



Quand l'informatique observe les réseaux

Alain Lelu

► To cite this version:

Alain Lelu. Quand l'informatique observe les réseaux. Eric Letonturier. Les Réseaux, CNRS Editions, pp.101-108, 2012, Les Essentiels d'Hermès, 978-2-271-07509-3. hal-00733075

HAL Id: hal-00733075

<https://hal.archives-ouvertes.fr/hal-00733075>

Submitted on 17 Sep 2012

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.

In « Les Réseaux », coord. Eric Letonturier, pp.101-108, collection « Les Essentiels d'Hermès », CNRS Editions, Paris, 2012

Quand l'informatique observe les réseaux

Alain Lelu

La généralisation tous azimuts de l'informatique donne accès à un nombre croissant d'entités qui « font réseau » : les pages Web pointent vers d'autres pages Web, les articles scientifiques citent d'autres articles scientifiques, les personnes inscrites sur un réseau social en ligne définissent leurs « amis », ou ceux qu'ils « suivent », etc. Au-delà de ces réseaux dont la structure consiste en un ensemble de liens explicites entre entités, et rien de plus, beaucoup d'autres réseaux peuvent être déduits à partir de *traces*. Ainsi la sociologie des sciences selon Callon, Latour et leurs continuateurs a-t-elle pu observer les réseaux d'alliances entre scientifiques, les reconfigurations de disciplines, les enrôlements de nouveaux acteurs, à partir du vocabulaire utilisé dans les titres et résumés d'articles ou brevets. A la limite, *toute trace informatique peut faire réseau*, comme l'ont compris les promoteurs des techniques de marketing individualisé : dis-moi ce que tu as consommé, et je te dirai ce qui pourra te plaire, compte tenu de ce qui a plu aux personnes de profil de consommation voisin du tien – de ce qui a plu à tes « voisins de réseau » virtuels, en quelque sorte. Ainsi fonctionnent les systèmes de recommandations (*recommender systems*) apparus dans la recherche informatique académique aux Etats-Unis dans les années 1990, aujourd'hui en pleine phase de mise en pratique sur la Toile.

Qui analyse les réseaux ?

La croissance ininterrompue du volume de ces données de réseau, explicite ou implicite, fait de leur analyse et de leur représentation un enjeu scientifique et technique majeur. Plusieurs spécialités scientifiques s'attaquent à ce défi, avec des approches parfois complémentaires, parfois concurrentes :

- Les mathématiciens développent des approches issues de la théorie des graphes. Par exemple l'efficacité de Google s'appuie sur l'indicateur de popularité PageRank¹, qui s'exprime en termes mathématiques comme le « vecteur singulier dominant » de l'immense matrice d'adjacence de la Toile entre pages Web pointantes et pages Web pointées.
- Les statisticiens s'intéressent aux lois de répartition des nombres de liens par « nœuds » dans un réseau, et aux propriétés-limites des réseaux aléatoires².
- Les physiciens, nouveaux venus dans le domaine, ne sont pas les moins bien armés. Ils s'appuient sur leur expérience séculaire de la prise en compte des interactions entre des milliards d'« individus » élémentaires dans les solides et les fluides. On leur doit les avancées récentes les plus remarquables, comme la découverte de la généralité et de l'omniprésence des structures de réseau *petit monde*³, tout comme celles des répartitions dites Zipfiennes⁴ de liens dans les réseaux, tous principes qui traversent les frontières des disciplines – on les retrouve en étudiant la longueur des fleuves, la taille des villes, les associations de mots dans les textes, les interactions entre gènes, les liens sur la Toile...
- Les informaticiens, moteurs dans ces domaines, piétinent sans complexe les plates-bandes de ces spécialités vénérables. Ils introduisent souvent sans détours et de façon candide les demandes managériales comme « identifier parmi les clients d'une banque les mauvais payeurs potentiels », « quels films suggérer à un abonné

¹ Google attribue à chaque page Web une note de notoriété, calculée automatiquement. En première approximation il s'agit du nombre de liens qui pointent vers cette page, qu'on corrige de façon itérative en pondérant chaque lien par les notes des pages pointantes : *une page a d'autant plus de notoriété que les pages qui pointent vers elle en ont*. On montre mathématiquement que ce processus converge vers une note stable pour chaque page. Cette description succincte passe sur certains détails mathématiques (connus), et sur beaucoup d'ajustements (non divulgués) de la part de Google.

² Etant donné un ensemble d'entités, un réseau aléatoire est obtenu en tirant au hasard des liens entre ces entités. Selon la loi de répartition choisie pour le nombre de liens par entité, divers types de réseaux sont possibles.

³ Il suffit d'une chaîne de six ou sept « amis d'amis » pour relier n'importe quel individu sur terre à n'importe quel autre, fût-il le Pape ou le président des Etats-Unis...

⁴ Sur la base de constats antérieurs G.K. Zipf a exprimé dans les années 1930 que, quelle que soit la langue d'un corpus textuel, l'emploi des mots obéissait à une loi très inégalitaire, dont la formulation suivante donne une idée : « quand le mot le plus fréquent revient 10 000 fois, alors le dixième plus fréquent se répète environ 1000 fois, le centième 100 fois, le millième 10 fois, etc. ».

dont on connaît les DVD loués antérieurement ? », « A quels clients un opérateur téléphonique peut-il espérer proposer avec succès des services supplémentaires ? », ... En bousculant les frontières et les points de vue, ils créent de nouvelles disciplines, comme l'apprentissage artificiel, la fouille de données, ou la visualisation de grands graphes. Toute une industrie de l'observation et de l'analyse des réseaux (*network analytics*) se met en place dans la foulée, emmenée par des start-up innovantes, comme LinkFluence en France, bientôt rejointes par les poids-lourds de l'informatique – le groupe Cap-Gemini vient de conclure un accord avec la start-up américaine Attensity pour offrir un service de suivi et d'alerte sur « ce qui se dit » sur les réseaux sociaux en ligne au sujet de tel ou tel produit ou entreprise cliente.

