La notion de réseau complexe : du réseau comme abstraction et outil à la masse de données des réseaux sociaux en ligne

Alain Barrat

Édition électronique

URL : <http://journals.openedition.org/communicationorganisation/4093>

DOI : 10.4000/communicationorganisation.4093

ISSN : 1775-3546

Éditeur

Presses universitaires de Bordeaux

Édition imprimée

Date de publication : 1 juin 2013

Pagination : 15-24

ISBN : 978-2-86781-883-7

ISSN : 1168-5549

Référence électronique

Alain Barrat, « La notion de réseau complexe : du réseau comme abstraction et outil à la masse de données des réseaux sociaux en ligne », *Communication et organisation* [En ligne], 43 | 2013, mis en ligne le 01 juin 2015, consulté le 19 avril 2019. URL : <http://journals.openedition.org/communicationorganisation/4093> ; DOI : 10.4000/communicationorganisation.4093

La notion de réseau complexe : du réseau comme abstraction et outil à la masse de données des réseaux sociaux en ligne

Alain Barrat¹

Les réseaux, la révolution des données, et la complexité

Le concept de réseau ou de graphe est très facile à définir : il s'agit de points, appelés sites ou sommets, reliés par des liens. Un tel niveau de généralité et d'abstraction a comme conséquence que cette représentation est adaptée à décrire de nombreux systèmes de natures très diverses. Les exemples de réseaux vont de l'Internet jusqu'aux interconnexions d'agents financiers ou bien aux réseaux d'interactions entre gènes, protéines et autres molécules à l'intérieur de la cellule, constituant élémentaire du vivant. On peut aussi citer les grandes infrastructures telles que le réseau électrique ou les réseaux de transports, dont dépendent de manière cruciale nos sociétés modernes. Le mot « réseau » peut donc correspondre à deux types d'entités. D'une part, on s'intéresse à l'objet mathématique, on développe des outils de description et de caractérisation. D'autre part, on parle d'objets bien concrets tels que les réseaux de transport ou de communication, ou bien de phénomènes comme les réseaux d'interactions sociales.

La théorie des graphes existe depuis Euler et sa représentation du problème des ponts de Königsberg en 1735 (comment trouver un chemin qui parte d'un point de la ville de Königsberg et y revienne en passant une fois et une seule par chacun des sept ponts de la ville), et s'est développée comme un pan entier des mathématiques. D'autre part, « l'étude des réseaux sociaux constitue un domaine de recherche bien établi depuis les travaux de Moreno (1934) et sa définition du « sociogramme » comme le diagramme des liens sociaux d'une personne donnée. De nombreux concepts ont été définis dans ce cadre,

¹ Alain Barrat est Directeur de Recherches CNRS au Centre de Physique Théorique de l'Université d'Aix-Marseille, UMR CNRS 7332. Il a obtenu le Prix Paul Langevin en 2009. Ses recherches portent sur l'étude de la structure des réseaux à des applications en épidémiologie dans le domaine interdisciplinaire des réseaux complexes. Il a élaboré LANETvi, un outil de visualisation de réseaux complexes ; alain.barrat@gmail.com

comme par exemple ceux concernant la centralité d'un site (Freeman, 1977 ; Wasserman, 1994).

Depuis quelques années, des scientifiques venant de nombreuses autres disciplines s'intéressent aux réseaux. La possibilité de représenter de nombreux systèmes par des réseaux n'est cependant pas nouvelle. En fait, le changement déterminant qui a eu lieu dans les 15-20 dernières années réside dans la possibilité effective d'obtenir des données empiriques décrivant ces réseaux, et de pouvoir les analyser, grâce aux nombreux développements technologiques et scientifiques. Dans certains domaines, les réseaux étaient inexistantes, comme pour Internet ou la Toile. Dans d'autres (biologie, réseaux d'infrastructures), les données étaient difficiles à obtenir ou à analyser. L'ère de l'avalanche de données, qui s'est ouverte dans les deux dernières décennies, a donc permis la description de nombreux réseaux d'origines diverses, et les chercheurs de divers domaines ont mis en évidence dans ces grands ensembles de données des ordres caches, des régularités et ressemblances inattendues, d'ordre statistique et donc facilement traduits dans le langage des mathématiques.

