que ces pétitions soient récentes ou non). Indépendamment de ses propres réseaux concrets et de ses propres appétences, un individu qui se rend sur ce site voit son attention attirée sur certaines pétitions et tout particulièrement celles qui ont été le plus récemment signées.

Dès lors, l'utilisateur « consommateur » se rendant sur le site *lapetition.be* le 15 novembre 2008 à partir de 13h constate la visibilité particulièrement forte des pétitions 655 et 845 ce qui l'incite à y porter son attention. De plus, ce type de structuration du site contribue à un effet auto-générateur de la mise en visibilité de pétitions particulières, puisqu'une pétition davantage signée à un moment donné est particulièrement mise en visibilité. Ceci favorise alors les signatures des usagers « consommateurs », ce qui renforce encore la visibilité de la pétition. La proximité des horodatages des signatures des deux pétitions peut donc s'expliquer par l'existence de réseaux structurés comme par un « effet de plateforme ».

Ceci étant, le fait de signer un grand nombre de pétitions sur une même thématique et de le faire précocement indique que le signataire est bien inséré dans un réseau structuré, voire qu'il y participe activement comme relais d'initiatives. Connaître l'existence d'une pétition est une chose. Connaître la plupart des pétitions sur une question donnée juste après leur mise en ligne et avant toute médiatisation massive en est une autre. Les données de la base montrent justement une corrélation claire entre le fait de signer précocement les pétitions 655 et 845 et le fait de signer beaucoup de pétitions sur la thématique de l'unité/scission de la Belgique.

Les deux pétitions semblent donc s'être appuyées sur le même réseau, mais l'une a vraisemblablement bénéficié d'une plus grande mise en visibilité sur le site : alors que la 655 doit surtout ses signatures à un réseau existant, la 845 bénéficie en plus d'un « effet plateforme ». Rien n'indique d'ailleurs que cette seconde pétition ait une propriété particulière expliquant cet état de fait. En effet, puisque le site met davantage en visibilité les pétitions les plus récemment signées, un simple effet contingent faisant que l'une ou l'autre des pétitions soit mise davantage en avant à un moment donné suffit à engendrer d'importantes différences.

Articuler l'analyse des données de la base avec celles d'un questionnaire et d'entretiens

Les données issues de la base ont donc permis de préciser nos hypothèses et d'y répondre pour partie. Pour autant, elles demeurent sous bien des aspects insuffisantes pour déterminer la nature des liens existants entre les signataires de mêmes pétitions et notamment dans quelle mesure les signatures dépendent des dispositions propres à chaque agent, de réseaux concrets ou encore d'interactions concrètes de face à face ou *via* les outils de communication numérique. Surtout, si l'existence d'une forte proportion d'individus liés au sein d'un réseau est démontrée, rien ne nous informe sur sa nature. S'agit-il d'un réseau professionnel ? S'agit-il d'un groupe fondé sur des affinités politiques ? Le cas échéant, ce groupe politique est-il un groupe spécifiquement fondé sur les réseaux sociaux ou a-t-il une existence en parallèle ?

Dans cette optique, il est nécessaire de produire d'autres données qui permettent, dans une démarche compréhensive, de restituer les logiques concrètes du recours à la pétition électronique. Le seul recours aux données de l'internet apporte en effet des données nombreuses, mais pas nécessairement riches sociologiquement, qu'il convient dès lors d'éclairer à partir d'autres dispositifs d'enquête (Boyadjian, 2016). En l'occurrence, dans notre dispositif, l'étude de la base

s'accompagne de la mise en place d'un questionnaire en ligne qui doit être diffusé à l'ensemble des signataires du site *lapetition.be* via la lettre d'information du site et d'une campagne d'entretiens.

