

Réseaux complexes et physique statistique



Les physiciens statisticiens portent une attention sans cesse croissante à des systèmes extérieurs à leur champ d'étude traditionnel. La physique statistique possède une longue tradition dans l'étude des comportements collectifs des systèmes à plusieurs corps, et elle a pour cela développé des outils mathématiques et des concepts qui ont permis la compréhension des transitions de phases et des systèmes critiques. L'exemple le plus récent de cette démarche concerne le domaine des réseaux complexes.

De nombreux systèmes, aussi bien naturels qu'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 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 sensibles telles que le réseau électrique ou les réseaux de transports, dont dépendent de manière cruciale nos sociétés modernes.

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'intervention de la physique statistique est également justifiée par ses liens avec la théorie des graphes, et par sa capacité à proposer des outils pour la caractérisation statistique de grands ensembles de données (voir [encadré 1](#)). L'analyse des réseaux complexes a été rendue possible grâce à l'apparition récente de grandes bases de données permettant d'étudier systématiquement la topologie des grands réseaux complexes : les premiers projets de cartographie ont concerné la Toile et l'Internet. Graduellement, d'autres « cartes » sont apparues et ont permis 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, qui peuvent être facilement traduits dans le langage des mathématiques. Ceci a montré la possibilité de l'existence de principes

sous-jacents allant au-delà des particularités de chaque domaine, et donc l'intérêt de développer un cadre général de modélisation.

Principales caractéristiques

De nombreux réseaux complexes représentant des systèmes extrêmement divers partagent des caractéristiques communes. En particulier, nombre de ces systèmes sont des « petits-mondes », ce qui traduit le fait que la distance topologique moyenne dans le réseau (qui mesure le nombre moyen de liens à franchir sur le réseau pour aller d'un site à un autre) varie très lentement avec le nombre total de sites (typiquement comme un logarithme). 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 (sites de degré k), qui est décrite par une loi de puissance, indiquant ainsi l'absence de toute échelle caractéristique (voir [encadré 1](#)). En d'autres termes, pour ces 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. Ce résultat a permis l'identification d'une nouvelle catégorie de réseaux dits « sans-échelle », et nous allons voir que leurs propriétés topologiques sont cruciales pour un grand nombre de propriétés physiques du système représenté, telles que sa résistance à l'endommagement, sa vulnérabilité face à une attaque ou bien encore la propagation d'épidémies.

Changement de paradigme

Pendant longtemps, on a considéré que, lorsqu'une représentation d'un système en termes de réseau était adaptée, on pouvait utiliser le paradigme d'un ensemble



Encadré 1

Caractérisation statistique des réseaux

Un graphe est défini par un ensemble V de N points (ou sommets ou nœuds) et E de liens entre ces points. Un lien est donc un couple de points $(i, j) \in V^2$. Pour tout $i \in V$, on appelle *degré* k_i de i le nombre de sommets j tels que $(i, j) \in E$, c'est-à-dire le nombre de voisins de i sur le graphe. Une première caractérisation statistique est la distribution $P(k)$ de ces degrés : par définition $P(k) = N_k/N$ où N_k est le nombre de sommets de degré k . La classe de degré k regroupe tous les sommets ayant le même degré k . On distinguera schématiquement les graphes ou réseaux homogènes pour lesquels $P(k)$ décroît très rapidement (exponentiellement ou encore plus vite) si k s'éloigne de la valeur moyenne $\langle k \rangle = \sum_k kP(k)$, des réseaux hétérogènes pour lesquels $P(k)$ s'étend sur de nombreux ordres de grandeur et décroît lentement à grand k , par exemple en loi de puissance $P(k) \sim k^{-\gamma}$ (voir **figure 1**). Le rapport $\kappa = \langle k^2 \rangle / \langle k \rangle$ entre fluctuations et moyenne de $P(k)$ intervient souvent pour déterminer le comportement de phénomènes dynamiques sur le réseau (voir **encadrés 2 et 3**).

