



## Great expectations

Tommaso Venturini

► **To cite this version:**

Tommaso Venturini. Great expectations. Fourmentraux, Jean-Paul. L'ère post-média. Humanités digitales et cultures numériques, Hermann, pp.39-51, 2012. <hal-01064258>

**HAL Id: hal-01064258**

**<https://hal-sciencespo.archives-ouvertes.fr/hal-01064258>**

Submitted on 15 Sep 2014

**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.

# Great Expectations

## méthodes quali-quantitative et analyse des réseaux sociaux

Tommaso Venturini

### Les ennuis de la richesse

Depuis quelques années, les sciences sociales se retrouvent dans une situation tout à fait nouvelle. Relativement jeunes et encore précairement établies, ces sciences étaient loin de se doter des énormes machines à données des sciences naturelles. Contrairement aux physiciens jonglant avec des milliards de particules dans leurs accélérateurs ou aux biologistes cultivant des millions de microbes sous leurs microscopes, les sociologues ne pouvaient suivre que quelques centaines d'êtres humains et étaient condamnés à deviner la forme des phénomènes collectifs par ces aperçus partiels.

Face à cette difficulté, les sciences sociales avaient introduit dans leur monde une distinction qui s'adaptait fort bien à leur manque de données. En distinguant micro-interactions et macro-structures, les sociologues s'étaient creusés deux niches confortables. D'un côté, les qualitativistes pouvaient s'intéresser à la vie intime des communautés locales, qu'ils dessinaient comme bien spécifiques et bien délimitées. De l'autre, les quantitativistes pouvaient esquisser les grandes tendances globales sans se préoccuper des détails (Creswell, 2002). Bien sûr, tout le monde savait que la distinction était fictive (Latour, 2005) et que la plupart des phénomènes collectifs se passent entre ces deux niveaux (Giddens, 1986). Et pourtant, faute de données qui leurs auraient permis de retracer comment des milliers d'interactions locales s'assemblent pour former des structures globales, les sociologues se sont installés dans leurs niches micro ou macro.

Dans les dix dernières années, tout cela a été bousculé par l'arrivée des médias numériques. Les médias numériques ont une caractéristique intéressante : toutes les interactions qui les traversent y laissent des *traces* et ces traces peuvent être facilement enregistrées, conservées et retransmises. La traçabilité des médias numériques a des conséquences capitales, non seulement pour la vie privée des individus, mais aussi pour les sciences sociales (Lazer et al., 2009). Au fur et à mesure que le numérique infiltre les sociétés modernes, la vie collective devient de plus en plus traçable (Mitchell, 2009). Jour après jour, de nouveaux réservoirs de données numériques sont rendus accessibles aux chercheurs: les archives publiques et privées sont avalées par la mémoire des ordinateurs, les transactions économiques migrent en ligne, les réseaux sociaux s'enracinent dans le web. La médiation numérique s'étale comme un immense papier-carbone offrant aux sciences sociales plus de données qu'elles n'en ont jamais rêvé (Rogers, 2009).

Soudainement, les sciences sociales se retrouvent confrontées à autant de données que les sciences naturelles, mais avec une différence cruciale : les sciences sociales n'ont rien fait pour le mériter. Elles n'ont pas construit leurs radiotélescopes,

leurs microscopes électroniques, leurs séquenceurs. Les données numériques que les sciences sociales commencent à explorer ont été recueillies pour des finalités autres que la recherche scientifique. Il s'agit, au contraire, d'informations récoltées pour des besoins de marketing (comme dans le cas des cartes de fidélité ou des cartes bancaires), de surveillance (comme dans le cas des déplacements aériens), d'optimisation techniques (comme dans le cas des réseaux de télécommunication) ou tout simplement parce que leur coût est marginal (comme dans le cas de *logs* des serveurs Internet). En tout cas, il s'agit de *données d'occasion*, qu'investissent les sciences sociales sans que les chercheurs ne puissent maîtriser leur production et surtout sans que les chercheurs puissent s'y préparer. La situation des sciences sociales ressemble à celle de certains pays ruraux poussés à une brusque industrialisation par les pressions de l'économie internationale. Nées dans une époque de pénurie, les sciences sociales accèdent à l'âge de l'abondance trop vite et sans préparation.

Au médialab de Sciences Po, nous sommes coutumier du fait d'avoir trop de données. Pas un jour ne passe sans que quelqu'un de notre équipe ne tombe sur une nouvelle base de données en ligne : hier celle de l'OCDE ([stats.oecd.org](http://stats.oecd.org)), aujourd'hui celle de Twitter ([an.kaist.ac.kr/traces/WWW2010.html](http://an.kaist.ac.kr/traces/WWW2010.html)). Pas un jour sans qu'un de nos partenaires ne nous propose de partager la réflexion sur leurs données : l'Agence France Presse, IPSOS, Exalead, Linkfluence. Pas un jour sans qu'un chercheur de Sciences Po ne demande notre aide pour archiver les matériaux d'une enquête qualitative, traiter les données d'une négociation internationale, analyser les milliers de pages publiées par les institutions qui s'occupent de santé, de sécurité, de migration... Face à une telle abondance, on se sent comme des voleurs qui, ayant pénétrés dans le coffre-fort d'une banque, réalisent qu'ils n'ont pas les forces pour transporter leur butin.