Ce questionnaire a été conçu pour nous informer sur les pratiques pétitionnaires en elles-mêmes, sur les usages des nouvelles technologies de l'information et de la communication ainsi que sur les différents engagements politiques et associatifs des individus et nous permettre de mettre ces informations en relation avec les variables socio-démographiques usuellement étudiées (sexe, âge, PCS, etc.). Il nous permettra alors de saisir les raisons concrètes qui ont motivé la participation des individus aux pétitions en ligne ainsi que les causes structurelles qui l'expliquent comme le degré de politisation de l'individu, son insertion dans des réseaux spécifiques et les usages qu'il fait d'internet. Il permettra aussi de comprendre comment s'articulent la signature de pétition en ligne, la signature de pétitions papier et d'autres modes de participation politique, de manière générale et sur des campagnes de mobilisation particulières.

Les entretiens sont quant à eux utiles pour saisir en quoi la forme électronique du pétitionnement s'inscrit ou non dans le prolongement du pétitionnement sous forme papier comme mode de participation politique. L'existence de plateformes internet permet en effet *a priori* aux individus de trouver plus facilement des pétitions sur les thèmes qui les intéressent tout particulièrement et d'être moins dépendants des offres de pétitions que lorsque le pétitionnement n'avait qu'une forme papier. Si cette hypothèse de la nouveauté de ce répertoire d'action doit être envisagée, la littérature sur la question suggère cependant qu'il n'est pas évident qu'elle soit validée (Mathieu, 2013), les structures traditionnelles de diffusion des pétitions jouant encore un rôle primordial comme semblent d'ailleurs le confirmer les premiers résultats de notre enquête.

À travers les entretiens, on pourra également étudier la polyphonie à l'intérieur d'une même pétition (Contamin, 2001 ; Boure et Bousquet, 2011), c'est à dire la diversité des usages et des sens conférés à un même texte pétitionnaire. À ce titre, il est utile de mettre en parallèle les discours et les représentations véhiculées lors des entretiens avec, le cas échéant, les présentations qui en sont faites sur les différents relais (blogs, journaux, etc.) qui font référence au texte pétitionnaire. Nous procéderons aussi à des analyses lexicométriques des commentaires et de leurs évolutions polyphoniques, travail qui pourra être utilement complété par la réalisation d'entretiens avec les signataires.

Enfin, l'analyse de l'« effet plateforme » pourra utilement être complétée par une comparaison entre le site étudié et d'autres qui présentent des architectures différentes. À titre d'exemple, le site « Avaaz » met en avant les dernières signatures de pétitions (« En ce moment ») comme *lapetition.be* tout en se distinguant puisque les dernières signatures sont actualisées en permanence, sans que l'internaute ait besoin d'agir. Sur le site *lapetition.be*, les « dernières signatures » ne sont actualisées que quand l'internaute recharge la page. Cette différence a sans doute d'importantes répercussions quant à la manière dont les « dernières signatures » influencent les pratiques et signatures des individus sur le site. L'analyse de la nature de ces différences reste en tout cas encore à être réalisée.

CONCLUSION

Face à la question du « Big data », les discours académiques comme issus du débat public sont aujourd'hui le plus généralement tranchés : les uns y voient le moyen de l'émergence d'une « physique sociale » mathématisée quand d'autres pensent au contraire que l'utilisation de ces mégadonnées entraîne une illusion de scientificité et une dégradation concrète de la production de savoirs (Boelaert, Ollion, 2015). Il nous semble pour notre part que le débat doit se décentrer de ce que *pourraient être* les « Big data » pour se centrer sur la question des *conditions de la scientificité de leurs usages*. Qu'on s'en réjouisse ou qu'on la redoute, la multiplication des données numériques est un fait bien établi sur lequel on n'a aucune prise.

Pour notre équipe de recherche, l'accès à la base de données des initiateurs et des signataires de pétitions du site *lapetition.be* constitue à la fois un casse-tête épistémologique et une opportunité ouvrant des potentialités *en partie* nouvelles. Il s'agit d'un casse-tête, dans la mesure où un tel matériau, en raison de son ampleur, mais peut-être plus encore en raison du mode de collecte non supervisé, tranche radicalement avec les démarches classiques mobilisées en sciences sociales. Mais, l'accès à cette base de données est aussi apparu comme une opportunité potentiellement intéressante car, d'une part, autorisant des méthodologies nouvelles et, d'autre part, n'excluant pas le recours aux méthodologies déjà éprouvées de longue date.