Corrélations : une caractérisation simple des corrélations entre les degrés de sommets voisins est donnée par le degré moyen des voisins d'un sommet i , $k_{nn, i} = (\sum_{j \text{ voisin de } i} k_j) / k_i$. A partir de cette quantité, une mesure commode des corrélations entre les degrés est obtenue par le degré moyen des voisins des sommets de degré k ,

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i/k_i=k} k_{nn, i} \quad (1)$$

Le comportement de cette fonction de k définit deux grandes classes de réseaux : on parle d'« assortativité » lors-

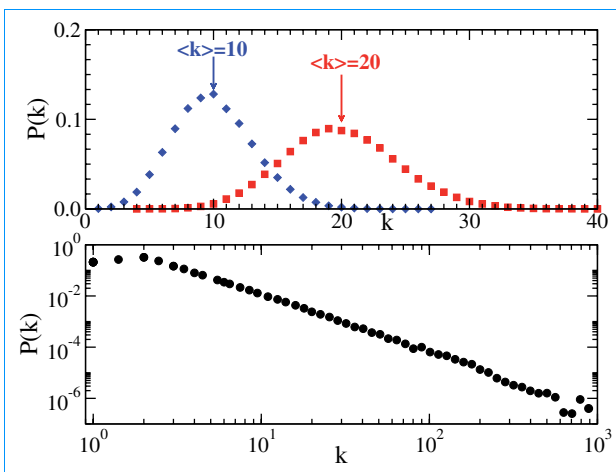


Figure 1 – Distribution $P(k)$ des degrés pour des réseaux homogènes de degré moyen $\langle k \rangle = 10$ et $\langle k \rangle = 20$ (haut) et un réseau hétérogène (bas) avec $P(k) \propto k^{-\gamma}$ (ici $\gamma = 2,3$).

que $k_{nn}(k)$ croît avec k , et de « disassortativité » si $k_{nn}(k)$ décroît. L'assortativité, présente par exemple dans les réseaux sociaux, correspond au fait que des sommets de fort degré sont préférentiellement liés à d'autres sommets de fort degré. Le cas contraire se rencontre dans des réseaux hiérarchiques pour lesquels les « hubs » sont connectés à de nombreux sommets de faible degré.

Le coefficient de « clustering » c_i mesure la densité locale de liens : il est défini pour chaque sommet i comme la fraction de voisins de i qui sont connectés entre eux. Le coefficient de clustering moyen $C = N^{-1} \sum_i c_i$ quantifie donc l'aspect cohésif du réseau, en donnant la densité moyenne de triangles présents. Dans de nombreux réseaux réels, un clustering important est observé, en fort contraste avec les graphes aléatoires.

Citons finalement la décomposition dite en « k -cores », qui permet de découvrir les éléments les plus « centraux » d'un réseau : le « core », ou noyau, d'indice k est le plus grand sous-graphe (c'est-à-dire un sous-ensemble de points de V , avec les liens correspondants) dans lequel tous les éléments ont un degré au moins k . L'étude de la structure hiérarchique ainsi définie permet de mieux caractériser les réseaux. Un algorithme appelé « LaNet-vi » exploite cette décomposition afin de proposer une visualisation de grands réseaux complexes, comme montré pour un exemple en **figure 2** (voir <http://xavier.informatics.indiana.edu/lanet-vi/> pour d'autres exemples).

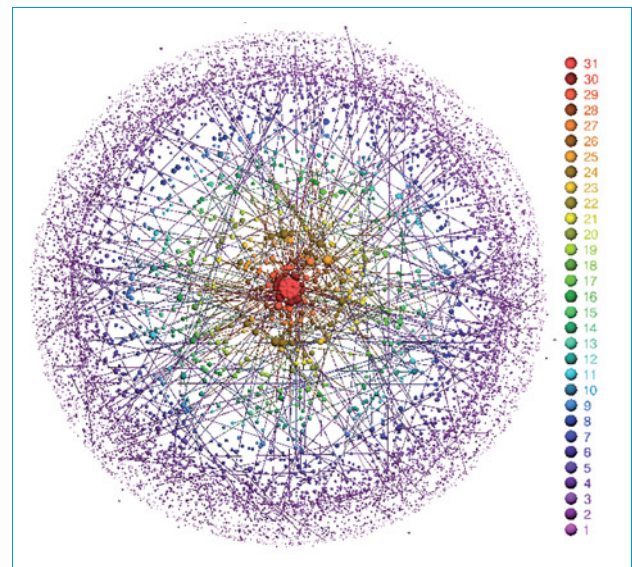


Figure 2 – Une image du réseau Internet, créée par l'algorithme LaNet-vi : chaque site représente un système autonome (ensemble de réseaux locaux sous le contrôle d'une seule et même entité, fournisseur d'accès à Internet par exemple), et un lien correspond à l'existence d'une communication directe entre deux systèmes autonomes.