- Les sciences humaines et sociales sont à la fois pionnières et peu motrices dans ce domaine : pionnières car l'effet « petit monde », entre autres, a été découvert par des psycho-sociologues dans les années 1960, et l'indicateur de notoriété de type PageRank l'a été dans le cadre des publications scientifiques par Eugene Garfield, un grand nom de la science des bibliothèques, suivi par le sociologue Philip Bonacich dans les années 1970 ; peu motrices car leur réticence générale envers les représentations formelles laisse le champ libre aux informaticiens ou physiciens pour ce qui est de l'analyse fine des mécanismes de propagation de rumeurs ou d'opinions, par exemple. En minorité, ceux qui utilisent des méthodes formelles le font souvent à partir d'un logiciel de référence qui a su s'imposer de longue date dans une « communauté » de recherche particulière : ainsi *Pajek* (de Vlad Batagelj) pour l'analyse des réseaux sociaux, depuis les années 1990, ou *Alceste* (de Max Reinert) pour l'analyse de textes, un peu antérieur.

Dans les tuyaux

A moyen terme, au delà de l'explosion prévisible des applications commerciales et sécuritaires de la « science des réseaux », on peut s'attendre à ce que les systèmes de recommandation s'étendent à la suggestion d'« amis », ou de façon plus ciblée, à celle de partenaires scientifiques, individus ou laboratoires, selon le modèle élitare de la « Google Science » - transposant à l'échelle mondiale et sur support électronique le modèle séculaire d'organisation de l'industrie audiovisuelle, où l'on s'efforce de rassembler « les bons » autour d'un projet précis, à durée limitée, financé ou à la recherche de financements.

A plus long terme de nouveaux défis apparaissent, auxquels sont consacrés une multitude de colloques, d'ateliers et de numéros spéciaux de revues. On peut citer :

- celui d'introduire la dimension temporelle : après avoir détecté des agrégats sur des réseaux immenses, identifié des individus « leaders », ou des « connecteurs » entre groupes, il s'agit maintenant de décrire comment se font, se défont, et évoluent ces groupes, voire de le prévoir.

- celui d'extrapoler les « appétences » de consommateurs dans un domaine sur lequel on ne dispose pas de données, après avoir « appris » leurs comportements dans un ou plusieurs autres domaine(s).

- celui d'extraire les modèles directement à partir des données, ce qui constituerait un changement de paradigme scientifique si cet effort aboutissait. Ainsi les *Deep Neural Networks* tentent de simuler le traitement multi-échelle de l'information réalisé en permanence par l'esprit humain, où les informations « de bas niveau » (phonèmes, formes des lettres, ...) sont intégrées dans plusieurs niveaux supérieurs successifs (mots, phrases, concepts, sens, ...), au sein d'un seul et même processus.

Ce type de recherche fait preuve d'un grand dynamisme, puisé sans doute aussi bien dans les fantasmes millénaires de re-création de la vie et de l'intelligence, que dans des considérations plus prosaïques, comme la possibilité de trouver des financements autour de problématiques posées par les industriels. A noter que le caractère d'ingénierie de beaucoup de ces recherches stimule l'organisation de compétitions, de *challenges*, aux résultats mesurables, ouverts à tous, scientifiques ou non. Ainsi en Suisse une équipe du laboratoire IDSIA de Jürgen Schmidhuber peut-elle faire état de sa première place en 2011 à la *Traffic Sign Recognition Competition*⁵, avec un taux d'erreur de 0,56%, inférieur au taux de 1,16% obtenu en moyenne par des humains... Encore une forme nouvelle de pratique scientifique – « libertarienne » ? – explicitement inspirée des compétitions et « Coupes » aéronautiques des débuts de l'aviation. Renouerait-elle aussi avec les « prix » décernés par les Académies des Sciences du 17^e au 19^e siècle ?

⁵ Il s'agissait d'étiqueter le plus correctement possible 13 000 photos de panneaux routiers (« Stop », « Chaussée glissante », « 60 », ...) prises sous différents angles, éclairages, conditions atmosphériques, résolutions, ... Les compétiteurs pouvaient s'entraîner, pour mettre au point leur méthode, sur 26 000 photos étiquetées par les organisateurs.

A ma connaissance, peu de chercheurs en sciences humaines sont impliqués dans ces initiatives. A noter cependant que quelques sociologues commencent à s'intéresser à l'utilisation de simulations de systèmes complexes multi-agents, à prétention d'explications phénoménologiques, et non de prédictions.

Biographie :

Alain Lelu, ingénieur Arts et Métiers de formation, a été impliqué dans l'essor de l'analyse des données « à la française » dans les années 1970-80, avec un fort intérêt pour les sciences sociales, en particulier au sein de la Direction Générale des Télécommunications pendant une douzaine d'années. Il a entamé dans les années 1990 une carrière universitaire de maître de conférences à l'université Paris 8, puis de professeur à celle de Franche-Comté, poursuivant des recherches sur l'application des statistiques et des modèles neuronaux aux sciences de l'information. Il est actuellement accueilli en délégation à l'ISCC.