

# 3 • Internet change-t-il

Les réseaux sociaux tels que Facebook et Twitter produisent quantité de données permettant l'étude de nos comportements. Des scientifiques réfléchissent à la meilleure façon d'en tirer parti.



**PAR Alain Barrat,** directeur de recherche au CNRS, travaille au Centre de physique théorique de Marseille.

**S**ur Facebook, rien de plus facile que d'avoir 1 000, 2 000, voire 5 000 « amis ». Cette facilité à étendre ses connexions entretient l'idée selon laquelle les réseaux permettent d'avoir beaucoup plus de relations que dans la vie réelle. Qu'en est-il vraiment ? Les réseaux sociaux sur Internet ont-ils bouleversé les relations sociales humaines ? Une récente étude portant sur Twitter, un réseau de micromessages en ligne, montre que cette idée reste un mythe : en pratique, on maintient des contacts réguliers avec pas plus de 200 relations [1].

Comment l'équipe menée par Alessandro Vespignani, aujourd'hui à l'université Northeastern, aux États-Unis, s'y est-elle prise pour estimer le nombre maximal de contacts réguliers ? À partir de 380 millions de « tweets », ces chercheurs ont extrait 25 millions de conversations, chacune correspondant à 2 ou 3 tweets envoyés en réponse les uns aux autres sur une période de six mois. Au final, le réseau étudié était formé de 1,7 million d'individus – les nœuds du réseau – reliés par ces 25 millions de conversations – les liens. En attribuant plus de poids à chaque lien selon le nombre de réponses et en étudiant la saturation des connexions en fonction du nombre de messages entrants, ils ont montré que le nombre d'utilisateurs avec lequel un échange a réellement lieu sature aux alentours de 100 à 200. Cette analyse est confirmée par un modèle de champ moyen, très courant en physique statistique, où l'on rem-

place les interactions entre tous les individus par une interaction moyenne. Ici, le modèle considère le point de vue d'un seul utilisateur qui reçoit les connexions « moyennes » de tous les autres.

Cette étude étaye donc la vision de l'anthropologue britannique Robin Dunbar proposée en 1993. Selon lui, pour des raisons cognitives liées à la taille du néocortex, le nombre de relations stables qu'un individu peut maintenir simultanément est de l'ordre de 150. Sur les réseaux sociaux, ajouter un ami, c'est facile, mais entretenir un lien durable prend du temps... Le travail met aussi en évidence la distinction entre 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. Une distinction que l'on retrouve sur les réseaux physiques où la place dont on dispose et le coût contraignent le fait d'ajouter une nouvelle liaison.

**Terrain d'investigation.** Même si les réseaux sociaux font apparaître de nouveaux usages – et tweeter en est un –, ils ne permettent pas de transcender les contraintes biologiques sur la communication humaine. La véritable révolution des réseaux sociaux est, nous l'avons dit, celle des données. Ainsi, l'essor du Web participatif et des réseaux sociaux en ligne fournit un nouveau terrain d'investigation pour les sciences humaines. Un terrain dans lequel la complexité est une part essentielle. Les expériences quantitatives à grande échelle dans les réseaux complexes deviennent possibles, notamment pour tester l'homophilie. L'homophilie est la traduction sociologique de cet adage bien connu : « Qui se ressemble s'assemble. » Dans les réseaux sociaux, deux individus qui ont une relation sont typiquement plus « similaires » que deux individus pris au hasard dans la population. La science des réseaux permet d'analyser plus finement cette similarité.

Eytan Bakshy et ses collègues, de l'université du Michigan, aux États-Unis, spécialistes de l'analyse des données des réseaux sociaux, viennent

## L'essentiel

- > LE FLOT D'INFORMATIONS échangées sur les réseaux sociaux fournit une masse de données pour analyser les comportements.
- > LES ÉTUDES montrent que le nombre de contacts avec lequel un échange a vraiment lieu plafonne à 200.
- > DE NOUVEAUX PROTOCOLES sont développés à différentes échelles pour tirer parti des traces numériques laissées quotidiennement.

# les relations humaines ?

ainsi de publier un travail portant sur la manière dont l'information est échangée parmi 235 millions de personnes inscrites sur Facebook [2]. Habituellement, lorsqu'un utilisateur publie sur son « mur » une information – un article de journal, une photo, une vidéo, etc. –, ses amis voient cet élément dans leurs données de mise à jour (le « feed ») et choisissent ou non de le partager à nouveau. Ce partage successif d'un même élément par deux utilisateurs est-il dû au fait qu'ils ont été tous deux exposés à la même information, indépendamment de leur utilisation de Facebook et de leur lien d'amitié, ou à une réelle propagation de type bouche à oreille ? Autrement dit, est-ce que deux personnes qui partagent la même information le font parce qu'elles ont lu le même journal, ou parce qu'elles ont vu l'information partagée par l'autre ?