Le lien entre vision abstraite des réseaux et réseaux bien concrets se nourrit donc avant tout des données empiriques. Diverses propriétés se sont révélées être très génériques. En particulier, nombre des systèmes étudiés sont des « petits-mondes », ce qui traduit le fait que la distance topologique moyenne dans le réseau (qui mesure le nombre moyen de sauts pour aller d'un site à un autre) varie très lentement avec le nombre total de sites (typiquement comme un logarithme), même si chaque site est relié à un faible nombre d'autres sites, et même si la cohésion locale reste forte. Une autre découverte particulièrement importante est le fait que de nombreux réseaux sont caractérisés par une abondance statistique de sites qui ont un très grand nombre de connexions avec d'autres éléments du réseau. En fait, pour de nombreux réseaux, la notion de site représentatif (ou typique) n'a pas de sens car les fluctuations de la connectivité d'un site à un autre sont gigantesques. D'autre part, la présence de structures hiérarchiques, à toutes les échelles, ou de communautés (ensemble de sites mieux connectés entre eux qu'avec le reste du réseau) est également commune.

En plus de leur caractère générique, ces propriétés ont deux aspects très importants. D'une part, elles diffèrent fortement de celles des graphes les plus connus et étudiés, à savoir les graphes aléatoires dans lesquels les sites sont reliés entre eux de manière indépendante les uns des autres (Erdos et Renyi, 1960). D'autre part, ce sont des propriétés émergentes : les réseaux considérés n'ont pas été construits a priori avec ces propriétés, mais elles émergent au niveau du réseau pris dans son ensemble à partir de l'évolution et de l'interaction non planifiées des nombreux éléments du réseau (sites et liens), ce qui justifie l'adjectif « complexe » souvent utilisé. Les régularités

statistiques observées peuvent ainsi être considérées comme la manifestation de lois sous-jacentes gouvernant l'évolution de ces systèmes complexes.

Ces deux aspects ont ainsi stimulé une forte activité interdisciplinaire, nourrie par les données, visant à construire de nouveaux outils de caractérisation des réseaux, à comprendre les mécanismes pouvant permettre l'émergence de leurs propriétés, et à étudier comment les phénomènes dynamiques dont ils sont le support (comme par exemple la propagation d'informations ou d'épidémies) sont influencés par ces propriétés (Barrat *et al.*, 2008).

Le cas des réseaux sociaux

Dans le cas des réseaux sociaux, l'obtention de données a longtemps représenté une difficulté majeure ; les données concernaient souvent des réseaux de petite taille, il s'agissait en général de réseaux personnels (*i.e.*, centrés sur un individu particulier) et non complets (un réseau complet concerne un ensemble d'individus entre lesquels tous les liens existants sont connus), souvent obtenus par des questionnaires et donc correspondant à la vision subjective de chaque individu, et il était rare d'avoir accès à leur évolution temporelle. L'avènement des nouvelles technologies, qui a conduit à la « révolution des données » dans de nombreux domaines, a donc un impact important également dans l'étude des réseaux sociaux. Les données concernant les comportements sociaux peuvent être grossièrement divisées en deux catégories. D'une part, les nouvelles technologies utilisées dans la vie quotidienne amènent les individus à laisser, volontairement ou non, de nombreuses traces numériques de leurs activités et de leur comportement. D'autre part, l'essor du Web participatif et des réseaux sociaux en ligne a créé de nouveaux types de relations sociales, et fournit un nouveau terrain d'investigations pour les sciences humaines. L'utilisation du web participatif permet également de concevoir et réaliser des expériences sociologiques à grande échelle grâce à des outils tels qu'Amazon Mechanical Turk. Des approches quantitatives à de nombreuses questions sociologiques deviennent ainsi possibles. Qu'il s'agisse des traces numériques laissées par les activités humaines ou de nos relations sociales en ligne, l'aspect masse de données nécessite l'intervention de nouveaux outils d'analyse de données et de modélisation ; elle permet de poser la question de l'émergence de propriétés globales à partir de l'action indépendante d'une multitude d'acteurs. Ceci explique l'intérêt et l'interaction de scientifiques venant de domaines divers, de la physique statistique à l'informatique. La combinaison de leurs expertises avec celle des sciences sociales permet actuellement l'entrée dans une ère prometteuse en ce qui concerne les sciences sociales quantitatives.