Les données sont là, disponibles et copieuses comme elles ne l'ont jamais été, mais cela n'implique pas que le travail des chercheurs soit plus facile. Les données sont là, mais avant de les exploiter il faut se poser et leur poser toutes sortes de questions. Si la banque mondiale met en ligne ses données statistiques ([data.worldbank.org](http://data.worldbank.org)), encore faut-il chercher à savoir comment ces données ont été constituées et saisir les raisons de leur divulgation. Si American Online publie par erreur vingt millions de requêtes à son moteur de recherche ([en.wikipedia.org/wiki/AOL\\_search\\_data\\_scandal](http://en.wikipedia.org/wiki/AOL_search_data_scandal)), il faut se demander si l'utilisation de ces données est éthiquement correcte (Ess C. and AoIR ethics working committee, 2002). Si Wikipedia met à disposition une API pour télécharger tous ses articles (avec l'histoire complète de leur rédaction) ([mediawiki.org/wiki/API](http://mediawiki.org/wiki/API)), il faut réfléchir sur le statut épistémique des traces de cette entreprise collective (Viegas, Wattenberg, Kriss, & Van Ham, 2007).

Les données sont là, mais il faut savoir les extraire, les nettoyer, les indexer, les préparer à l'analyse. Il faut, en d'autres termes, les constituer en corpus. Le fait que de plus en plus de données soient disponibles grâce à la traçabilité du numérique ne

signifie pas qu'elles soient plus faciles à traiter. En ce sens on peut dire que les données ne sont jamais *données*, elles sont toujours construites par le travail des chercheurs. Pour utiliser les mots de Bruno Latour : « Décidément, on ne devrait jamais parler de "données" mais toujours d'"obtenues" » (Latour, 1993). Pour être exact, ce que les médias numériques offrent aux sciences sociales ne sont pas des données, mais des *traces*. Ces traces permettent d'obtenir énormément d'informations sur les phénomènes collectifs, mais au prix d'être transformées en données par les chercheurs qui les recueillent.

Enfin et surtout, les données sont là, mais il faut disposer d'outils d'analyse à la hauteur de leur abondance et pour cela les sciences sociales sont encore largement démunies. Face aux nouvelles données numériques les méthodes traditionnelles se révèlent complètement inadéquates. Développées dans une époque de pénurie, ni les méthodes quantitatives, ni les méthodes qualitatives ne semblent capables de gérer (et digérer) ce déluge de traces numériques.

### **Les méthodes quali-quantitatives**

En renversant sur les chercheurs une quantité sans précédents de traces, les médias numériques s'imposent à la réflexion sociologique. Aucun sociologue ne peut ignorer que

- a. De plus en plus de phénomènes sociaux se passent désormais principalement sur les réseaux numériques (par ex. là où les bases des données bibliographiques ont remplacé les catalogues papier, les bibliothèques peuvent tracer non seulement l'emprunt des livres, mais aussi leur recherche).
- b. Même les phénomènes qui ne se passent pas principalement sur les réseaux numériques y laissent souvent des traces (par ex. toutes les fois que nous achetons un billet d'avions avec une carte « fréquent flyers » ou que nous utilisons un système de télépéage sur une autoroute, nos déplacements physiques deviennent numériquement traçables).
- c. Les initiatives pour numériser les traces recueillies sur des médias non-numériques se multiplient (ce qui permet, par ex, à Google de nous offrir une volumétrie de l'utilisation littéraire de n'importe quel mot à partir du 1800 et jusqu'à aujourd'hui - [ngrams.googlelabs.com](http://ngrams.googlelabs.com), Michel et al., 2010).

Pourtant, si aucun sociologue ne peut ignorer les médias numériques, aucune des pratiques classiques de nos disciplines ne se prépare à la gestion de ce nouveau type de données. Comment faire face aux nouvelles traces? Comment les transformer en données ? Comment traiter sociologiquement les réseaux électroniques ?

Imaginons, par exemple, que nous nous intéressons au succès ou à la faillite des campagnes de vaccination. On sait que la campagne de vaccination contre la grippe H1N1 a eu des résultats très différents selon les pays (Lagarde & Door, 2010). Alors que dans les pays scandinaves une large partie de la population s'est soumise à la vaccination (64,5% en Suède, 45% en Norvège), les taux de l'Europe continentale sont beaucoup plus bas (10% en Allemagne, 6,6% en Belgique, 8,5% en France). La plupart

des observateurs attribuent cette disparité de résultats aux différentes dispositions des opinions publiques à l'égard des indications des autorités médicales nationales (Keck, 2010). Mais comment investiguer cette hypothèse ?

