

Sommaire

Editorial	4
Comité de l'UPS élu en 2005	6
Etude des réseaux par la physique statistique	11
Par Alain BARRAT, Marc BARTHÉLÉMY et Alessandro VESPIGNANI	
Classification des phénomènes de halos	20
Par Luc DETTWILLER	
Rapport sur les problèmes posés aux concours 2005	35
Chimie	35
Mathématiques	37
Physique	44
Assemblée générale de l'UPS	45
Réunions des physiciens	45
Olympiades internationales de Physique	50
Olympiades internationales de Mathématiques	54
Comptes rendus de stages	56
Stage de perfectionnement à l'E.N.S.I.C	56
Forum scientifique et pédagogique	57
Forum Informatique	57
Forum Mathématique	59
Bibliographie	66
Mathématiques	66
Entre nous	67
Claudine MÉRESSE	67
Bruno NOËL	67

Etude des réseaux par la physique statistique

Par Alain BARRAT

Laboratoire de Physique théorique
CNRS et université d'Orsay

Marc BARTHÉLÉMY

Département de Physique théorique et appliquée, CEA
et School of Informatics, Indiana University, USA

Alessandro VESPIGNANI

School of Informatics, Indiana University, USA

1. Introduction

Depuis un peu plus d'une décennie, la communauté des physiciens statisticiens porte une attention sans cesse croissante à des systèmes en dehors de son champ d'application 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 très puissants qui ont permis la compréhension des transitions de phases et des systèmes critiques. L'exemple le plus récent de cette démarche concerne le domaine actuellement très actif des réseaux complexes.

Un réseau est constitué d'un ensemble de sites (ou sommets), pouvant être liés entre eux (figure 1). Dans cet exemple, les liens sont symétriques : si A est lié à B alors B est lié à A ¹. Considérons un exemple très simple : l'ensemble de sites est constitué des personnes vivant sur Terre à une date donnée, et on considère qu'il existe un lien entre deux personnes si elles se connaissent.

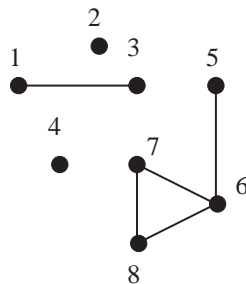


Figure 1. — Exemple de réseaux à 8 sites (sommets) et 5 liens.

1. Un exemple de liens non réciproques est donnée par la Toile (WWW) : une page peut contenir un hyperlien vers une autre page sans que la réciproque soit vraie.

L'étude des réseaux est par nature interdisciplinaire car ceux-ci jouent un rôle central dans la compréhension de nombreux systèmes 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 la structure de l'Internet ou de la Toile jusqu'aux interconnexions d'agents financiers ou bien aux réseaux trophiques des écosystèmes. On peut aussi citer les grandes infrastructures critiques telles que le réseau électrique ou le réseau des transports aériens, dont dépendent de manière cruciale nos sociétés modernes. Enfin, la cellule, constituant élémentaire des êtres vivants, n'est pas une exception et son organisation ainsi que ses fonctions sont aussi le résultat d'un réseau compliqué d'interactions entre gènes, protéines et autres molécules.

2. Une première approche : graphes aléatoires

Pendant longtemps, on a considéré que tous ces systèmes pouvaient être décrits par un ensemble de points reliés aléatoirement, suivant le prototype du graphe aléatoire proposé par les mathématiciens Erdős et Rényi [1] dans les années 60. Deux points distincts choisis aléatoirement possèdent une certaine probabilité p d'être connectés. Cette probabilité est alors indépendante du choix des points.

De tels graphes sont homogènes, dans le sens que la distribution de probabilité $P(k)$ du nombre de voisins ou connectivité k est centrée autour de la valeur moyenne $\langle k \rangle = pN$ (N est le nombre de sommets), avec des fluctuations limitées, c'est-à-dire que $\langle k^2 \rangle$ est fini, et $P(k)$ décroît exponentiellement rapidement à grand k (figure 2).

L'apparition ces dernières années de grandes bases de données, l'augmentation de la taille de la mémoire et des capacités de calcul des ordinateurs ont radicalement modifié cette vision. En particulier, les projets de cartographie de la Toile et de l'Internet ont offert une première chance d'étudier la topologie des grands réseaux complexes.

