



RÉSEAUX COMPLEXES ET PHYSIQUE STATISTIQUE

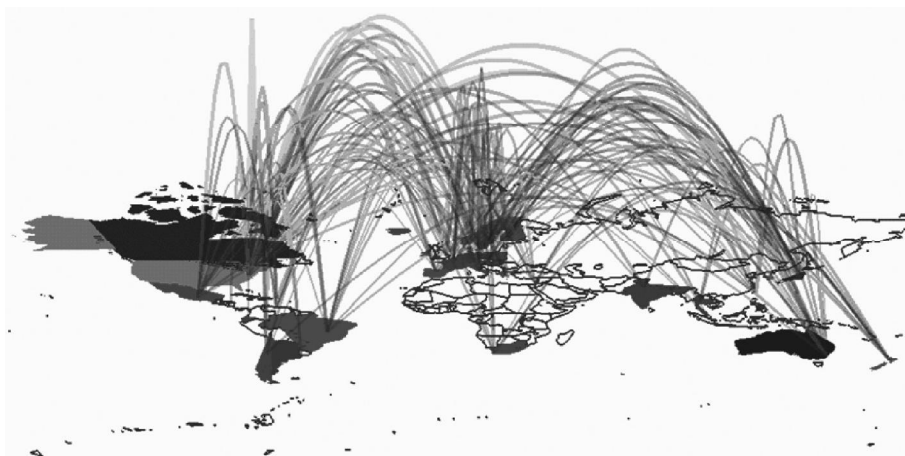
Alain Barrat (1990 s)

Sa sortie de l'ENS s'est effectuée en plusieurs étapes : il a commencé en 4^e année une thèse de doctorat au Laboratoire de physique théorique (sous la direction de M. Mézard) et fréquenté les locaux les deux années suivantes. Depuis, il a plutôt évité l'École, en passant deux ans en Italie puis au CNRS, à Orsay. Il s'en éloigne d'ailleurs de plus en plus, étant actuellement directeur de recherche au Centre de physique théorique à Marseille et chercheur associé à la Fondation ISI de Turin.



Depuis un certain temps, les physiciens statisticiens s'intéressent à des systèmes en dehors de leurs champs d'étude traditionnels. La physique statistique possède en effet une longue tradition dans l'étude des comportements collectifs, et elle a pour cela développé de nombreux outils mathématiques et des concepts qui peuvent être utilisés dans des contextes divers. L'exemple le plus récent de cette démarche concerne le domaine des réseaux complexes.

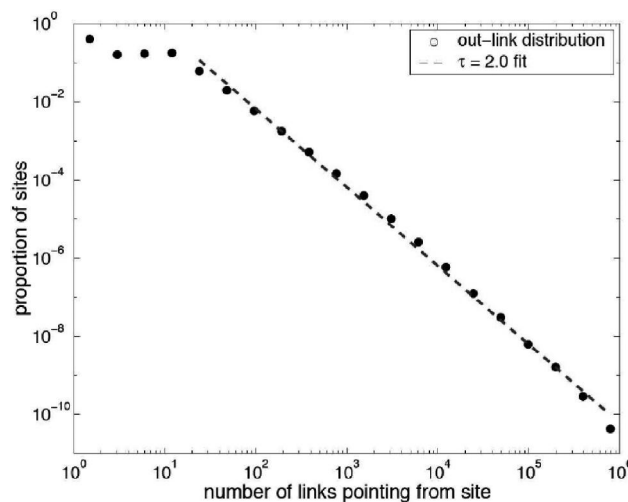
De nombreux systèmes, naturels ou artificiels, peuvent être représentés par des réseaux, c'est-à-dire des sites ou sommets reliés par des liens. L'étude de ces réseaux est par nature interdisciplinaire car ceux-ci apparaissent dans des domaines scientifiques aussi variés que la physique, la biologie, l'informatique ou bien les technologies de l'information. Les exemples vont d'Internet et d'autres infrastructures critiques (réseaux de transport, réseau électrique) jusqu'aux interconnexions d'agents financiers ou bien aux réseaux d'interactions entre gènes, protéines et autres molécules à l'intérieur de la cellule.