Des traces toujours plus nombreuses et détaillées du comportement humain

L'utilisation des nouvelles technologies, et en particulier de la téléphonie mobile, laisse des traces non seulement des communications entre individus,

mais également de leur mobilité, comme déduite en particulier des positions des antennes de téléphonie mobile successivement utilisées par le téléphone de chaque utilisateur. Ces informations permettent de déduire des informations statistiques sur les déplacements, à l'échelle d'un pays, de populations de millions d'utilisateurs (Gonzalez, 2008) : des travaux ont ainsi mis en évidence l'existence de régularités statistiques entre les structures de déplacements d'individus à différentes échelles, permettant de définir des lois générales gouvernant les flux d'individus entre régions urbaines, en fonction des densités de populations ; il est également possible de quantifier la prédictibilité statistique des déplacements (Song *et al.*, 2010), c'est-à-dire jusqu'à quel point la trajectoire future d'un individu peut être prévue, connaissant ses déplacements passés. La capacité prédictive est très élevée, correspondant au fait que la plupart des déplacements sont en fait très réguliers d'un jour sur l'autre (domicile-école-travail par exemple).

Les données issues de la téléphonie mobile sont également des traces de la communication entre individus, et peuvent donc être considérées comme des indications de relations sociales (en particulier si l'on considère qu'il n'existe un lien entre deux individus qu'à condition qu'il y ait eu plusieurs échanges d'appels pendant une certaine période, filtrant ainsi les centrales d'appels automatiques). Diverses études ont ainsi examiné les réseaux de communications entre individus, montrant par exemple l'existence de communautés de liens forts et l'importance des liens faibles (Onnela *et al.*, 2007) dans la transmission de l'information. L'application d'algorithmes de détection de communautés a également montré d'une part l'importance (attendue) des frontières linguistiques dans ce réseau de communication (Blondel *et al.*, 2010) mais également qu'en France comme en Belgique, 85 % des communications se font au sein d'une région et seulement 15 % en sortent. Bien que l'existence de structures liées à la géographie ne soit pas une surprise, le fait que les territoires mis en évidence soient les régions et non les départements n'étaient pas attendus, ni que le découpage empirique colle aussi bien à des entités a priori purement administratives.

Alors que les données décrites ci-dessus résultent de l'utilisation quasi universelle de certains outils de communication, le recueil d'autres types d'informations concernant le comportement humain nécessite la conception et le déploiement d'infrastructures spécifiques. C'est le cas des interactions entre individus à l'échelle par exemple d'un bâtiment, d'une école, d'un hôpital... Plusieurs groupes de recherche se sont attaqués à ce problème. A. Pentland, au MIT, a développé les badges sociométriques qui mesurent la proximité entre les individus qui les portent, enregistrent leurs mouvements et leurs paroles, avec le but de mieux comprendre le rôle des interactions dans le management et l'efficacité d'organisations (Pentland *et al.*, 2008). De son côté, la collaboration « SocioPatterns » (www.sociopatterns.org), rassemblant des chercheurs et ingénieurs allemands, italiens et français, a