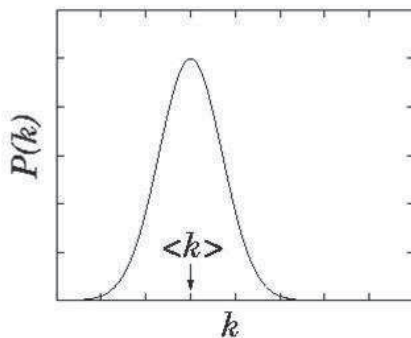


Figure 2. — Distribution de connectivité pour un réseau aléatoire : piquée autour de la valeur moyenne, avec décroissance exponentielle à grand k , typique de réseaux homogènes.

3. Vers un nouveau paradigme : réseaux sans échelle

Graduellement, d'autres « cartes » sont apparues et ont permis la description de nombreux réseaux intervenant dans les sciences sociales, les infrastructures critiques et la biologie. Les chercheurs de ce domaine ont alors commencé à étudier de manière systématique ces grands ensembles de données et ont mis en évidence des ordres

cachés et des motifs qui peuvent être considérés comme la manifestation de lois sous-jacentes gouvernant l'évolution de ces systèmes complexes. En effet, lors de l'étude de la structure des réseaux complexes, on observe qu'en dépit de leur apparente complexité et de leur caractère aléatoire, des motifs et des régularités statistiques apparaissent.

3.1. Caractéristiques

Plus spécifiquement, nombre de ces systèmes sont des « petits-mondes », ce qui traduit le fait que la distance topologique moyenne dans le réseau (qui est définie par 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). D'autre part, ces réseaux présentent en général une surabondance de triangles (triplets interconnectés) par rapport à des graphes aléatoires de même taille. Enfin, une découverte particulièrement importante est le fait que de nombreux réseaux sont caractérisés par une abondance statistique de « hubs », c'est-à-dire des sites qui ont un très grand 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 en première approximation par une loi de puissance, i.e. $P(k) \sim k^{-\gamma}$ avec typiquement $\gamma \in [2, 3]$, indiquant ainsi l'absence de toute échelle caractéristique (figure 3). En d'autres termes, pour ces réseaux, la notion de moyenne ou de site représentatif (ou typique) a peu de sens car les fluctuations de la connectivité d'un site à un autre sont gigantesques (en particulier si $\gamma \leq 3$, $\langle k^2 \rangle$ diverge avec la taille du système). Ce résultat a permis l'identification d'une nouvelle catégorie de réseaux dits « sans-échelle » dont les propriétés topologiques sont cruciales pour un grand nombre de propriétés physiques du système global. Par exemple, les hubs, à cause de leur grand nombre de voisins, sont très efficaces dans la propagation aussi bien d'informations dans le contexte de la propagation de nouvelles que de virus dans le contexte épidémiologique. D'autre part, dans un contexte de réseau technologique (Internet, réseau de transport d'énergie...), la défaillance de quelques uns de ces sites très connectés suffit à mettre en péril l'ensemble du fonctionnement du réseau.

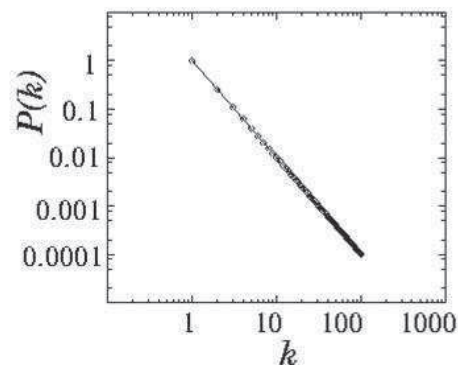


Figure 3. — Distribution de connectivité pour un réseau sans échelle : distribution en loi de puissance caractéristique (échelles logarithmiques).

3.2. Changement de paradigme : approche dynamique

Les tentatives de modélisation et de compréhension de l'origine des propriétés topologiques observées ont mené à un changement radical de perspective en passant des graphes statiques reproduisant le réseau à un instant donné à une vision dynamique

qui tente de reproduire l'évolution des réseaux. Cette nouvelle approche s'est basée sur deux observations fondamentales : (i) nombre de réseaux résultent d'un processus de croissance ; (ii) cette croissance ne se fait pas de façon totalement aléatoire : des sommets ayant déjà de nombreux voisins (donc « connus », ou « importants ») sont plus susceptibles d'en gagner de nouveaux. Cela peut se comprendre facilement par exemple pour les pages de la Toile : une nouvelle page va probablement créer des liens vers un site déjà populaire. Le nouveau paradigme est représenté par le modèle de Barabási et Albert (BA) [2], dont la construction combine la croissance du réseau et un ingrédient connu sous le nom d'attachement préférentiel : les nouveaux sites introduits dans le réseau se lient préférentiellement à des sites ayant déjà de nombreux voisins.