Jusqu'à il y a quelques années, les sociologues n'avaient que deux possibilités : administrer un questionnaire à un échantillon des populations concernées ou recueillir les discours de quelques individus particulièrement intéressants. Aucune de ces méthodes n'est exempte de difficultés. Le désavantage de la première méthode est que, afin de toucher un échantillon suffisamment large, il faut se contenter d'un questionnaire très simple, inadapté au caractère hétérogène et éphémère de l'opinion publique. Le désavantage de la seconde méthode est que, afin d'observer les interactions dans le temps et avec le détail nécessaire, il faut se contenter de suivre un nombre très limité d'individus, sans aucune garantie de leur représentativité. Jusqu'à il y a quelques années, il fallait choisir : peu d'informations sur beaucoup d'acteurs ou beaucoup d'informations sur peu d'acteurs ? *Tertium non datur*.

Aujourd'hui, les chercheurs n'ont pas besoin de se poser la question. Grâce aux traces numériques, ils peuvent observer dans les détails les interactions d'une vaste population d'individus. On peut par exemple recueillir sur le web toutes les conversations autour de la vaccination H1N1. Ou encore on peut s'adresser à un des nombreux services commerciaux offrant ce type de veille. Ici, par exemple, l'excellent service offert par la société Linkfluence, un des partenaires du médialab de Sciences Po :

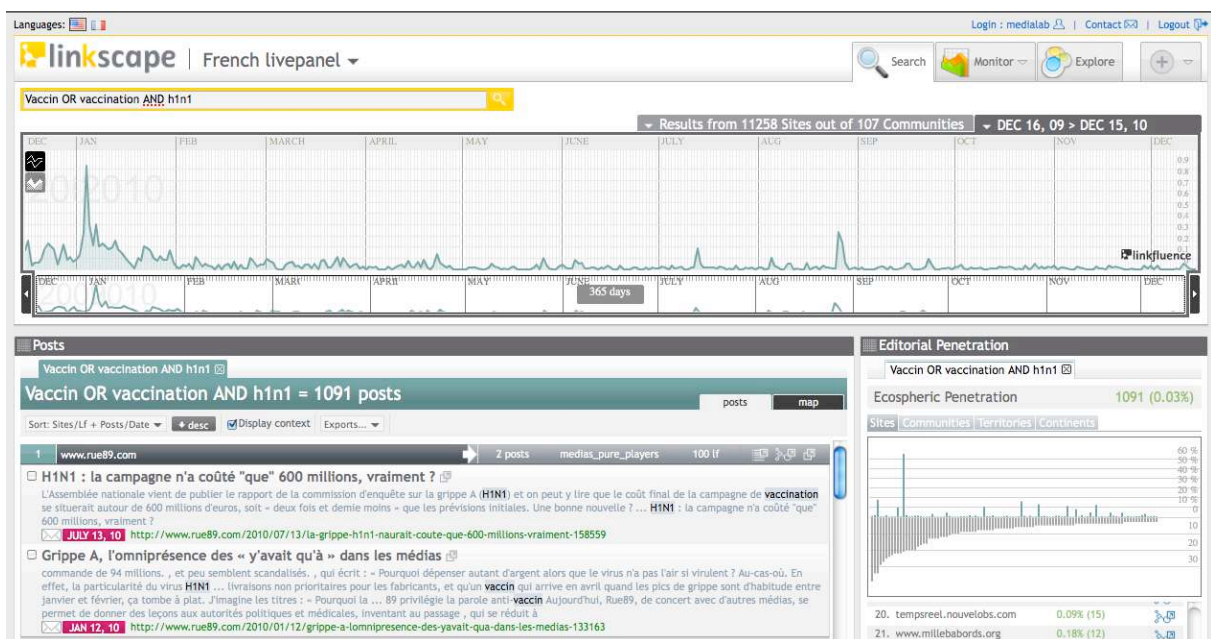


Fig 1. Exploration des billets des blogs français mentionnant le mot H1N1 et le mot vaccine ou vaccination (source linkfluence.net 19/08/10).

L'archive de Linkfluence recueille et sauvegarde tous les flux RSS des sites les plus influents de la blogosphère (plus de 14.000 sites pour les États Unis, 11.000 sites pour la France, 6.000 pour l'Angleterre et l'Allemagne, 5.000 pour l'Italie, 3.000 pour le Pays Bas...) et permet de chercher tous les billets mentionnant un ou plusieurs mots-clés sur une période d'un an. Grâce à cette archive, on a donc à disposition la

(quasi) intégralité du débat en ligne sur un sujet donné. Mais comment exploiter cette masse de données ?

Ni les méthodes quantitatives ni les méthodes qualitatives ne semblent capables de profiter pleinement de l'abondance de ces nouvelles traces numériques. Avec les méthodes quantitatives on pourrait faire émerger quelques tendances globales : démontrer, par exemple, qu'il y a corrélation entre les taux de vaccination et le débat public sur le web autour des risques de la maladie ou du vaccin. Toutefois, l'agrégation statistique rend difficile de revenir aux *verbatim* pour interroger le sens des corrélations observées. Inversement, les méthodes qualitatives pourraient nous aider à identifier les arguments pour et contre la vaccination dans un échantillon de billets, mais sans aucune garantie de pouvoir généraliser les résultats.