Tout d'abord, si les données auxquelles nous avons accès sont lacunaires sur un certain nombre d'informations, elles sont à l'inverse plus complètes sur d'autres aspects, puisque renseignant sur l'ensemble du processus de signature depuis sa création. L'horodatage des signatures constitue l'exemple archétypique d'une information particulièrement utile dans l'analyse des dynamiques de mobilisation et auxquels les méthodes classiques d'analyse sociologique des pétitions (en ligne ou non), comme d'autres actions protestataires, ne permettent pas de répondre de manière pleinement satisfaisante. Plus généralement, le croisement de données habituellement inaccessibles constitue *a minima* un outil aux vertus heuristiques indéniables dans la construction argumentée d'hypothèses.

Le recours à ce type de matériau ne doit cependant pas aboutir à exclure les méthodes traditionnelles de traitement et d'analyse des données. À ce titre, l'étude détaillée de cas de pétitions, à la manière de ce qui a déjà pu être fait sur les pétitions sous forme papier mais aussi sur les pétitions en ligne, n'est aucunement remise en question et demeure au contraire facilitée sur certains aspects. C'est ainsi que, à travers l'étude de cas spécifiques de pétitions, nous avons pu produire de premiers résultats qui permettent de questionner l'existence de pratiques de « consommation politique » ainsi que de réseaux concrets dans les usages pétitionnaires.

Enfin, nous avons également montré que ces données demeurent pourtant insuffisantes pour répondre à nombre d'interrogations. Il est à ce titre nécessaire de compléter les données issues de la base par la mobilisation d'autres méthodes pouvant permettre de restituer adéquatement les pratiques des usagers du site et ce dans une démarche compréhensive. À cet égard, le recours au questionnaire, aux entretiens semi-directifs et aux analyses lexicométriques conservent tout leur intérêt.

Ainsi, le changement technologique appelle toujours à des commentaires fantasmés sur les transformations radicales qu'il serait censé apporter, que ces fantasmes soient vus comme une opportunité révolutionnaire pour les sciences ou comme la cause de son déclin annoncé. Au contraire, il nous semble plus raisonnable de considérer que le développement du « Big data » engendre en lui-même des transformations toutes relatives.

BIBLIOGRAPHIE

AGRIKOLIANSKY E., SOMMIER I., (eds) (2005), *Radiographie du mouvement altermondialiste*, Paris La dispute.

ALATHUR, S. (2007). « Citizen Participation and Effectiveness of E-petition: Sutharyakerlam –India », in *Transforming Government: People, Process and Policy*, vol.6, p. 392-403.

BARATS C. (2013), *Manuel d'analyse du web en sciences humaines sociales*, Paris, A. Colin, coll. U.

BARATS C., DISTER A., GAMBETTE P., LEBLANC J.-M., PERES-LEBLANC M., « Analyser des pétitions en ligne : potentialités et limites d'un dispositif d'étude pluridisciplinaire », (2016), in *Actes des Journées internationales d'Analyse statistique des Données Textuelles (JADT 2016)*, Nice, 7-10 juin 2016

BAYGERT N. (2014), « L'activisme numérique au regard du consumérisme politique : Pirates et *Tea Partiers* sous la loupe », in *Participations*, vol.8, p. 75-95.

BERNARD H.R., KILLWORTH P., KRONENFELD D., SAILER L. (1984), « The Problem of Informant Accuracy: The Validity of Retrospective Data », in *Annual Review of Anthropology*, vol.13, p. 495-517.

BOHLE, K., RIEHM, U. (2013), « E-petition Systems and Political Participation: About Institutional Challenges and Democratic Opportunities », in *First Monday*, vol.18, <http://firstmonday.org/ojs/index.php/fm/article/view/4220/3694>.

BOURDONCLE F. (2014), « Peut-on créer un écosystème français du Big Data ? », in *Le journal de l'école de Paris du management*, N° 108, p. 8-15.

BOURE R., BOUSQUET F. (2010), « Enjeux, jeux et usages d'une pétition politique. La pétition Vauzelle », in *Réseaux*, vol.164, p. 127-160.