La règle de construction est la suivante :

- on part d'un certain nombre N_0 de sites, liés entre eux ;
- à chaque pas de temps, un nouveau site est introduit. Celui-ci établit m liens (m étant un paramètre du modèle) avec les sites préexistants selon une probabilité proportionnelle à leur connectivité : la probabilité de choisir un site i est donc $k_i / \sum_j k_j$ où k_i est le nombre de voisins du site i . Par exemple, dans le cas de la figure 1, la probabilité pour un nouveau site de se lier au site 1 est 0.1 alors qu'elle est de 0.3 pour le site 6 ;
- le processus est ensuite renouvelé par l'arrivée d'un nouveau site, jusqu'à l'obtention de la taille désirée du réseau.

Ces règles simples permettent d'obtenir un réseau sans-échelle, caractérisé par une distribution de connectivités en loi de puissance avec $\gamma = 3$. De nombreuses variations de ce modèle ont été proposées et étudiées par la suite : les réseaux complexes sont maintenant typiquement décrits comme des systèmes dynamiques qui évoluent par addition (ou soustraction) de liens et de nœuds. L'ensemble des règles dynamiques qui définit ces processus résume la théorie dynamique nécessaire à la compréhension des propriétés macroscopiques des réseaux.

C'est dans ce contexte que la physique statistique se propose de relier les diverses propriétés à grande échelle aux interactions locales qui gouvernent l'évolution des éléments constitutifs du système. Ce n'est donc pas par hasard que de nombreux articles fondateurs de ce domaine ont été publiés par des physiciens et ont eu un grand impact sur d'autres disciplines. Cependant, une approche véritablement interdisciplinaire est nécessaire afin de comprendre les spécificités de réseaux particuliers. Cette interdisciplinarité a conduit à l'émergence d'un nouveau domaine de recherche parfois défini comme la « science des réseaux » [3, 4, 5].

4. Réseaux pondérés

L'étude de nombreux réseaux réels ne saurait être complète sans une caractérisation de la diversité des liens entre les nœuds du réseau, et des éventuelles corrélations entre la topologie et les intensités des liens, ou le trafic prenant place sur ces liens. Par exemple, dans le réseau Internet, certaines connexions offrent une bande passante bien supérieure à d'autres ; dans les réseaux sociaux, certains liens entre personnes ou entités sociales sont nettement plus forts que d'autres ; dans les réseaux de nature biologique également, certaines réactions chimiques (par exemple) portent des flux énormes tandis que d'autres sont marginales. Les réseaux pondérés prennent naturellement en compte ces facteurs.

4.1. Quelques définitions et outils

— **Poids**

Les propriétés d'un graphe sont encodées dans la matrice d'adjacence (a_{ij}) , dont les éléments valent 1 s'il existe un lien connectant le site i au site j , et 0 sinon (avec $i, j = 1, \dots, N$ où N est la taille du réseau). Les réseaux pondérés sont de même décrits par une matrice de poids (w_{ij}) qui spécifie les poids des liens (avec $w_{ij} = 0$ si i et j ne sont pas connectés). On considérera seulement le cas de poids symétriques et positifs $w_{ij} = w_{ji} \geq 0$.

— **Distributions de connectivité et de poids**

La caractérisation habituelle d'un réseau est obtenue par l'analyse de la distribution de probabilité $P(k)$ qu'un site possède k voisins. Les réseaux complexes exhibent souvent une distribution en loi de puissance $P(k) \sim k^{-\gamma}$ avec $2 \leq \gamma \leq 3$, en fort contraste avec les graphes aléatoires pour lesquels $P(k)$ décroît exponentiellement. De façon similaire, une première caractérisation des poids des liens est obtenue par leur distribution de probabilité $P(w)$.

— **Connectivité pondérée**

Un site donné i est tout d'abord caractérisé par son nombre de voisins k_i . Certains voisins peuvent cependant être plus importants que d'autres, ce pourquoi il est nécessaire de considérer la connectivité pondérée définie par

$$s_i = \sum_{j \in \mathcal{V}(i)} w_{ij} \quad (1)$$

où la somme porte sur l'ensemble $\mathcal{V}(i)$ des voisins de i . Cette quantité concentre l'information à la fois sur la connectivité et sur l'importance des poids des liens du site i : c'est la généralisation naturelle de la connectivité. Dans le cas où les poids et la topologie sont indépendants, on obtient pour la connectivité pondérée des sommets de degré k : $s(k) \approx \langle w \rangle k$ où $\langle w \rangle$ est le poids moyen. En présence de corrélations on peut s'attendre à $s(k) \approx Ak^\beta$ avec $\beta > 1$ ou $\beta = 1$ mais $A \neq \langle w \rangle$.