de points reliés aléatoirement, c'est-à-dire du graphe aléatoire proposé par les mathématiciens Erdős et Renyi dans les années 60. Ces graphes sont cependant homogènes, dans le sens où le nombre de voisins de chaque site fluctue très peu autour d'une valeur moyenne. Les observations empiriques récentes ont 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 une approche de modélisation de processus microscopiques qui permettent l'émergence spontanée de ces caractéristiques. Citons la plus fameuse de ces approches : l'attachement préférentiel du modèle de Barabási et Albert. Dans ce modèle, le réseau est créé à partir de deux principes simples : (i) des nouveaux sites sont ajoutés un par un, et (ii) chaque nouveau site se connecte avec m liens au réseau préexistant, en choisissant les sites auxquels se connecter avec une probabilité proportionnelle au nombre de voisins des sites préexistants. En d'autres termes, pour chaque sommet préexistant i , ayant un nombre de voisins k_i , la probabilité de recevoir une nouvelle connexion est proportionnelle à k_i . Ce principe d'attachement préférentiel traduit l'idée que pour Internet par exemple, un nouveau fournisseur d'accès va établir préférentiellement des connexions vers d'autres fournisseurs bien connectés, ou qu'une nouvelle page web va typiquement contenir des liens vers d'autres pages déjà très connues. Il est particulièrement remarquable que ce principe très simple permet d'obtenir des réseaux ayant une distribution des degrés en loi de puissance, précisément $P(k) \sim k^{-3}$, donc très hétérogène et avec des fluctuations divergentes (voir [encadré 1](#)). Bien entendu, ce mécanisme est trop simple pour être réaliste, mais il présente l'énorme avantage de pouvoir servir de base à une pléthore d'enrichissements possibles.

Phénomènes dynamiques

La mise en évidence du caractère hétérogène de la topologie des réseaux complexes a mené à la question naturelle de l'influence de ces caractéristiques sur les phénomènes dynamiques prenant place sur ces réseaux. De nombreuses études ont donc été réalisées dans ce cadre, depuis des problèmes fondamentaux jusqu'à des processus plus appliqués.

Modèle d'Ising

L'étude des transitions de phase représente probablement l'exemple le plus frappant des potentialités de la physique statistique, et le domaine où se trouvent ses réalisations les plus marquantes. En particulier, le modèle d'Ising représente le paradigme le plus connu de l'émergence d'un comportement global (aimantation macroscopique) à partir de règles d'interactions micro-

scopiques (interactions entre spins voisins). Le comportement macroscopique résulte en effet de la compétition entre ces interactions microscopiques tendant à aligner les spins et la tendance au désordre due à l'agitation thermique. A haute température, le désordre est prépondérant tandis qu'à basse température (en dessous d'une certaine température critique) apparaît un ordre spontané à longue portée. L'existence, la nature et les propriétés de la transition entre les phases de haute et basse température dépendent de la symétrie des interactions microscopiques et de la topologie de ces interactions. Deux cas ont été particulièrement étudiés. D'une part, si les spins sont situés sur un réseau complet, pour lequel chaque sommet est connecté à tous les autres, chaque spin ressent le « champ moyen » de tous les autres, ce qui permet de négliger les fluctuations et d'obtenir des solutions analytiques. D'autre part, les réseaux réguliers en dimension finie (réseau carré, réseau cubique, etc.) permettent une modélisation plus proche de systèmes magnétiques réels. D'un point de vue fondamental, il est intéressant de considérer l'effet de différentes topologies d'interaction entre les spins situés aux sommets du réseau sur l'existence et la nature d'une transition de phase. Il a ainsi été montré que le fait que le nombre de voisins puisse varier beaucoup d'un site à l'autre a des conséquences importantes sur la transition ferromagnétique. En particulier, une fraction infime de liens à longue portée ajoutée à un réseau régulier en dimension finie conduit à changer la nature de la transition vers un régime de type champ moyen. D'autre part, si les spins sont situés sur les sites d'un réseau hétérogène pour lequel les fluctuations de degré divergent, la température de transition peut elle aussi diverger, car les sites à forte connectivité polarisent le système à toute température. On arrive ainsi à un effet particulièrement frappant : la disparition complète de la phase « haute température ».