Représentation du trafic internet mondial sous forme d'arcs
(d'après Stephen G. Eick <http://www.cybergeography.org/atlas/geographic.html>)

Contrairement à des systèmes qui peuvent être très « compliqués », mais en suivant un plan prédéfini, les réseaux dits « complexes » sont en général le résultat d'une évolution décentralisée et non planifiée. Le fait qu'une telle auto-organisation aboutisse, à partir de mécanismes « microscopiques » (c'est-à-dire au niveau des sites et des liens), à l'émergence de propriétés statistiques macroscopiques permet de comprendre l'implication naturelle de la physique statistique dans ce domaine de recherche.

L'analyse des réseaux complexes a été rendue possible grâce à l'apparition récente de grandes bases de données : les premiers projets de cartographie ont concerné la Toile et Internet. Graduellement, d'autres cartes sont apparues et ont rendu possible la description de nombreux réseaux intervenant dans les sciences sociales, les infrastructures et la biologie. Ces recherches ont alors permis de mettre en évidence, en dépit de leurs origines très diverses, des motifs et des régularités statistiques communes à nombre de ces réseaux. En particulier, ces systèmes sont des « petits-mondes », ce qui traduit le fait que la distance moyenne dans le réseau (qui mesure le nombre moyen de liens à franchir sur le réseau pour aller d'un site à un autre) est très petite par rapport au nombre total de sites, bien que chaque site soit connecté avec seulement quelques autres. 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 degré, où le degré d'un site est défini comme son nombre de connexions avec d'autres éléments du réseau. Cette caractéristique ressort clairement lors de l'observation de la fréquence d'apparition de sites avec k voisins qui est décrite par une loi de puissance, indiquant ainsi l'absence de toute échelle caractéristique (d'où le qualificatif souvent employé de réseaux « sans-échelle »).



Nombre de liens pointant d'un site web (d'après L.A. Adamic et A. Huberman).



Ce nombre suit une loi quasi linéaire dans un diagramme log-log (logarithmes en abscisses et en ordonnées), sur une grande échelle (de 10 à plus de 1 million de liens sortants). Ce type de répartition est connue dans de nombreux domaines, par exemple sous le nom de loi de Zipf en linguistique ou de Parreto en économie.

En d'autres termes, pour ces réseaux, la notion de site représentatif ou typique n'a pas de sens car les variations de la connectivité d'un site à un autre sont gigantesques.

Nouvelles approches de modélisation

Ces résultats empiriques ont montré la possibilité de l'existence de principes sous-jacents allant au-delà des particularités de chaque domaine, et donc l'intérêt de développer des cadres généraux de modélisation. Pendant longtemps en effet, on a considéré que, lorsqu'un système pouvait être décrit en termes de graphes ou réseaux, on pouvait utiliser le paradigme d'un ensemble de points reliés aléatoirement, c'est-à-dire du graphe aléatoire proposé par les mathématiciens Erdős et Renyi dans les années 1960. Ces graphes sont cependant homogènes, dans le sens où les sites sont à peu près équivalents les uns aux autres : le nombre de voisins de chaque site fluctue très peu autour d'une valeur moyenne. L'observation empirique de nombreux réseaux sans-échelle a montré l'inadéquation de ce modèle à nombre de systèmes réels. D'autre part, le caractère auto-organisé et évolutif des réseaux complexes a imposé un changement total de perspective : on est passé d'une modélisation *ad hoc* où on impose les caractéristiques statistiques à des approches de modélisation de processus microscopiques qui permettent l'émergence spontanée de ces caractéristiques. Dans les dernières années, la physique statistique a ainsi joué un rôle important dans la définition de nouvelles classes de modèles, et la description de mécanismes permettant de comprendre l'émergence des propriétés des réseaux complexes observées empiriquement.