développé une infrastructure plus légère qui peut être déployée aisément et de façon plus versatile. L'infrastructure, basée sur des badges d'identification par radio-fréquences (RFID) actifs qui échangent des paquets de données à faible puissance, est capable de détecter la proximité en position face-à-face des individus qui portent ces badges, avec une haute résolution temporelle (Cattuto *et al.*, 2010). Elle a été déployée dans des environnements aussi divers qu'une école primaire, des conférences scientifiques, un musée ou des hôpitaux. Les données recueillies permettent petit à petit de créer une sorte « d'atlas » des interactions humaines, de détecter des propriétés génériques et des caractères spécifiques à chaque environnement. Par exemple, la durée d'une interaction entre deux personnes peut varier énormément selon les personnes concernées, mais la statistique des durées est très robuste d'un environnement à l'autre. Dans un hôpital d'autre part, on peut mettre en évidence les différences dans les structures des contacts entre les différentes catégories de personnels. Un important domaine d'utilisation de telles données est en effet l'épidémiologie : les modèles de propagation de maladies infectieuses dépendent crucialement des informations disponibles sur les contacts entre individus. Disposer de plus d'informations permet de rendre les modèles plus réalistes, et de mieux comprendre l'efficacité relative de différentes mesures de santé publiques, comme par exemple la fermeture ciblée de classes dans une école, ou la vaccination prioritaire de certaines catégories de personnel dans des hôpitaux. Dans un contexte comme celui de l'école, on peut aussi s'intéresser à des problèmes d'intérêt social, comme quantifier les phénomènes d'homophilie liés à l'âge ou au genre. La multiplication des sources de données concernant différentes échelles temporelles et spatiales indique clairement les challenges futurs : comment mettre en commun et combiner ces différentes sources, les comparer, afin d'arriver à une description multi-échelles des mobilités et interactions humaines. Une tâche gigantesque et ardue qui pose de nombreuses questions et problématiques nouvelles.

Le Web 2.0 et les nouveaux réseaux sociaux

Une autre « révolution » a eu lieu dans les dernières années dans les relations sociales : l'avènement du web participatif et des réseaux sociaux en ligne. Ces nouveaux outils et les masses de données correspondantes, qui comportent dans certains cas un aspect temporel, posent d'une part de nouvelles questions sur la nature des relations sociales en ligne, et leur éventuelle similarité ou différence avec les autres types de relations sociales, et permettent également de tester des théories sociologiques de manière beaucoup plus détaillée et quantitative que par le passé.

Un aspect fort bien connu et documenté des réseaux sociaux est par exemple le phénomène d'homophilie : deux individus qui ont une relation sociale sont typiquement plus « similaires » que deux personnes prises au hasard dans la population. La similarité peut correspondre à des comportements,

