LES GRANDS RÉSEAUX D'INTERACTIONS ET LES PETITS MONDES

Emmanuelle Lebhar

Résumé. — On appelle *réseau d'interactions* tout ensemble d'entités interagissant de façon individuelle. On qualifie le réseau de *grand* lorsque le nombre d'entités mises en jeu est de l'ordre du million. Les grands réseaux d'interactions recouvrent ainsi des réseaux aussi divers que le réseau des connexions Internet, le réseau des pages web, le réseau des contacts sociaux entre individus, ou encore le réseau des réactions chimiques entre protéines dans le métabolisme d'un être vivant. Alors que les interactions locales sont généralement bien connues (la communication entre deux ordinateurs, la réaction entre deux protéines), le résultat *global* de l'ensemble des interactions est encore mal compris. La compréhension de ces propriétés globales touche pourtant à des problématiques essentielles : la dynamique des interactions dans un réseau social ou un réseau informatique est par exemple liée à la problématique de la propagation des virus (informatiques ou biologiques), celle d'un réseau de distribution d'électricité, au problème de la robustesse d'un grand réseau. L'augmentation récente des capacités de traitement et de collecte d'un grand nombre de données statistiques sur ces réseaux a permis l'essor des études de ces objets. En particulier, on a observé expérimentalement que ces réseaux, *a priori* éloignés, partageaient des propriétés macroscopiques communes, dont l'effet *petit monde*.

1. Les grands réseaux d'interactions

On appelle *réseau d'interactions* tout ensemble d'entités, ou *nœuds*, interagissant de façon individuelle. On qualifie le réseau de *grand* lorsque le nombre d'entités mises en jeu est de l'ordre du million. Les grands réseaux d'interactions recouvrent ainsi des réseaux aussi divers que le réseau des connexions Internet, le réseau des pages web, le réseau des contacts sociaux entre individus, ou encore le réseau des réactions chimiques entre protéines dans le métabolisme d'un être vivant.

Alors qu'il y a quelques années, on ne disposait pas d'un ensemble de données suffisant sur ces réseaux pour les étudier statistiquement, l'augmentation récente des capacités de traitement et de collecte d'un grand nombre de données statistiques sur ces réseaux a permis l'essor des études de ces objets. L'observation majeure qui a motivé le développement de ces études a été de constater que ces réseaux, sémantiquement éloignés, partageaient des propriétés communes au niveau de leur comportement global, comme l'existence d'un chemin très court entre toutes les paires d'entités.

L'étude conjointe des propriétés macroscopiques de différents types de réseaux contient plusieurs problématiques. Pour donner un sens à une propriété observée sur plusieurs ensembles de données statistiques, il faut s'assurer de sa pertinence. En particulier, il faut garantir que la méthode de collecte des données n'est pas biaisée et que la propriété est spécifique, c'est-à-dire qu'on ne l'observe pas sur tout ensemble de données arbitraire, ou aléatoire uniforme. Une fois qu'une propriété pertinente est isolée se pose la question de son apparition dans le réseau : est-ce dû à un processus dynamique de construction du réseau ? Comment peut-on la reproduire ? Révèle-t-elle un principe général sous-jacent ? Les réponses à ces questions peuvent avoir des applications intéressantes si la propriété présente un avantage pratique (on peut par exemple penser aux mécanismes de diffusion de rumeurs dans un réseau social pour une entreprise de marketing), mais aussi s'il s'agit d'une propriété cruciale (comme la diffusion rapide des virus) que l'on souhaiterait prédire pour la contrôler. Pour résumer, les études sur les grands réseaux d'interactions ont trois objectifs principaux : isoler les propriétés communes pertinentes, les reproduire, et les expliquer.

De ce point de vue, la modélisation mathématique de ces réseaux est un outil puissant. Elle peut permettre de tester une hypothèse d'explication de la présence d'une propriété en créant un modèle basé sur cette hypothèse et en vérifiant que la propriété macroscopique est présente. Elle peut également permette de prévoir le comportement d'un réseau, une fois celui-ci correctement reproduit par le modèle.

2. Les propriétés observées

Nous listons dans cette section les trois principales propriétés statistiques qui ont été observées expérimentalement : la distance très courte entre toutes les paires de nœuds (reliée à l'effet petit monde), la distribution des degrés particulière et la forte densité locale des voisins.

1. Distance très courte entre toutes les paires de nœuds. Dans de nombreux réseaux d'interactions, la distance moyenne observée entre deux objets est, de façon surprenante, de l'ordre du logarithme du nombre total de nœuds, donc très faible. Pourtant, le nombre de connexions, ou *liens*, reste très inférieur au carré du nombre de nœuds, ce qui signifie que la densité des connexions est faible. Newman [**New01**] donne l'exemple d'un réseau de co-auteurs d'articles de biologie de 1 520 251 nœuds et 11 803 064 liens, dans lequel la longueur moyenne d'un chemin est 4,9. Un autre exemple est l'étude du réseau Internet de 10 597 nœuds et 31 992 liens effectuée par Faloutsos et al. [**FFF99**], où la longueur moyenne d'un chemin est de 3,3.