Phénomènes dynamiques

La plupart des réseaux complexes sont le siège de phénomènes dynamiques. Par exemple, Internet et la Toile supportent un flux continu d'informations ; les réseaux sociaux sont le siège de phénomènes de formation d'opinion ou de propagation d'informations, ainsi que d'épidémies. La mise en évidence du caractère hétérogène de la topologie des réseaux a donc naturellement mené à la question de l'influence de telles caractéristiques sur ces phénomènes dynamiques.

Il a été compris tout d'abord empiriquement que la structure très hétérogène des réseaux complexes implique une grande résistance à des pannes, c'est-à-dire à un phénomène où des sites deviennent inopérants et donc sont retirés du réseau de manière aléatoire, mais en même temps une grande fragilité devant des attaques





ciblées sur les sites les mieux connectés. Pour un réseau hétérogène, diverses études se sont intéressées à l'évolution de la taille de la plus grande composante connexe en fonction de la fraction de sommets du réseau qui sont supprimés (une composante d'un graphe est dite connexe si et seulement s'il existe un chemin entre chaque paire de sommets de la composante). Si les sommets sont retirés au hasard, la dégradation de la taille du réseau survivant, et donc par exemple de ses capacités de transmission, est lente. Ceci est dû au fait que de telles pannes aléatoires touchent typiquement des sites faiblement connectés (car ce sont les plus nombreux), plutôt périphériques ; l'intégrité du réseau est dans ce cas peu remise en question. Au contraire, l'élimination d'une petite fraction des sites les plus connectés conduit à une très rapide désintégration du réseau en petites composantes isolées. En effet, ces sites de fort degré mettent en communication de nombreuses parties du réseau et leur suppression a des conséquences beaucoup plus importantes. Il est possible de comprendre quantitativement cette phénoménologie par une approche de physique statistique, en la traduisant en termes de problème de « percolation » (c'est-à-dire de comportement d'un graphe lorsque l'on retire une fraction de sites ou de liens au hasard). Les fluctuations de connectivité apparaissent alors clairement comme responsables de cette ambivalence entre robustesse et vulnérabilité des réseaux complexes.

Mentionnons les phénomènes de cascades, qui sont par exemple à la base des gigantesques pannes électriques ayant eu lieu dans divers pays ces dernières années. Dans le cadre d'un réseau servant de support à un transport d'énergie ou d'information, de tels phénomènes peuvent se décrire de la façon suivante : supposons que chaque élément du réseau a une capacité finie, au-delà de laquelle il devient inopérant. Si un des éléments entre en surcapacité, les informations (par exemple) qu'il devrait transmettre doivent être déviées vers d'autres sites ou liens du réseau ; ces autres éléments peuvent, du fait de cette surcharge, entrer à leur tour en surcapacité et devenir inopérants. Un tel processus d'avalanche peut soit s'arrêter rapidement, soit se propager largement, causant une saturation globale. De nombreuses études récentes ont ainsi modélisé le trafic sur des réseaux complexes, pour comprendre dans quelles conditions un incident initial, éventuellement mineur, peut causer un dysfonctionnement à grande échelle. Des travaux récents tendent aussi à développer des stratégies de réaction rapides afin d'empêcher une cascade de se développer.

Un autre exemple particulièrement intéressant de phénomène dynamique sur réseau concerne la propagation d'épidémies. En effet, il a été montré empiriquement que les réseaux de contacts entre humains (par exemple les réseaux d'interaction sexuelle, selon lesquels se propagent les maladies sexuellement transmissibles) sont hautement hétérogènes. D'autre part, les virus ou les vers informatiques se propagent respectivement sur le réseau Internet ou sur le réseau d'échange de courriers électroniques. La physique statistique et l'épidémiologie se rejoignent pour comprendre comment