BOURE R., BOUSQUET F. (2011), « La construction polyphonique des pétitions en ligne. Le cas des appels contre le débat sur l'identité nationale », in *Questions de communication*, vol.20, p. 293-316. BOURE R., BOUSQUET F., MARCHAND P. (2012), « Médiateurs et signataires des pétitions en ligne. L'exemple de trois pétitions sur l'identité nationale », in *Les Enjeux de l'information et de la communication*, n° 13, p. 99-118.

BOYADJIAN J. (2016), *Analyser les opinions sur internet. Enjeux théoriques et défis méthodologiques*, Paris, Dalloz.

BRIGHT J., BERMUDEZ S., PILET J.-B., SOUBIRAN T. (2016), « Power Users in Online Democracy: their origins and impact », working paper.

CARDON D. (2015), *À quoi rêvent les algorithmes. Nos vies à l'heure des big data*, Paris, Seuil, La République des idées.

CONTAMIN J.-G. (2001), *Contribution à une sociologie des usages pluriels des formes de mobilisation. L'exemple de la pétition en France*, Thèse, Université Paris 1.

CONTAMIN J.-G. (2013), « Petitions » in Snow D., Della Porta D., Klandermans B., McAdam D. (eds), *The Blackwell Encyclopedia of Social and Political Movements*, Oxford, Wiley-Blackwell.

DE LEEUW E., DE HEER W. (2002), « Trends in household survey nonresponse: a longitudinal and international comparison », in Groves R., Dillman D., Eltinge J., Little R. (eds), *Survey Nonresponse*, New York, Wiley, p. 41-54.

DELLA PORTA D. (eds) (2009), *Democracy in Social Movements*, London, Palgrave Macmillan.

DESROSIÈRES A., THÉVENOT L. (2002), *Les Catégories socio-professionnelles*, Paris, La Découverte.

DUMAS, C., LAMANNA, D., HARRISON T.M., RAVI, S.S., HAGEN, L., KOTFILA, C. and CHEN, F. (2015), « Examining Political Mobilization of Online Communities Through E-Petitioning Behavior in We the People », In *Big Data and Society*, vol.2, p. 1-26.

ESCHER T., RIEHM U. (2016), « Petitioning the German Bundestag: Political Equality and the Role of the Internet », In *Parliamentary Affairs* : <https://academic.oup.com/pa/article/doi/10.1093/pa/gsw009/2823719/Petitioning-the-German-Bundestag-Political>

FAVRE P., FILLIEULE O., MAYER N. (1997), « La fin d'une étrange lacune de la sociologie des mobilisations. L'étude par sondage des manifestants. Fondements théoriques et solutions techniques », *Revue Française de Science Politique*, vol.47, p. 3-28.

FELLEGI I., SUNTER A. (1969), « A theory for record linkage », in *Journal of the American Statistical Association*, vol. 64, p. 1183-1210.

GREFFET F. (2012), « Le web dans la recherche en science politique. Nouveaux terrains, nouveaux enjeux », in *Revue de la BNF*, n°40, p. 78-83.