Modèles de formation d'opinion

Les modèles de type Ising ont en fait un champ d'application qui va bien au-delà du magnétisme. En particulier, ils forment la base de nombreux modèles d'interaction sociale, de formation d'opinion ou de propagation de connaissances. Citons par exemple le modèle des électeurs, où chaque individu peut avoir deux opinions différentes (donc pouvant être modélisées par une variable de spin valant + ou -1). A chaque instant, un individu pris au hasard choisit un de ses voisins et adopte son opinion. La convergence vers un état uniforme d'opinion a été étudiée par les physiciens statisticiens en particulier pour des individus pouvant interagir sur un réseau régulier en dimension finie. Cependant, l'étude des réseaux complexes a mis en évidence la très forte hétérogénéité des interactions sociales : une majorité d'individus sont peu connectés tandis qu'une fraction non négligeable de personnes a de nombreuses connaissances. En conséquence, de nou-



velles études se sont intéressées à la dynamique et à l'évolution de tels modèles définis sur des réseaux d'interactions plus réalistes, fortement hétérogènes. En particulier, ces études ont permis de mieux caractériser le rôle fondamental de l'hétérogénéité dans la propagation d'opinions ou de connaissances.

Résistance à des pannes ou à des attaques

Dans de nombreux domaines, il est essentiel de comprendre et caractériser la stabilité des réseaux, et leur réponse à une perturbation extérieure comme par exemple la suppression d'un certain nombre de sommets. Par exemple, dans le cas de grandes infrastructures comme Internet ou le réseau de distribution d'électricité, l'arrêt ou le dysfonctionnement local de serveurs ou de générateurs peuvent affecter les propriétés globales de communication et de transport. La modélisation et l'étude de telles perturbations permettent en particulier d'identifier les parties les plus sensibles du réseau et de mieux les protéger.

Il a été réalisé tout d'abord empiriquement que la structure très hétérogène des réseaux complexes implique une grande résistance à des pannes, c'est-à-dire au cas où des sites sont retirés aléatoirement, mais en même temps une grande fragilité devant des attaques ciblées sur les sites les mieux connectés. Un tel comportement est observé aussi bien pour des réseaux réels comme ceux représentant Internet que pour les modèles de réseaux sans échelle. La [figure 1](#) montre, pour un réseau hétérogène, 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

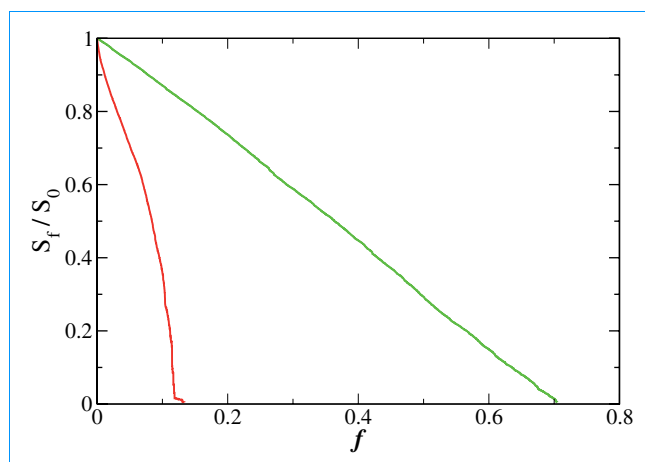


Figure 1 – Résistance d'un réseau complexe à la suppression de sommets. Les courbes représentent la taille de la plus grande composante connexe du réseau S_f (normalisée à la taille initiale) en fonction de la fraction f de sites éliminés. Les sites sont retirés soit au hasard (courbe verte) soit par nombre de voisins décroissant (courbe rouge).

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), et donc 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 clairement à une très rapide désintégration du réseau en petites composantes isolées. En effet, ces sites de fort degré sont plus centraux car ils mettent en communication de nombreuses parties du réseau et leur suppression a donc des conséquences beaucoup plus importantes. Comme expliqué dans l'[encadré 2](#) pour le cas des pannes, il est possible de comprendre plus 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 après suppression aléatoire d'une fraction de sites ou de liens). Les fluctuations de connectivité apparaissent ainsi clairement comme responsables de cette ambivalence entre robuste et vulnérabilité des réseaux complexes.