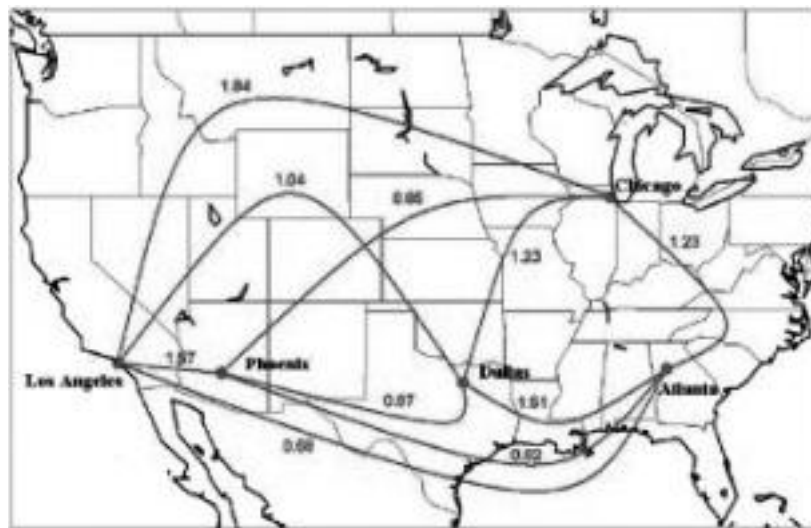
de tels phénomènes épidémiques se propagent. En effet, les modèles communément utilisés en épidémiologie peuvent être exprimés sous la forme de modèles de type réaction-diffusion, couramment étudiés en physique statistique : les individus peuvent être dans différents états, par exemple sain (S) ou infecté (I), et un individu infecté peut contaminer ses voisins sur le réseau. La topologie du réseau de contacts le long desquels se transmet l'infection a donc un rôle important, et les outils de la physique statistique permettent de l'appréhender. Ces systèmes présentent généralement une transition entre une phase stationnaire active où l'épidémie se propage et une phase inactive où elle meurt. Le paramètre de contrôle permettant de changer de phase est la transmissibilité de l'infection ; on s'intéresse généralement à la détermination de sa valeur seuil, pour comprendre si une infection est dangereuse et risque de contaminer une grande partie de la population ou, au contraire, va disparaître sans se propager. Des études récentes ont montré, dans le cas des modèles paradigmatiques utilisés en épidémiologie théorique, la conséquence drastique de l'hétérogénéité du réseau sur le seuil de propagation, qui tend vers 0 si les fluctuations de degré sont très grandes. Aussi faible que soit la transmissibilité, l'épidémie peut alors s'étendre et survivre dans le réseau, grâce au fait qu'il existe toujours une probabilité non nulle que l'infection rejoigne un site très fortement connecté qui a ainsi un fort pouvoir infectieux. Ces développements ont en fait permis de mieux comprendre les données existantes sur les virus informatiques, qui montrent que de nombreux virus sont encore présents dans le réseau Internet pendant des durées très longues, c'est-à-dire de nombreux mois après la mise à disposition des antivirus. Ces constatations impliquent en effet que l'immunisation de sommets pris au hasard par l'antivirus est inadéquate pour bloquer la propagation, et qu'une stratégie efficace consiste à protéger en premier lieu les sites fortement connectés, qui sont des « superpropagateurs ».

Pour modéliser des propagations d'épidémies à grande échelle, on utilise d'autre part souvent des modèles dits de « méta-populations », pour lesquels on considère que chaque site du réseau ne représente pas seulement un individu mais possède une sous-structure. Par exemple, on peut penser à des villes connectées entre elles par le réseau de transport. On doit alors modéliser l'épidémie à deux niveaux : une dynamique d'infection à l'intérieur de chaque ville, et une dynamique de voyage entre les villes, donnée par la structure complexe et hétérogène du réseau de transport. Où intervient la physique statistique, peut-on se demander ? D'une part, les modèles utilisés s'écrivent avec un type d'équations bien connues des physiciens dans d'autres contextes, et pour lesquelles les techniques d'étude pour les processus stochastiques s'appliquent. D'autre part, le bagage culturel des physiciens a permis de proposer des moyens pour caractériser et quantifier certains aspects importants de la propagation d'une épidémie dans un tel réseau. Par exemple, l'hétérogénéité de la propagation, c'est-à-dire le fait que des régions différentes soient atteintes à des moments et des degrés