Pour mener à bien notre exploration du débat public autour de la vaccination H1N1, on doit imaginer une nouvelle approche : une méthodologie quali-quantitative capable de réunir les avantages des deux types de méthodes. Tant que les chercheurs ne disposent pas d'une nouvelle génération de méthodes capable de combler la discontinuité entre micro et macro, les sciences sociales seront incapables de profiter de l'abondance des traces numériques. Développer ce nouveau type de méthodes, permettant de dépasser la distinction micro/macro et de retracer comment les phénomènes globaux sont construits par la coordination d'une multiplicité d'actions locales, telle est précisément l'ambition du médialab de Sciences Po.

Admettons, mais à quoi ressemble une méthode quali-quantitative ? Pour illustrer le potentiel d'innovation de cette démarche, revenons à notre exemple du débat autour de la vaccination H1N1. Comme on l'a vu, le succès de ces campagnes de vaccination contre la grippe est très variable et semble dépendre d'interactions complexes entre les indications des autorités médicales et les réactions de l'opinion publique. L'analyse des discours sur le web peut être très utile pour interpréter ces réactions collectives, à condition toutefois de mettre en place un dispositif de recherche capable de lier l'exploration des tendances globales à l'observation détaillée des interactions remarquables. Dans le cas des vaccinations contre les pandémies grippales, une méthode quali-quantitative pourrait se composer des étapes suivantes :

1. Combiner les données sur l'adhésion aux campagnes de vaccination avec les traces des discussions en ligne sur le vaccin H1N1.
  - Désagréger les deux séries de données et les projeter sur la même échelle spatio-temporelle.
  - Comparer les deux séries de données en affichant sur le même graphe le taux de vaccination et la volumétrie des discours mentionnant le mot « vaccin ».
2. Zoomer sur l'analyse d'un échantillon non-aléatoire des discours web.
  - Identifier les points critiques : les temps et lieux où les discours web et l'adhésion aux campagnes de vaccination semblent être le plus liés.

- Identifier les discours particulièrement visibles dans les points critiques et suivre leur diffusion et transformation dans la blogosphère.
- 3. Généraliser les intuitions venant de l'analyse de l'échantillon choisi.
  - Proposer une interprétation des dynamiques influant sur l'adhésion à la vaccination basées sur les phénomènes observés dans les points critiques.
  - Traduire l'interprétation proposée en indicateurs mathématiques qui puissent être calculés pour le corpus entier.
- 4. Valider (ou invalider) la solidité et l'intérêt des indicateurs construits
  - Calculer les indicateurs construits pour le corpus entier pour évaluer leur généralité.
  - Employer les indicateurs construits pour identifier d'autres points critiques où les indicateurs ne s'appliquent pas ou s'appliquent particulièrement bien.

La procédure décrite, évidemment, ressemble moins à une séquence qu'à une boucle, puisque les résultats de la dernière phase peuvent (et même doivent) relancer un nouveau cycle de *zooming-in* et *zooming-out*. L'exemple que nous avons donné illustre bien la particularité des méthodes quali-quantitatives : l'oscillation incessante entre agrégation et désagrégation qui efface toute distinction entre micro et macro. Il n'est pas difficile d'imaginer des méthodes de ce type pour beaucoup d'autres phénomènes sociaux et, grâce à la traçabilité numérique, les données ne manquent pas. Toutefois, pour que ce genre d'approche puisse effectivement entrer dans les pratiques des sciences sociales, il faudrait disposer d'outils conceptuels, mathématiques et informatiques aussi solides et standards que ceux des méthodes traditionnelles. Pour le moment, nous en sommes encore loin : la plupart des méthodes quali-quantitatives restent des efforts isolés et liés à des cas d'étude et des questions de recherche spécifiques. Cependant il y a au moins une approche qui semble correspondre aux ambitions d'une sociologie quali-quantitative. Il s'agit de l'analyse de réseaux, une approche qui a rapidement colonisée plusieurs disciplines dont les sciences sociales.

### **Réseaux, graphes, cartes, interfaces**

Bien que les bases mathématiques de l'analyse des réseaux existent depuis la célèbre promenade d'Euler sur les ponts de Königsberg (*Solutio problematis ad geometriam situs pertinentis*, 1736), ce n'est qu'avec les sociogrammes de Jacob L. Moreno (Moreno, 1934) que l'analyse de réseaux entre dans l'équipement des sciences sociales. Depuis quelques années, cette approche a connu une diffusion vertigineuse grâce à deux développements parallèles :

1. La mise au point d'une série des techniques mathématiques permettant d'analyser des réseaux aussi complexes que ceux qu'on observe dans les phénomènes collectifs (Newman, Barabasi, & Watts, 2006).
2. La disponibilité croissante de logiciels (parfois gratuits et *open source*) pour l'analyse et la visualisation des graphes comme Pajek, Ucinet, Gephi (Combe, LARGERON, & EGYED-ZSIGMOND, 2010).