2. Distribution des degrés. Le degré d'un nœud est le nombre de connexions qui le rattachent au réseau. On dit qu'un réseau présente une *distribution des degrés* suivant une loi de puissance si le nombre de nœuds de degré k est proportionnel à $1/k^{\alpha}$, pour une constante $\alpha > 0$. En 1999, Faloutsos et al. [**FFF99**] ont observé que le réseau Internet présentait cette propriété. Par la suite, elle a également été observée dans des réseaux de pages web, et des réseaux de distribution d'électricité [**New03**]. Cette découverte a été cruciale pour les travaux sur la propagation des virus dans les réseaux réels.

3. Forte densité locale des voisins. On parle de forte densité locale d'un réseau lorsque les voisins d'un même nœud sont très reliés entre eux. Dans un réseau social, par exemple, cela signifie que les amis d'un même individu ont une grande probabilité d'être amis entre eux. Pour quantifier cette propriété, Watts et Strogatz ont introduit en 1998 la notion formelle de *coefficient de clustering* [**WS98**] qui donne une mesure formelle de cette densité. Ils observent que ce coefficient s'élève à 0,2 dans un réseau de collaborations entre acteurs de cinéma comprenant 449 913 nœuds et 25 516 582 liens, alors qu'en construisant un réseau purement aléatoire uniforme ayant le même nombre de nœuds et de liens, on obtient un coefficient de l'ordre de 10^{-4} seulement.

3. L'effet petit monde

Nous en venons à présent à la définition de l'effet petit monde. Si ce terme est parfois employé pour décrire la présence de l'ensemble des propriété citées ci-dessus dans un réseau, nous nous focaliserons sur une caractéristique bien spécifique de l'effet petit monde : le fait que l'on soit capable de découvrir, sans connaître la carte du réseau, les chemins très courts qui relient toutes les paires de nœuds.

L'effet petit monde tient son nom de l'expression populaire *le monde est petit* désignant la surprise de constater que deux connaissances d'un même individu, *a priori* sans rapport, se connaissent entre elles. Stanley Milgram a effectué une expérience sociologique en 1967 [**Mil67**] en demandant à 300 habitants du Nebraska et de Boston de faire parvenir une lettre à un habitant de Boston dont ils ne connaissaient que le lieu d'habitation et la profession, en ne la retransmettant qu'à une personne qu'ils connaissaient personnellement (et en itérant le processus jusqu'à atteindre la personne cible). Relativement peu de lettres sont arrivées à destination (environ un quart), mais le résultat surprenant fut que la longueur moyenne d'une chaîne de porteurs du message de son origine à sa destination était très faible (5,2) en regard du nombre d'individus et de leur éloignement géographique et social. On pourrait penser qu'il n'est pas vraiment surprenant que les chaînes soient si courtes, puisqu'en supposant que chaque individu ait seulement 10 connaissances, chacun pourrait *a priori* atteindre 10^6 individus en 6 sauts. Mais les réseaux sociaux présentent une forte densité locale, il est donc probable que les connaissances immédiates d'un individu n'aient qu'une seule connaissance étrangère à ce voisinage, et que l'on atteigne seulement 11 individus en 2 sauts par exemple, c'est pourquoi cette propriété est remarquable.

La notion de petit monde n'a pas aujourd'hui de définition formelle ; elle est définie, dans certains articles ([**WS98**]), comme la combinaison d'un fort coefficient de *clustering* et d'un petit diamètre. C'est seulement récemment que l'on a découvert un aspect supplémentaire fondamental dans l'expérience de Milgram, qui a été mis en valeur par Jon Kleinberg en 2000 [**Kle00**], il s'agit de la notion de *navigabilité*. Dans l'expérience de Milgram, les individus n'utilisent que leurs contacts locaux pour renvoyer la lettre ; il s'agit donc d'une découverte de chemin qui n'utilise qu'une vue locale du réseau. C'était aussi le cas de la navigation à travers le réseau des pages web il y a quelques années, qui se faisait d'une page à l'autre sans connaître la carte globale du réseau [**AJB99**]. Par ailleurs, la découverte des chemins de façon décentralisée est une nécessité pour les réseaux d'interactions réels qui comportent un très grand nombre de sommets et où une recherche classique des plus courts chemins n'est pas possible, car très coûteuse en temps.

En mettant en valeur cet aspect, Jon Kleinberg a proposé le premier modèle mathématique de réseau qui reproduisait ce phénomène de découverte décentralisée des chemins courts. L'étude de ce modèle et les recherches sur sa généralisation ouvrent un ensemble de perspectives d'applications pour les grands réseaux de télécommunication actuels (comme les réseaux d'échange de fichiers peer-to-peer).

Références

[AJB99] R. Albert, H. Jeong, and A.-L. Barabási. The diameter of the world wide web. *Nature*, 401(130–131), 1999.

[FFF99] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *Computer Communications Rev.*, 29 :251–262, 1999.

[Kle00] J. Kleinberg. The Small-World Phenomenon : An Algorithmic Perspective. In *Proceedings of the 32nd ACM Symposium on Theory of Computing (STOC)*, pages 163–170, 2000.

[Mil67] S. Milgram. The small world problem. *Psychology Today*, 61(1), 1967.

[New01] M.E.J. Newman. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*, 98 :404–409, 2001.

[New03] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2) :167–256, 2003.

[WS98] D. Watts and S. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(440–442), 1998.

Emmanuelle Lebhar

CNRS, LIAFA - Université Paris 7, 2 place Jussieu, Case 7014, 75251 PARIS Cedex 05.

E-mail : `elebhar@gmail.com`