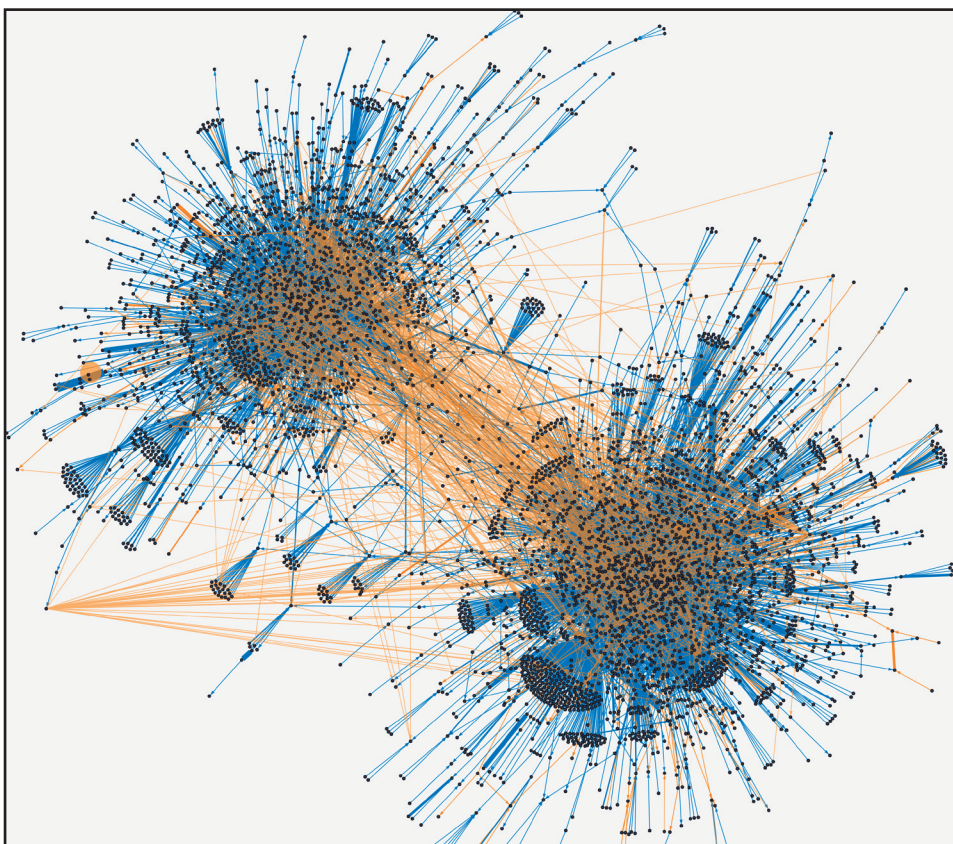
Afin de quantifier le rôle des liens dans cette diffusion de l'information, les auteurs ont modifié l'exposition d'un grand nombre d'individus aux signaux partagés par leurs amis. En temps normal, toute information exposée par un individu apparaît sur le « fil d'actualité » de ses amis. Ici, les chercheurs ont supprimé cette exposition automatique de manière aléatoire pour une petite fraction d'informations.

Ils montrent ainsi que, lorsqu'un utilisateur partage une information, ceux de ses amis qui y ont été exposés *via* Facebook ont une probabilité supérieure de le partager. Et, ils le font plus rapidement, que ceux n'ayant pas reçu l'information sur leur fil d'actualité. De plus ils montrent l'importance des liens dits « forts », quantifiés par exemple par le nombre de messages échangés.

Ainsi, la probabilité de repartager une information est plus élevée si elle a été d'abord exposée par un ami avec qui l'individu a ce lien fort. Toutefois, le

rôle des liens faibles est également intéressant. Car plus ces liens sont nombreux, plus les stimuli sont variés. Du coup, un individu dans un tel réseau est ainsi plus exposé à des informations différentes de celles reçues par son voisinage « fort ».

**Données protégées.** Aussi stimulantes soient-elles, ces études de grande ampleur se heurtent à un obstacle de taille : le problème de la confidentialité des données. De fait, 3 des 4 signataires de cet article appartiennent au Centre de recherche sur les données de la société Facebook. Comme beaucoup d'acteurs commerciaux du Web, l'entreprise garde jalousement ses propres données. Aucun autre groupe n'a donc accès à ces données pour les analyser indépendamment. Ce qui pose un sérieux problème de reproductibilité >>>



**Ce réseau est généré par Twitter, outil de micro-messages en ligne. Il correspond à la diffusion des messages contenant un identifiant lié au parti républicain aux États-Unis. Chaque nœud (point noir) est un utilisateur. Selon la manière dont est diffusé le message, les chercheurs distinguent deux populations, l'une plutôt orientée à gauche (en orange), et l'autre plutôt à droite (en bleu).**



Des chercheurs ont équipé les visiteurs d'un musée de puces RFID. Deux personnes équipées qui se croisent produisent un « contact ». Sur la figure, chaque point représente un visiteur, et l'arc de cercle indique le nombre de contacts entre individus à l'intérieur du musée toutes les deux minutes. © SOCIOPATTERN.ORG

» des résultats, pierre angulaire de la démarche scientifique... Pour contourner cette difficulté, les chercheurs travaillent aussi avec des données publiques, souvent sur des réseaux plus modestes en taille que Facebook.