Grâce à ces deux développements, les sciences sociales commencent finalement à disposer d'outils conceptuels et techniques capables de gérer la masse croissante de traces numériques. C'est donc vers les réseaux que s'adresse une partie des espoirs de développer des méthodes quali-quantitatives et aussi la plupart des recherches conduites au médialab de Sciences Po. Les techniques d'analyse de réseaux, bien sûr, ne sont pas les seules techniques quali-quantitative possibles et encore moins le seul moyen d'exploiter les traces numériques. Pour le moment, elles représentent néanmoins le jeu d'outils le plus abouti à disposition des nouvelles sciences sociales. C'est pour cette raison que nous leur dédions le reste de cet article.

Les réseaux sont des objets hybrides qui rassemblent les avantages des graphes, des cartes et des interfaces. En tant que graphes, les réseaux ont l'avantage de focaliser l'attention des chercheurs sur le phénomène élémentaire de la vie collective : l'association et la dissociation des acteurs sociaux. Dans ce sens, l'analyse des réseaux réalise l'invite de la Théorie de l'Acteur-Réseau à refonder la sociologie comme science des associations (Latour, 2005). Au lieu de prendre pour acquise l'existence des groupes, des institutions, des structures sociales, l'analyse de réseau ne postule que l'existence d'acteurs (nœuds ou sommets) et leur capacité à s'attacher ou se détacher (arcs ou arêtes).

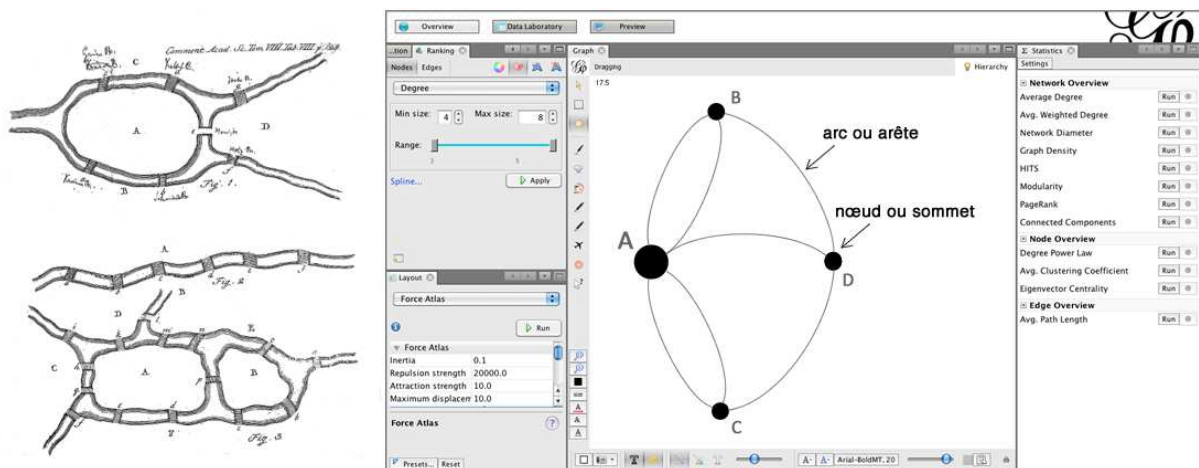


Fig 2. Le réseau des ponts de Königsberg dans la représentation originale d'Euler et dans l'interface de Gephi.

La distinction entre micro et macro n'existe pas dans le monde des graphes, puisque les phénomènes globaux sont inséparables des connexions locales. Les propriétés globales d'un réseau n'émergent pas de la somme des associations locales (au sens de la théorie des systèmes), tout simplement parce que les associations d'un graphe ne s'additionnent pas. Considérez, par exemple la plus simple des propriétés d'un graphe – son diamètre. Défini comme la plus longue des distances entre deux nœuds, le diamètre d'un graphe n'a aucune existence autonome de ses arcs. Il peut suffire de changer un de ces arcs pour que le diamètre soit doublé ou réduit de moitié. Bien sûr cela dépend de quel arc on modifie. La modification de chaque nœud ou de chaque arc peut avoir un effet différent sur la forme du graphe et, pourtant, cela ne dépend que de son profil de connectivité. Un nœud dans un graphe



n'est défini que par ses relations et une relation n'est définie que par les nœuds qu'elle connecte. Dans un graphe, il n'y a pas de macro-structures, ni de micro interactions, mais seulement une organisation à la fois locale et globale.

Plus important encore, les graphes permettent une économie conceptuelle. N'étant formés que de nœuds et d'arcs, les graphes ont permis le développement d'une puissante mathématique discrète. Grâce à la formalisation de ses techniques, l'analyse de graphes peut s'appliquer à une multitude de phénomènes différents : des dynamiques sociales aux réseaux biologiques, des réactions chimiques aux forces physiques (Barabási, 2003). Cela a encouragé le développement d'un espace de collaboration multidisciplinaire, en favorisant l'échange des algorithmes, des logiciels et des pratiques d'analyse<sup>1</sup>.