- GROVES R., PEYTCHEVA E. (2008), « The impact of nonresponse rates on nonresponse bias a meta-analysis », in *Public Opinion Quarterly*, vol. 72, p. 167–189 .
- HAGEN L. et al. (2016), « E-petition popularity: Do linguistic and semantic factors matter? », in *Government Information Quarterly*, 33(4), p. 783-795.
- ION J. (1997), *La fin des militants ?*, Paris, Les Éditions de l'Atelier.
- JONSSON M., (2014), « Explaining deliberative participation in e-petition systems: the case of Malmö City », Conference paper, Conference Statsvetenskapliga förbundets metodinternat, Enköping, Sweden, June 9-10.
- JUNGHERR A., JÜRGENS P., (2010), « The Political Click: Political Participation through E-Petitions in Germany », In *Policy & Internet*, 2 (4), p. 131–164.
- LASSERRE F. (2000), « Les hommes qui voulaient être rois Principautés et nations sur Internet », in *Cybergeo : European Journal of Geography* URL : <http://cybergeo.revues.org/4397> ; DOI : 10.4000/cybergeo.4397.
- LATAPY M., MAGNIEN C., DEL VECCHIO N. (2008), « Basic notions for the analysis of large two-mode networks », in *Social Networks* , vol. 30, p. 31-48.
- MACINTOSH A., MALINA A., FARRELL S. (2002), « Digital Democracy through Electronic Petitioning », in W. J. McIver W.J., Elmagarmid A.K. (eds), *Advances in Digital Government*, Springer US, p. 137-148.
- MARGETTS H.Z. (2009), « The Internet and Public Policy », in *Policy & Internet*, 1 (1), p. 1–21.
- MARGETTS H.Z., JOHN P., HALE S.A., REISSFELDER S. (2015), « Leadership without leaders? Starters and followers in online collective action », in *Political Studies*, Vol.23, p. 278-299.
- MATHIEU L. (2013), « Un militantisme qui n'a de "nouveau" que le nom », in Frère B., Jacquemain M. (eds), *Résister au quotidien ?*, Paris, Presses de Science Po, p. 223-240.
- MILLER L. (2009), « E-Petitions at Westminster: the Way Forward for Democracy? », in *Parliamentary affairs*, vol. 62, no. 1, p. 162-178.
- MORVA O. (2016), « Are E-Petitions Operative for Change? On the Effectiveness and the Transformative Potential of E-Petitioning », Conference Paper, 22nd International Academic Conference, Lisbon, 22 March.
- MOSCA L., SANTUCCI D. (2009), « Petitioning Online: The Role of EPetitions in Web Campaigning », In Baringhorst S., Kneip V., Niesyto J. (eds), *Political Campaigning on the Web*, Bielefeld, Transcript, p. 121–146.
- NEIMAN M., GOTTDIENER M. (1982), « The Relevance of the Qualifying Stage of Initiative Politics : the Case of Petition Signing », in *Social Science Quarterly*, n°63, p. 582-588.
- PIERCE J.C., LOVRICH N.P. (1982), « Survey Measurement of Political Participation : Selective Effects of Recall in Petition Signing », in *Social Science Quarterly*, n°63, p. 164-171.
- OLLION É. et BOELAERT J. (2015), « Au-delà des big data. Les sciences sociales face à la multiplication des données numériques », in *Sociologie*, vol. 6.
- PANAGIOTOPOULOS P., SAMS S., ELLIMAN T., FITZGERALD G. (2011), « Do social networking groups support online petitions? », in *Transforming Government: People, Process and Policy*, Vol. 5 No. 1, p. 20-31.
- PENTLAND A. (2014), *Social Physics*, New York, Penguin
- PPC (Public Petitions Committee to the Scottish Parliament) (2006), « The Assessment of the Scottish Parliament's Public Petition System 1999–2006 », Study on behalf of the Scottish Parliament Information Centre for the Public Commissions Committee, Edinburgh, SP Paper 654.
- RUBIN D., LITTLE R. (2002), *Statistical analysis with missing data*, New York, Wiley.
- RUPPERT, E., LAW, J. et SAVAGE, M. (2013), « Reassembling social science methods: The challenge of digital », in *Theory, culture & society*, vol. 30.
- SCHLOZMAN K.L., VERBA S., BRADY H.E. (2010), « Weapon of the strong? Participatory inequality and the Internet », in *Perspectives on Politics*, vol.8, p. 487–509.

SCHMIDT J.-H., JOHNSEN K. (2014), « On the Use of the E-Petition Platform of the German Bundestag », in *SSRN Electronic Journal* : <http://dx.doi.org/10.2139/ssrn.2444672>.

SIMÉANT J., POMMEROLLE M.-E., SOMMIER I. (eds) (2015), *Observing Protest from a Place : the WSF in Dakar (2011)*, Amsterdam, Amsterdam University Press

WEISBERG H.F. (2005), *The Total Survey Error Approach: A Guide to the New Science of Survey Research*, Chicago, University of Chicago Press.

YASSERI T. HALE S.A., MARGETTS H.Z. (2013), « Modeling the Rise in Internet-based Petitions », working paper, consulté le 9/04/2017 : <http://arxiv.org/abs/1308.0239>.