Enfin, il convient de mentionner les phénomènes de cascades, qui sont à la base par exemple des gigantesques pannes électriques qui se sont produites dans plusieurs 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 modéliser 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 – si la surcharge initiale est faible par exemple – soit se propager sur une grande portion du réseau, causant une saturation globale. De nombreuses études récentes ont ainsi cherché à modéliser le trafic sur des réseaux complexes et à comprendre dans quelles conditions un incident initial, éventuellement mineur, peut causer un dysfonctionnement à grande échelle. Des travaux récents tendent également à développer des stratégies de réaction rapides afin d'empêcher une cascade de se développer.

Propagation d'épidémies

Un autre exemple de l'impact des études des réseaux complexes réside dans la meilleure compréhension des phénomènes de propagation d'épidémies sur des réseaux hétérogènes. 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



Encadré 2

Résistance aux pannes

Une panne dans un réseau technologique est définie par le fait qu'un certain nombre de sommets cessent d'exercer leur fonction, qui peut être par exemple celle de transmettre de l'information. L'étendue des dommages causés au réseau est en général mesurée par la taille de la plus grande composante connexe du graphe après qu'une fraction f de sommets en a été retirée. Si cette composante (ou amas) a une taille extensive, c'est-à-dire qui diverge si la taille du réseau initial diverge, on parle d'amas géant et on considère que le réseau peut encore fonctionner.

L'effet de la suppression de sites dans un réseau peut être compris et quantifié une fois le problème traduit en termes de percolation. En effet, retirer une fraction f de sites au hasard est équivalent à écrire que chaque site est présent avec probabilité $p = 1 - f$. Un amas géant existe alors si et seulement si p est plus grand qu'une valeur critique p_c , appelée seuil de percolation.

On note q la probabilité qu'un lien choisi au hasard n'amène pas à un sommet appartenant à un amas géant. En négligeant l'existence de cycles et de corrélations, on peut écrire q comme la moyenne sur tous les degrés k possibles du produit de deux probabilités : (i) la probabilité que le lien choisi mène à un site de degré k , soit $kP(k)/\langle k \rangle$ (ii) la probabilité qu'aucun des $k - 1$ liens restant ne mène à un sommet appartenant à un amas géant, soit q^{k-1} . Ainsi, on obtient :

$$q = \sum_k \frac{kP(k)}{\langle k \rangle} q^{k-1} \quad (1)$$

qui admet toujours la solution $q = 1$. Une composante géante existe s'il existe une autre solution à cette équation, c'est-à-dire si :

$$\left. \frac{d}{dq} \left(\sum_k \frac{kP(k)}{\langle k \rangle} q^{k-1} \right) \right|_{q=1} \geq 1. \quad (2)$$

sur le réseau Internet ou sur le réseau d'échange de courriers électroniques. Les modèles communément utilisés en épidémiologie peuvent en fait être exprimés sous la forme de modèles de type réaction-diffusion : 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 clairement un rôle important, et les outils de la physique statistique hors d'équilibre permettent de l'appréhender (voir [encadré 3](#)). Ce genre de systèmes présente généralement une transition de phase hors d'équilibre 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, et on s'intéresse généralement à la détermination de sa valeur critique ainsi qu'à des exposants critiques associés et à leur possible universalité. L'[encadré 3](#) montre, dans le

Cette condition se ré-écrit :

$$\langle k^2 \rangle \geq 2 \langle k \rangle \quad (3)$$

qui est le critère de Molloy-Reed pour l'existence d'un amas géant dans un graphe aléatoire sans corrélations, de distribution de degrés $P(k)$.

Le critère précédent permet de comprendre le problème des pannes. Considérons un réseau initial de distribution de degrés $P_0(k)$ (degré moyen $\langle k \rangle_0$, second moment $\langle k^2 \rangle_0$). On peut déterminer la distribution des degrés $P_f(k)$ (et les moments $\langle k \rangle_f$ et $\langle k^2 \rangle_f$) après la suppression aléatoire d'une fraction f des sites. En effet, pour chaque site non supprimé, de degré initial k_0 , ce processus aléatoire implique la suppression avec probabilité f de chacun de ses voisins et donc un degré final k ($< k_0$) avec probabilité $C_{k_0}^k (1-f)^k f^{k_0-k}$. La distribution $P_f(k)$ est obtenue en sommant cette expression sur toutes les valeurs de k_0 , pondérées par $P_0(k_0)$, et on obtient $\langle k \rangle = (1-f) \langle k \rangle_0$ et $\langle k^2 \rangle_f = (1-f)^2 \langle k^2 \rangle_0 + f(1-f) \langle k \rangle_0$. La valeur critique f_c à laquelle le réseau est détruit est telle que pour $f > f_c$ aucun amas géant n'existe, ce qui, d'après le critère de Molloy-Reed, correspond à $\langle k^2 \rangle_{f_c} = 2 \langle k \rangle_{f_c}$, soit :

$$f_c = 1 - \frac{\langle k \rangle_0}{\langle k^2 \rangle_0 - \langle k \rangle_0}. \quad (4)$$