De plus, les réseaux (et particulièrement les réseaux sociaux) ne sont pas que des graphes, ils sont aussi des cartes. En tant que cartes, les réseaux héritent de deux notions essentielles de la tradition géographique : la notion de proximité et la notion de limite. Bien sûr, ces deux notions sont profondément redéfinies par l'analyse de réseaux. À la différence des cartes géographiques, les graphes ne sont pas des projections sur un fond de carte prédéterminé. Dans un réseau, la position d'un nœud n'est pas définie par des coordonnées spatiales. C'est au contraire l'espace même du graphe qui est défini par les relations qu'il contient. Néanmoins, afin de maximiser la lisibilité et minimiser les croisements des arcs, la plupart des logiciels de visualisation de graphes sont conçus pour rapprocher les nœuds connectés et éloigner les nœuds non-connectés. Les algorithmes de spatialisation fonctionnent en attribuant une force de répulsion aux nœuds (chacun d'eux tend à se distancer le plus possible des autres) et une force d'attraction aux arcs (qui, comme des élastiques, rapprochent les nœuds qu'elles connectent).

Le résultat est que, dans un graphe spatialisé, deux nœuds sont proches s'ils sont directement liés ou s'ils ont un profil de connexion semblable (ils sont connectés aux mêmes nœuds sans forcément être connectés entre eux). La proximité dans un graphe spatialisé n'est donc pas dépourvue d'information. Comme dans une carte géographique, la contiguïté spatiale indique souvent d'autres types de proximité (la facilité de passer d'un nœud à l'autre, l'appartenance au même regroupement, une certaine air de famille...).

De la même manière, il est possible d'identifier des frontières dans un réseau. Souvent, les espaces dessinés par les réseaux présentent une densité variable : les nœuds ne sont pas distribués d'une façon homogène dans l'espace, mais rassemblés dans des régions densément connectées séparées par des régions moins densément connectées. On appelle clusters, les regroupements des nœuds qui sont plus connectés entre eux que vers l'extérieur. Les clusters peuvent être identifiés par des algorithmes mathématiques (comme le calcul de la modularité) et leurs frontières

---

<sup>1</sup> Aux Etats-Unis le travail de construction de cet espace de partage à été mené surtout par la Network Science Collaborative Technology Alliance (NS CTA), <http://www.ns-cta.org>.

sont généralement visibles dans un graphe bien spatialisé.

Le fait que la proximité des nœuds soit significative et qu'il soit possible d'observer les limites des clusters permet de lire les réseaux comme des cartes et de leur appliquer une partie des conventions d'interprétation de la géographie. Par exemple, on retrouve dans les réseaux tous les éléments qui constituent, selon Kevin Lynch, l'image de la cité : les voies, les limites, les quartiers, les nœuds, les points de repères (Lynch, 1960). Les réseaux ne montrent que des connexions et, pourtant, ils dessinent des territoires, bien que des territoires de type nouveau (Boullier, 2009). La fascination que les réseaux exercent sur les chercheurs émane aussi de la possibilité de les explorer comme on explore des cartes géographiques. Avant les réseaux, on pouvait décrire les groupes sociaux. Avec les réseaux, on a un moyen pour les montrer. Bien sûr il s'agit d'une représentation aussi artificielle et conventionnelle qu'une description textuelle et pourtant la possibilité de donner une représentation graphique des phénomènes sociaux ouvre des possibilités nouvelles pour leur analyse (Ghitalla, 2009).

Enfin, les réseaux qu'on dessine avec les logiciels d'analyse de réseaux peuvent aussi servir d'interface de navigation dans les données. Étant des structures logiques relativement simples accompagnées d'une mathématique très puissante, les réseaux se prêtent particulièrement bien à un traitement informatique. Depuis quelques années, les solutions numériques pour la visualisation et l'analyse des réseaux se multiplient sous toutes les formes et toutes les licences, tout en devenant de plus en plus performantes et faciles à utiliser. Les logiciels pour l'exploration des graphes sont aujourd'hui si avancés qu'ils permettent non seulement de spatialiser rapidement des réseaux de dizaines de milliers de nœuds, mais aussi de les manipuler en temps réel. Avec ces logiciels, le travail de *zooming-in* et *zooming-out* qui caractérise les méthodes quali-quantitatives devient facile et instantané. Et le zoom ce n'est qu'une des multiples possibilités d'interagir avec les graphes offertes par ces logiciels.

Un logiciel gratuit et open source comme Gephi ([gephi.org](http://gephi.org)) permet, par exemple, de gérer de graphes de tout type (simples, orientés, pondérés) et de toute dimension. Il offre des dizaines d'algorithmes de spatialisation et permet facilement d'en rajouter de nouveaux. Il permet de changer la taille et la couleur des nœuds et des arcs manuellement ou sur la base d'une multitude de variables. Il facilite la création de requêtes complexes pour filtrer les graphes. Il permet de grouper ou dégrouper les nœuds en un clic. Il calcule tous les indicateurs de base de la mathématique des graphes et permet de les utiliser comme paramètres de la visualisation. Enfin, il exporte les résultats des analyses en plusieurs formats statiques et dynamiques (Bastian, Heymann, & Jacomy, 2009).