Une autre stratégie consiste à collecter soi-même des données issues de traces numériques laissées, volontairement ou non, par les nouvelles technologies utilisées dans la vie quotidienne. Avec des collègues allemands, italiens et français, de la collaboration « SocioPatterns » nous avons ainsi mis au point une infrastructure à base de badges RFID actifs qui échangent des paquets d'informations lorsque deux individus équipés se retrouvent face à face [3]. Nous avons déployé cette infrastructure dans des

environnements divers : une école primaire, des conférences scientifiques, un musée ou des hôpitaux. Le déploiement est rapide : dans une école par exemple, on peut installer badges et matériels le lundi, et récupérer les données prêtes à être analysées le vendredi suivant. Grâce à ces traces, on dresse petit à petit une sorte d'atlas des interactions humaines. Peut-on y discerner des propriétés génériques et des caractères spécifiques aux différents environnements ? Comme on pouvait s'y attendre, les structures des réseaux de contacts diffèrent entre une journée au musée et une semaine à l'école par exemple [fig. 1]. Cependant, quand on étudie en détail les durées d'interaction entre les personnes, elles présentent des similarités : la statistique des durées d'interactions est analogue dans les différents environnements [4].

**Régularités statistiques.** À plus grande échelle, les traces de la téléphonie mobile permettent de déduire des informations statistiques sur les déplacements au sein d'une ville ou d'un pays. En suivant des millions d'utilisateurs, on découvre des régularités statistiques dans ces déplacements et ce, à différentes échelles. Ces régularités servent à définir les lois générales qui gouvernent les flux d'individus entre régions urbaines, selon les densités de population. Le degré de désordre des trajectoires individuelles de 50 000 personnes, obtenues via la localisation approximative

de leurs téléphones portables, durant trois mois, a même permis de quantifier jusqu'à quel point la trajectoire future d'un individu peut être prédite, connaissant ses déplacements passés [5]. Cette capacité prédictive est étonnamment élevée, avec une moyenne de 93 %, malgré de fortes différences dans les trajectoires individuelles.

Qu'il s'agisse des traces numériques laissées par les activités humaines ou de nos relations sociales en ligne, l'aspect masse de données requiert le développement de nouveaux outils d'analyse de données et de modélisation ; elle permet de poser autrement la question de la complexité : comment émergent des propriétés globales à partir de l'action individuelle d'une multitude d'acteurs. Ces masses de données posent aussi de nouveaux problèmes. Dans le cas du suivi par RFID, on peut collecter des dizaines de milliers de contacts de durées très variées entre des centaines d'individus en les enregistrant toutes les minutes. Jusqu'à quel point peut-on simplifier ces données de manière à ne garder que l'information essentielle pour que la modélisation soit possible, sans perdre les aspects dynamiques importants ? La question reste ouverte.

Toutes ces interrogations expliquent 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 ouvre la voie à l'ère des sciences sociales quantitatives. Le Web participatif est un élément important de ce domaine : jusqu'à quel point les internautes sont-ils prêts à participer à ces études de sciences sociales ou de santé publique ? Le lancement récent du réseau de surveillance de la grippe – grippenet.fr – vise à répondre à ce type de question. Et quand les sociétés privées entrent dans la partie, cela peut donner un réseau de travailleurs en ligne, tel celui imaginé par la société Amazon. Avec Amazon Mechanical Turk, vous pouvez inclure dans les logiciels que vous êtes en train de concevoir des tâches répétitives pour lesquelles l'intelligence humaine reste irremplaçable. Par exemple, si votre logiciel doit effectuer une reconnaissance faciale, tâche mieux réalisée par un être humain qu'un ordinateur, vous pouvez « recruter » des milliers de personnes à travers le monde qui feront cette reconnaissance moyennant une petite rémunération. Encore un nouveau domaine d'exploration pour la sociologie des réseaux complexes... et celle du travail. ■

[1] B. Gonçalves et al., *PLoS One*, 6 (8):e22656, 2011.

[2] E. Bakshy et al., in *The Proceedings of ACM WWW 2012*, Lyon, France, arXiv.1201.4145v2.

[3] [www.sociopatterns.org](http://www.sociopatterns.org)

[4] C. Cattuto et al., *PLoS One*, 5(7):e11596, 2010.

[5] C. Song et al., *Science*, 327, 1018, 2010.