des catégories socio-professionnelles, etc. (McPherson *et al.*, 2001). Diverses études concernant des réseaux sociaux en ligne, comme par exemple aNobii, un réseau spécialisé autour de la lecture (Aiello *et al.*, 2010), ont montré que ce phénomène reste très présent dans les réseaux sociaux en ligne : la similarité entre les profils des utilisateurs décroît avec leur séparation dans le réseau social (les amis tendent à être plus semblables que les amis d'amis, etc.). L'accès à des données décrivant l'évolution du réseau et des profils d'utilisateurs permet de plus d'aller plus loin et de mettre en évidence les deux phénomènes qui ont cette homophilie pour résultat : d'une part, le phénomène de sélection, selon lequel les utilisateurs tendent à créer des liens avec d'autres utilisateurs qui ont un profil similaire ; d'autre part, le phénomène d'influence réciproque entre utilisateurs, qui tendent à devenir encore plus similaires une fois qu'ils sont liés par un lien d'amitié. Il devient également possible de mettre en évidence l'influence des voisins sur le réseau social pour le choix des livres qu'un utilisateur lit : la probabilité qu'un utilisateur lise un livre donné croît en fonction du nombre de voisins qui l'ont déjà lu, en particulier s'il s'agit de voisins avec lesquels une communication régulière a lieu. Le phénomène de bouche-à-oreille semble donc exister également dans ces nouveaux réseaux « sociaux ». Une autre expérience à grande échelle a également permis de mesurer ce phénomène dans le réseau Facebook (Bakshy *et al.*, 2012) : normalement, lorsqu'un utilisateur de Facebook expose un élément sur son « mur » (un article de journal, une vidéo, etc.), ses amis voient cet élément dans leur fil d'actualité (« feed »), et peuvent le re-partager. Le partage successif par deux utilisateurs du même élément peut cependant également être dû au fait qu'ils ont tous deux été exposés, de façon indépendante de leur lien d'amitié sur Facebook, à la même information, qu'ils ont tous deux choisi d'exposer sur leur page Facebook (l'homophilie rend cela plus probable, car les individus liés sur un réseau social ont typiquement les mêmes centres d'intérêt et sont donc en partie soumis aux mêmes stimuli). Afin de quantifier le rôle des liens dans la diffusion de l'information, les auteurs de l'étude ont modifié de façon aléatoire l'exposition d'un grand nombre d'individus aux signaux partagés par leurs amis, qui est automatique en temps normal. Ils ont ainsi montré que, lorsqu'un utilisateur partage une information, les amis de cet utilisateur qui ont été exposés à cette information via Facebook ont une probabilité supérieure de la partager à nouveau, et le font plus rapidement que ceux qui n'ont pas été exposés.

D'autre part, la probabilité qu'un individu a de re-partager un élément est plus élevée si l'élément a été d'abord exposé par un ami avec qui l'individu a un lien fort (quantifié par exemple par le nombre de messages échangés), mais les liens faibles ont également un rôle très important car leur abondance permet à chaque individu de recevoir de nombreux stimuli de natures différentes et d'être ainsi exposé à des informations différentes de celles reçues par son

voisinage « fort », ce qui permet de maintenir une certaine diversité dans les flux et échanges d'informations.

Finalement, un autre exemple intéressant est fourni par l'étude des liens et des communications sur le réseau social Twitter (Gonçalves *et al.*, 2011). L'examen des échanges entre 1,7 millions d'individus pendant 6 mois montre que, même si un utilisateur de Twitter peut « suivre » un nombre arbitraire d'autres utilisateurs, le nombre d'autres utilisateurs avec lesquels un échange a réellement lieu sature aux alentours de 100 à 200, en accord avec la théorie de Dunbar limitant le nombre de relations sociales stables qu'un individu peut maintenir simultanément aux alentours de 150, pour des raisons de capacités cognitives. Cette étude met bien en évidence l'importance de distinguer entre les liens « apparents » des réseaux sociaux en ligne, qui sont a priori en nombre illimité et n'ont aucun « coût » de création, et les liens plus « réels » le long desquels existe une réelle relation de type sociale, dont le maintien a un certain coût en termes de cognition et de temps.

Ces différents exemples, parmi d'autres, montrent un phénomène intéressant : les réseaux sociaux en ligne, contrairement à ce qui a pu être suggéré, ne semblent en fait pas avoir changé de manière drastique la nature des relations sociales. Des études toujours plus nombreuses montrent en effet la validité persistante de propriétés connues, et de diverses théories concernant les interactions entre individus. Le phénomène d'homophilie reste par exemple fort, et les systèmes de recommandation de nouveaux liens en ligne ont en fait même tendance à renforcer cet effet, puisqu'ils suggèrent des liens entre amis d'amis ou entre utilisateurs choisis pour leur similarité. Similairement, les relations « en-ligne » tendent à exister de façon préférentielle dans le voisinage géographique de chaque individu, et n'ont pas aboli les frontières.