La possibilité de modifier des centaines de paramètres en un clic et d'observer en temps réel leurs conséquences sur la forme des réseaux fait des logiciels de manipulation de graphes des outils idéaux pour l'analyse exploratoire des données (Tukey, 1977). Bien sûr, la plupart des analyses offertes par ces logiciels pourraient être faites à la main ou avec des outils traditionnels, mais l'aisance offerte par des

logiciels comme Gephi est impressionnante. Aucun des outils de sciences sociales classiques ne permet d'agrèger et désagrèger avec une telle fluidité. Clic : milliers d'acteurs et de relations apparaissent sur l'écran. Clic : la position des acteurs est réarrangée selon leur profil de connectivité. Clic : les positions cruciales prennent le devant de la scène. Clic : l'interface zoome sur un seul acteur et ses relations. Click : l'acteur est remis dans le cluster qui le contient. Clic : les acteurs sont colorés sur la base des groupes auxquels ils appartiennent. Clic : la scène zoome sur les détails d'un cluster.

Désagrèger les données d'une enquête quantitative classique, bien sûr, a toujours été possible, mais il fallait passer des résultats de l'enquête au tableau des données, du tableau aux questionnaires, des questionnaires aux notes de l'investigateur, des notes aux entretiens. Également, agréger les résultats d'une ou plusieurs enquêtes qualitatives a toujours été possible, mais il fallait passer des entretiens aux transcriptions, des transcriptions à une grille de codage unique, de la grille au tableau de données, du tableau aux indicateurs statistiques. Les approches quali-quantitatives n'ont jamais été impossibles. Tout simplement, il fallait trop de temps pour les mettre en place.

Grâce à l'abondance de traces numériques et à la souplesse des réseaux, on dispose des données et des outils nécessaires pour observer le tissu social dans toute son extension et, dans le même temps, de suivre chacun des fils qui le constituent. Pour la première fois dans l'histoire des sciences sociales, les chercheurs ne sont plus obligés de choisir entre l'ampleur et la finesse de leurs analyses, mais peuvent poursuivre les deux simultanément. La promesse des méthodes quali-quantitatives semble enfin à la portée des sciences sociales.

Tout en célébrant les avantages des réseaux comme graphes, cartes et interfaces d'exploration de la complexité sociale, nous ne voulons pas cacher leurs inconvénients. Le plus sérieux de ces inconvénients est la difficulté que l'analyse de réseaux rencontre encore dans l'exploration de la *dynamique* des phénomènes sociaux. Ni la mathématique de graphes, ni la sémiotique de cartes n'ont été développées pour traiter les phénomènes évolutifs. Par conséquent, la plupart des outils pour la manipulation de réseaux gèrent encore relativement mal la variable 'temps', obligeant les chercheurs à découper les phénomènes qu'ils étudient dans une série de clichés successifs. Bien que cette technique puisse suffire pour étudier les transformations plus simples, elle se révèle souvent inadéquate pour explorer les dynamiques complexes qui caractérisent la vie collective.

Des extensions de la théorie de graphes ont été proposées pour surmonter cette difficulté (par exemple, les réseaux de Petri), mais dans la plupart des cas on se limite à l'analyse de la circulation d'éléments mobiles à l'intérieur d'un réseau statique, sans se demander comment cette circulation peut rétroagir sur l'organisation des connexions. Par exemple, on étudie la diffusion des discours dans la blogosphère, comme on étudie la transmission des informations sur Internet (routing), sans s'apercevoir qu'il s'agit, en effet, de deux phénomènes complètement différents. Là

où les paquets TCP-IP se déplacent sur un réseau de connexions relativement stables (et encore !), la circulation des discours définit l'articulation même de la blogosphère (Leskovec, Kleinberg, & Faloutsos, 2005). En séparant les flux de connexions, on ne fait que reproduire la distinction entre une structure relativement stable et les interactions qui la traversent : précisément la distinction micro/macro que les réseaux promettent de dépasser.

A cause de ces difficultés, les réseaux, tout en étant des instruments extraordinaires d'exploration, restent encore inefficaces en tant qu'outils de narration. L'exploration, l'analyse, le déploiement ne sont qu'une partie de la recherche scientifique. L'autre partie est l'effort de description et représentation qui passe par le récit d'une *histoire*. Cette fonction est particulièrement importante pour les sciences sociales qui ne peuvent pas se limiter à modéliser leurs sujets d'études, mais doivent aussi donner du sens à leurs modèles en les intégrant dans une narration *crédible*. Pour cela les réseaux restent encore relativement inutiles. Dans la situation actuelle, personne ne sait comment raconter une histoire avec un réseau. Pour le moment, le mieux qu'on puisse faire avec les réseaux c'est d'accompagner ou illustrer le récit d'une histoire textuelle. Le langage graphique des réseaux reste encore trop primitif pour assumer des fonctions narratives. La plupart des logiciels se limitent à proposer quelques éléments hérités de la sémiotique photographique ou cinématographique (le zoom, le cadrage, les mouvements de camera...) (Metz, 1990) sans vraiment réfléchir à leur signification dans le monde des réseaux. Cela limite encore fortement l'utilisation des réseaux dans les sciences sociales. Pour maintenir leur promesse, autre que des graphes, des cartes et des interfaces les réseaux doivent devenir aussi des outils de narration. Le chantier des réseaux est ouvert et, malgré les grandes espérances que les graphes ouvrent aux sciences sociales, la route vers les méthodes quali-quantitatives reste encore longue.