différents, peut être quantifiée par une « entropie » ; la prévisibilité de la propagation, et donc la capacité des modèles à la prévoir avec plus ou moins de précision, peut de plus être mesurée par le recouvrement entre deux réalisations stochastiques. Ces approches permettent également de comprendre, parmi tous les éléments caractérisant le réseau de transport, lesquels sont les plus influents dans cette hétérogénéité et cette prévisibilité partielle. Ainsi, l'hétérogénéité du réseau de transport aérien, où des sites à grand nombre de connexions cohabitent avec de petits aéroports, diminue *a priori* la capacité de prédiction ; cependant, l'hétérogénéité des flux de passagers entre les différentes liaisons aériennes possibles crée des chemins de propagation privilégiés et renforce au contraire les capacités prédictives des modèles.



Quelques liens du réseau des lignes aériennes reliant les aéroports nord-américains, le poids des connexions mesurant le nombre de sièges disponibles (millions/an).
© LPT, CNRS-CEA.

Dans les dernières années, le développement de ces approches, l'accès à des bases de données toujours plus complètes ainsi qu'à des ordinateurs de plus en plus puissants ont permis, par des efforts communs aux physiciens et aux épidémiologistes, de construire des modèles de plus en plus réalistes permettant d'évaluer des scénarios pandémiques ainsi que l'efficacité et l'intérêt de diverses stratégies de lutte (antiviraux, vaccinations...). De tels modèles sont par exemple utilisés pour fournir des prédictions sur les possibles scénarios de propagation de la pandémie actuelle de grippe A/H1N1 (voir <http://www.gleamviz.org>).



Perspectives

Le domaine des réseaux complexes s'est énormément développé dans les dernières années. Des centaines d'articles scientifiques ont été écrits ainsi que quantités de livres. De nombreuses collaborations interdisciplinaires ont été initiées. Alors que certains aspects de l'étude des réseaux semblent avoir été bien explorés et compris, de nombreuses questions restent totalement ouvertes, et de nouvelles directions de recherche ont vu le jour récemment. Un premier sujet intéressant concerne le fait qu'un réseau complexe existe rarement de manière isolée : plusieurs réseaux sont souvent couplés entre eux, comme par exemple un réseau de distribution électrique et le réseau informatique de contrôle correspondant, ou bien différents réseaux sociaux. Il faut donc généraliser les approches précédentes au cas de réseaux couplés, ce qui est loin d'être simple. De même, de nombreux phénomènes se déroulent simultanément à plusieurs échelles, comme par exemple les épidémies, qui font intervenir des réseaux de transport aussi bien à l'échelle locale qu'à l'échelle continentale ou intercontinentale. Finalement, beaucoup de réseaux sont intrinsèquement dynamiques, et le fait de tenir compte de cette dynamique ouvre de nombreuses perspectives, que ce soit en termes de définition de nouveaux outils de description ou d'une meilleure connaissance des phénomènes dynamiques prenant place sur réseaux (voir par exemple <http://www.sociopatterns.org>).

INTERNET : POURQUOI, COMMENT ?

*Christine Nora (1963 S)
Avec le concours de son mari, Hervé Nora,
et de nombreux amis, tous acteurs de ce domaine.*

Depuis son arrivée à l'École, en 1963, Christine Nora n'a jamais officiellement quitté l'Éducation nationale. En pratique, elle a effectivement, pour commencer, passé trois mois à l'UPMC (à l'époque Jussieu) et un an pour finir comme chargée de mission à la présidence de l'UPMC. Entre temps, elle a toujours été « du côté d'ailleurs » (CNRS, MIT, ENST, X, UPMC).



L'origine d'Internet passionne de nombreux historiens. Il suffit, pour s'en convaincre, de lire notre rapide bibliographie donnée en annexe et surtout de consulter Google.

Mieux comprendre l'arrivée d'Internet suppose de se placer dans le contexte de l'informatique et des réseaux depuis leur origine, ce que, en tant qu'acteur ou observateur de ces évolutions, je vais essayer de faire. Mais comme c'est souvent le cas pour les événements récents, la mémoire des acteurs est souvent sélective et ils ne se rappellent que de leur propre point de vue, dans des versions parfois inexactes.