L'avènement de nombreuses données et de nouveaux outils du web participatif permet en revanche d'aller plus loin dans les tests et l'analyse de théories sociologiques. Un exemple est donné par Amazon Mechanical Turk (<https://www.mturk.com/mturk/welcome>), qui permet de demander à des humains de réaliser des tâches éventuellement répétitives, faisant appel à l'intelligence humaine et que des ordinateurs ne sont pas capables de faire. Cette possibilité a donné l'idée à des chercheurs de réaliser en ligne à grande échelle des expériences idéalisées portant sur le comportement humain dans des conditions précises, alors que de telles études faites dans un laboratoire sont coûteuses et difficiles à mettre en place. Ainsi, le blog <http://experimentalturk.wordpress.com/> recueille des expériences faites dans le domaine de l'économie, de la psychologie et des sciences sociales en général, faites par le biais d'Amazon Mechanical Turk. De grandes quantités de données sont ainsi obtenues de façon économique et efficace. Comme aucun contrôle sur l'identité et les caractéristiques des individus participant aux expériences ne peut être effectué, ce blog a pour but de recenser et documenter

les expériences ainsi réalisées, afin de les tester, de valider ce nouvel instrument et de comprendre comment l'utiliser au mieux (Paolacci , *et al.*, 2010).

Conclusion

Les exemples cités ci-dessus ne représentent qu'une petite partie du développement actuel de l'étude des sciences sociales « computationnelles » (Lazer *et al.*, 2009). L'ère de l'abondance des données offre de magnifiques perspectives pour affronter de manière quantitative et parfois renouvelée des sujets traditionnels de sciences humaines et sociales. Bien entendu, l'abondance de données ne fait pas tout et on ne peut espérer remédier à la qualité des données par la quantité. En d'autres termes, des données abondantes ne sont pas forcément des données de qualité, et l'accès à des données ne signifie pas que l'on sait forcément quoi en faire. Tout d'abord, des données, même abondantes, peuvent receler des biais potentiels : avoir un échantillon de grande taille ne signifie pas qu'il est représentatif. D'autre part, l'accès aux données n'est pas toujours facile, pour diverses raisons : protection de la vie privée, données jalousement gardées par les entreprises développant les réseaux en ligne, etc. La possibilité pour d'autres groupes de recherche de reproduire les résultats est donc parfois limitée. Des problèmes juridiques de protection et divulgation de données se mêlent aussi à des interrogations plus techniques sur la possibilité de rendre les données réellement anonymes : un simple remplacement du nom par un nombre ne suffit évidemment pas lorsque d'autres données descriptives sont disponibles, et les scientifiques s'interrogent par exemple sur la façon de remplacer les données concernant un réseau social en ligne par un réseau artificiel conservant un certain nombre des propriétés empiriques du réseau réel. Ces différents aspects, associés à la nécessité de collecter et de fouiller de grandes masses de données, de définir de nouveaux outils et cadres d'analyse et de modélisation multi-échelles, de comprendre comment interroger les données et donner du sens à leur analyse, montrent clairement la nécessité d'approches multidisciplinaires associant sciences humaines et sociales aux sciences dures afin d'exploiter au mieux les nouvelles perspectives offertes par l'avalanche de données concernant le comportement humain et les relations humaines.

BIBLIOGRAPHIE

AIELLO L. M., BARRAT A., CATTUTO C., RUFFO G., SCHIFANELLA R., *Link creation and profile alignment in the aNobii social network*, Proceedings of the Second IEEE International Conference on Social Computing SocialCom 2010, Minneapolis, USA, 2010, <http://arxiv.org/abs/1006.4966>

BAKSHY E., ROSENN I., MARLOW C., ADAMIC L., *The Role of social networks in information diffusion*, Proceedings of ACM WWW, 2012, <http://arxiv.org/abs/1201.4145>

BARRAT A., BARTHELEMY M., VESPIGNANI A., *Dynamical processes on complex networks*, Cambridge University Press, 2008

BLONDEL V., KRINGS G., THOMAS I., « Regions and borders of mobile telephony in Belgium and around Brussels », *Brussels Studies*, 42, 2010.