## Bibliographie

- Barabási, A.-L. (2003). *Linked: how everything is connected to everything else and what it means for ...* (p. 294). Plume. Retrieved from <http://books.google.com/books?id=rydKGwfs3UAC&pgis=1>.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media* (pp. 361-362). Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/download/154/1009>.
- Boullier, D. (2009). Au-delà des territoires numériques en dix thèses. In F. Rowe (Ed.), *Sociétés de la connaissance et prospective. Hommes, organisations et territoires*. Nantes: Lemna.
- Combe, D., LARGERON, C., & Egyed-Zsigmond, E. (2010). A comparative study of social network analysis tools. *Social Networks*, 2, 1-12. Retrieved from <http://hal.archives-ouvertes.fr/hal-00531447/>.
- Creswell, J. W. (2002). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (2nd Edition)* (p. 246). Sage Publications, Inc. Retrieved from <http://www.amazon.com/Research-Design-Qualitative-Quantitative-Approaches/dp/0761924426>.

- Ess C. and AoIR ethics working committee. (2002). Ethical decision-making and Internet research. Recommendations from the AOIR ethics working committee. Retrieved from [www.aoir.org/reports/ethics.pdf](http://www.aoir.org/reports/ethics.pdf).
- Ghitalla, F. (2009). La « Toile Européenne » Parcours autour d'une cartographie thématique de documents web consacrés au thème de l'Europe et à ses acteurs sur le web francophone. *Communication & langages*, 2008(158), 61. doi: 10.4074/S0336150008004067.
- Giddens, A. (1986). *The Constitution of Society: Outline of the Theory of Structuration* (p. 417). University of California Press. Retrieved from <http://www.amazon.com/Constitution-Society-Outline-Theory-Structuration/dp/0520057287>.
- Keck, F. (2010). *Un monde grippe* (Flammarion.). Paris, France.
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., et al. (2010). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014), 176-182.
- Lagarde, J.-C., & Door, J.-P., et al. (2010). *Rapport fait au nom de la Commission d'enquête sur la manière dont a été programmée, expliquée et gérée la campagne de vaccination contre la grippe A (H1N1)*.
- Latour, B. (1993). Le topofil de Boa Vista ou la référence scientifique–montage photo-philosophique. *Raison Pratique*, 4, 187–216. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Le+topofil+de+Boa+Vista+ou+la+référence+scientifique+montage+photo-philosophique#0>.
- Latour, B. (2005). *Reassembling the Social* (Oxford Uni.). Oxford.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., et al. (2009). Computational social science. *Science (New York, N.Y.)*, 323(5915), 721-3. doi: 10.1126/science.1167742.
- Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005). Graphs over time: densification laws, shrinking diameters and possible explanations. *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (p. 177–187). New York, New York, USA: ACM. doi: 10.1145/1081870.1081893.
- Lynch, K. (1960). *The image of the city*. Cambridge Mass.: MIT Press. Retrieved from [http://books.google.com/books?hl=it&lr=&id=\\_phRPWSpAgC&pgis=1](http://books.google.com/books?hl=it&lr=&id=_phRPWSpAgC&pgis=1).
- Metz, C. (1990). *Film Language: A Semiotics of the Cinema* (p. 286). University Of Chicago Press. Retrieved from <http://www.amazon.com/Film-Language-Semiotics-Christian-Metz/dp/0226521303>.
- Mitchell, T. M. (2009). Mining our reality. *Science (New York, N.Y.)*, 326(5960), 1644-5. doi: 10.1126/science.1174459.
- Moreno, J. (1934). *Who Shall Survive?* (Nervous an.). Washington, DC.
- Newman, M., Barabasi, A.-L., & Watts, D. J. (2006). *The Structure and Dynamics of Networks: (Princeton Studies in Complexity)* (p. 624). Princeton University Press. Retrieved from <http://www.amazon.com/Structure-Dynamics-Networks-Princeton-Complexity/dp/0691113572>.
- Rogers, R. (2009). *The End of the Virtual: Digital Methods* (p. 36). Amsterdam University Press. Retrieved from <http://books.google.com/books?id=ZHFis5sEAicC&pgis=1>.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading: Addison-Wesley.
- Viegas, F. B., Wattenberg, M., Kriss, J., & Van Ham, F. (2007). Talk before you type: Coordination in Wikipedia. *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on* (p. 78). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4076527](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4076527).