Cette formule permet de comprendre la distinction importante entre réseaux homogènes pour lesquels les fluctuations de degré sont limitées, et les réseaux hétérogènes où $\langle k^2 \rangle$ peut être très grand, voire diverger quand la taille du réseau augmente, ce qui donne $f_c = 1$, donc une extrême robustesse à des pannes aléatoires.

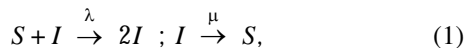
cas d'un modèle simple, la conséquence drastique de l'hétérogénéité du réseau sur le seuil critique de propagation, qui tend vers 0 si les fluctuations de degré divergent. Aussi faible que soit la transmissibilité, l'épidémie peut alors s'étendre et survivre dans le réseau, grâce au fait qu'il existe 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 « super-propagateurs ».



Encadré 3

Propagation d'épidémies

Les modèles communément utilisés en épidémiologie considèrent que chaque individu peut à chaque instant être dans un état précis décrivant un stade de l'infection. Le modèle le plus simple considère seulement deux états possibles, à savoir « sain » (S) et « infecté » (I), tandis que de nombreuses variantes peuvent être envisagées, en introduisant les états « remis » (R) et donc immunisé contre une réinfection, « latent » (L), « infecté a-symptomatique », etc. Le passage d'un individu entre ces différents stades se fait selon des lois qui peuvent s'écrire comme des réactions chimiques. Le modèle SIS par exemple s'écrit



où la première réaction indique qu'un individu sain en contact avec un infecté peut devenir infecté avec taux λ (c'est-à-dire avec probabilité $\lambda \Delta t$ à chaque intervalle de temps $\Delta t \ll 1$), tandis qu'un infecté guérit spontanément avec taux μ , et redevient ainsi sain, et susceptible d'être infecté de nouveau (une guérison avec immunisation correspondrait à $I \xrightarrow{\mu} R$, c'est-à-dire au modèle dit SIR).

L'évolution d'une épidémie, dans une population de N individus, est mesurée par la densité d'infectés ou *prévalence* $i(t) = I(t)/N$ ($I(t)$ étant le nombre d'individus infectés au temps t). L'équation gouvernant cette évolution peut être écrite, dans le cas où le nombre de voisins k de tous les individus est à peu près le même (c'est-à-dire dans l'approximation $k = \langle k \rangle$):

$$\frac{di(t)}{dt} = \lambda \langle k \rangle i(t) s(t) - \mu i(t). \quad (2)$$

Le premier terme correspond à une augmentation du nombre $I(t)$ d'infectés proportionnelle au nombre de susceptibles ($s(t) = 1 - i(t)$ est la densité de susceptibles), multiplié par la probabilité $\langle k \rangle i(t)$ que ces susceptibles aient un infecté parmi leurs voisins, et par la probabilité de transmission. Le deuxième terme correspond à la guérison spontanée. Dans l'état stationnaire, on voit donc que $i = 0$ ou $i = 1 - \mu/(\lambda \langle k \rangle)$, la deuxième solution étant possible seulement si la transmissibilité est assez grande, donc pour $\lambda > \lambda_c$ avec $\lambda_c = \mu/\langle k \rangle$.

Dans le cas de réseaux hétérogènes, certains sites ont un degré k très différent de $\langle k \rangle$, et on doit considérer non seu-

lement $i(t)$ mais aussi la densité d'individus infectés de degré k , i_k . Il est possible d'écrire l'équation d'évolution des $i_k(t)$ et d'en déduire le critère pour l'existence d'une prévalence i non nulle dans l'état stationnaire, qui s'écrit :

$$\lambda > \lambda_c = \mu \frac{\langle k \rangle}{\langle k^2 \rangle} \quad (3)$$