CATTUTO C., VAN DEN BROECK W., BARRAT A., COLIZZA V., PINTON J.-F., VESPIGNANI A., « Dynamics of person-to-person interactions from distributed RFID sensor networks », *PLoS One*, 5(7), 2010.

ERDÖS P., RENYI A., « On random graphs I », *Publ. Math. Inst. Hung. Acad. Sci.*, 5, 1960, p. 17-61.

FREEMAN L., « A set of measures of centrality based on betweenness », *Sociometry*, 40, 1977, p. 35-41.

GONÇALVES B., PERRA N., VESPIGNANI A., « Modeling users' activity on twitter networks : validation of Dunbar's number », *PLoS One*, 6(8), 2011.

GONZALEZ M., HIDALGO C., BARABASI A.-L., « Understanding individual human mobility patterns », *Nature*, 453, 2008, p. 779-782. http://perso.uclouvain.be/gautier.krings/docs/EN_129_BruS42EN.pdf

LAZER D. *et al.*, « Life in the network : the coming age of computational social science », *Science*, 6.323, 2009, p. 721-723.

MC PHERSON M., SMITH-LOVIN L., COOK J., « Birds of a Feather : Homophily in Social Networks », *Annual Review of Sociology*, 27, 2010, p. 415-444.

MORENO J.-L., *Who Shall Survive ? Foundations of Sociometry, Group Psychotherapy, and Sociodrama*, Beacon House, NY, 1934.

ONNELA J.-P. *et al.*, « Structure and tie strengths in mobile communication networks », *Proc. Natl. Acad. Sci. USA*, 104, 2007, p. 7332-7336.

PAOLACCI G., CHANDLER J., IPEIROTIS P., « Running Experiments on Amazon Mechanical Turk », *Judgment and Decision Making*, 5(5), 2010, p. 411-419.

PENTLAND A., PENTLAND S., *Honest Signals : how they shape our world*, MIT Press, Cambridge, 2008.

SONG C., QU Z., BLUMM N., BARABASI A.-L., « Limits of predictability in human mobility », *Science*, 327, 2010, p. 1018-1021.

WASSERMAN S., FAUST K., *Social Network Analysis : Methods and applications*, Cambridge University Press, 1994.

Résumé : Le mot de réseau ou graphe correspond à un objet mathématique abstrait mais aussi à des systèmes bien concrets, aussi divers que l'Internet ou les interactions entre molécules. Les données empiriques permettent de faire le lien entre les différents aspects. En SHS, l'obtention de données a toujours représenté une difficulté mais aujourd'hui, la « révolution des données » commence à trouver un écho. Les nouvelles technologies donnent accès à des informations sur la communication entre individus, sur leur mobilité, et sur certaines interactions de type social. L'étude des réseaux sociaux en ligne fournit

parallèlement un laboratoire où diverses théories sociologiques peuvent être testées quantitativement. Il devient donc possible de réaliser des expériences sociologiques à grande échelle. Des approches quantitatives en SHS deviennent ainsi possibles mais soulèvent de nombreuses questions : la nature des données, leur statut juridique... Autant de questions qui nécessitent un dialogue interdisciplinaire entre sciences dures et sciences humaines.

Mots-clés : abstrait, concret, données empiriques.

Abstract : *The word network or graph corresponds to an abstract mathematical object but also many practical systems, as diverse as the Internet or the interactions between molecules. The empirical data allow to link the different aspects. In SHS, obtaining data has always been a problem but now the «data revolution» beginning to find an echo. New technologies provide access to information on communication between individuals, their mobility, and some kind of social interactions. The study of online social networks in parallel provides a laboratory where various sociological theories can be tested quantitatively. It becomes possible to perform experiments on a large scale sociological. Quantitative approaches in social sciences thus become possible but raise many questions : the nature of the data, their legal status... These questions necessities an interdisciplinary dialogue between sciences and humanities.*

Keywords : *Abstraction, Concrete, Empirical data.*