On voit apparaître, comme dans l'*encadré 2*, le rapport entre les deux premiers moments de la distribution des degrés. Pour un réseau homogène, ce rapport est fini et donc un seuil critique $\lambda_c > 0$ sépare une phase active où la prévalence est finie à $\lambda > \lambda_c$ d'une phase inactive où l'épidémie s'éteint spontanément (dans l'approximation où tous les sommets ont degré $\langle k \rangle$ on retrouve $\lambda_c = \mu/\langle k \rangle$). Dans le cas d'un réseau hétérogène au contraire, $\langle k^2 \rangle$ peut diverger (par exemple si $P(k) \sim k^{-\gamma}$ avec $\gamma \leq 3$) et donc $\lambda_c \rightarrow 0$. Aussi faible que soit la transmissibilité, l'épidémie peut alors se propager. Bien entendu, la prévalence i dans l'état stationnaire est d'autant plus faible que λ est faible. La raison fondamentale de l'annulation du seuil critique réside dans le rôle des sites très connectés, qui peuvent infecter de nombreux autres sites du réseau. Ceci implique également que les stratégies de vaccination ou de protection doivent être ciblées prioritairement sur ces sites fortement connectés.

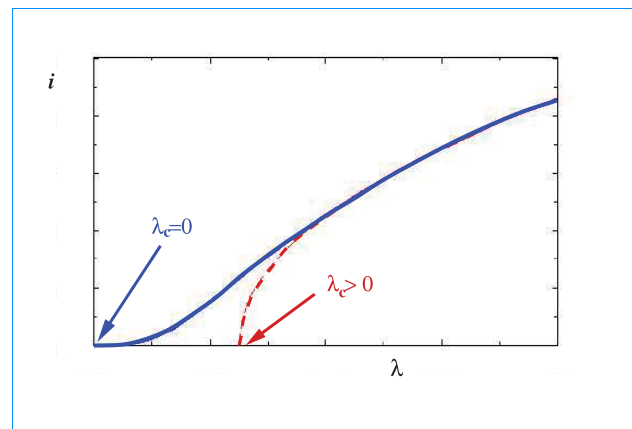


Figure 1 – Variation schématique de la prévalence i en fonction de la transmissibilité λ , dans l'état stationnaire du modèle SIS, pour un réseau homogène (traits interrompus) et pour un réseau hétérogène (trait plein).

Un pas supplémentaire dans l'interaction entre approches de physique statistique et épidémiologie est franchi en considérant 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 (où par exemple on utilise l'approximation de type champ moyen que chaque individu est potentiellement en contact avec tous les autres) et une dynamique de voyage entre les villes, données par la structure

complexe et hétérogène du réseau de transport. On peut se demander où intervient la physique statistique ; d'une part, les modèles utilisés s'écrivent en fait comme une série d'équations de Langevin, pour lesquelles les techniques d'intégration numérique 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 peut être quantifiée par une « entropie », et sa prévisibilité 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.

Conclusion et perspectives

L'étude des réseaux complexes est une activité en plein essor. Nous avons donné un aperçu des progrès réalisés récemment, en particulier grâce aux apports de la physique statistique. Notons que de nouvelles directions se développent, par exemple vers l'étude des réseaux pondérés, des réseaux dynamiques (de type pair à pair), et des réseaux de réseaux. Au-delà de la compréhension de l'émergence de la topologie des réseaux et de la dynamique de processus se déroulant sur ces réseaux, l'interaction réciproque entre topologie du réseau et dynamique des processus reste également un sujet largement ouvert.

POUR EN SAVOIR PLUS

- Albert (R.), Barabási (A.-L.), « *Rev. Mod. Phys.* », *74*, 2002, 47.
- Dorogovtsev (S.N.), Mendes (J.F.F.), « *Adv. Phys.* », *51*, 2002, 1079.
- Barabási (A.-L.), « *Linked : The new science of networks* », *Perseus publishing*, 2002.
- Dorogovtsev (S.N.), Mendes (J.F.F.), « *Evolution of networks : From biological nets to the Internet and WWW* », *Oxford University Press*, 2003.
- Pastor-Satorras (R.), Vespignani (A.), « *Internet : structure et évolution* », *Belin*, 2004.
- Albert (R.), Jeong (H.), Barabási (A.-L.), « *Nature* », *406*, 2000, 378.
- Pastor-Satorras (R.), Vespignani (A.), « *Phys. Rev., Lett.* », *86*, 2001, 3200.
- Colizza (V.), Barrat (A.), Barthélemy (M.), Vespignani (A.), « *Proc. Natl. Acad. Sci. USA* », *103*, 2006, 2015.