— **« Clustering » et clustering pondéré**

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 (dans le cas de la figure 1, $c_6 = 1/3$ et $c_7 = c_8 = 1$). 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 triplets interconnectés. Des informations plus détaillées sont obtenues par l'étude du coefficient $C(k)$ moyenné sur les sommets de degré k .

Cette caractérisation purement topologique ne tient cependant pas compte du fait que certains voisins peuvent être plus importants que d'autres. On peut donc introduire un coefficient de clustering pondéré, qui combine les informations sur la topologie et sur les poids [7] :

$$c_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{w_{ij} + w_{ih}}{2} a_{ij} a_{ih} a_{jh}. \quad (2)$$

Pour chaque triangle formé dans le voisinage d'un sommet i , c_i^w compte donc le poids des deux liens issus de i . Ainsi, on ne considère pas seulement le nombre

de triangles auxquels participe i mais aussi leur poids total par rapport à la connectivité pondérée du sommet i . Le facteur de normalisation $s_i(k_i - 1)$ assure que $0 \leq c_i^w \leq 1$. On définit également C^w et $C^w(k)$ comme le coefficient de clustering pondéré moyenné sur tout le réseau ou sur la classe de sommets de degré k . Dans le cas soit de poids homogènes $w_{ij} = \text{constante}$, soit de poids aléatoires sans corrélations avec la topologie, on retrouve $C^w = C$ et $C^w(k) = C(k)$. Dans des réseaux réels cependant, on peut avoir *a priori* des cas différents : si $C^w > C$, les triangles sont plutôt formés par des liens avec des grands poids ; dans le cas contraire, le clustering topologique est en fait engendré par des liens peu importants et l'existence de triangles a une importance moindre dans l'organisation du réseau puisque les poids les plus importants sont répartis sur d'autres liens.

4.2. Résultats empiriques

Grâce aux outils précédemment introduits, nous avons analysé deux réseaux pondérés de natures différentes :

- le réseau des collaborations scientifiques est un exemple de réseau social ; les sommets du réseau sont les auteurs et les liens représentent l'existence d'une collaboration, matérialisée par des articles cosignés. Le réseau analysé comprend les personnes ayant soumis des manuscrits à l'archive de prépublications de matière condensée² entre 1995 et 1998. Il est formé de $N = 12\,722$ sommets, avec un degré moyen (nombre moyen de collaborateurs) $\langle k \rangle \approx 6.28$, et degré maximum 97. Le poids d'un lien a été défini comme suit : l'intensité d'une collaboration entre deux auteurs i et j est $w_{ij} = \sum_p \delta_i^p \delta_j^p / (n_p - 1)$, où p désigne les articles, n_p est le nombre d'auteurs de l'article p , et δ_i^p est 1 si l'auteur i a signé le papier p , et 0 sinon. Cette définition semble assez objective : elle donne un poids qui augmente avec le nombre d'articles cosignés, mais la contribution d'un article donné est inversement proportionnelle au nombre de ses auteurs.
- le réseau des connexions aériennes est par contre un exemple de grand réseau d'infrastructure. Il comprend $N = 3\,880$ aéroports et les liens correspondent aux liaisons aériennes directes³. Le degré moyen est $\langle k \rangle \approx 9.7$, et le degré maximum 318. L'intensité d'une connexion est donnée par le nombre de places disponibles sur cette connexion sur une année.

► Connectivité et connectivité pondérée

L'analyse des propriétés topologiques de ces réseaux a montré l'existence de lois larges pour les distributions de connectivité [6] (figures 4, page suivante). Il s'agit donc clairement de deux réseaux hétérogènes. De plus, la distribution des connectivités pondérées présente également des queues larges, ce qui n'est pas très surprenant puisque s_i (connectivité pondérée du sommet i) croît avec k_i (nombre de voisins du sommet i). D'autre part, des informations intéressantes sont obtenues en étudiant les corrélations entre topologie et poids des liens : la figure 5 (page suivante) montre le comportement obtenu aussi bien pour les vrais réseaux que pour une version « aléatoire » dans laquelle les intensités des connexions sont redistribuées au hasard. Pour le réseau des collabo-

2. <http://xxx.lanl.gov/archive/cond-mat>.

3. Ces données viennent de la base de données IATA (International Air Transportation Association, <http://www.iata.org>).

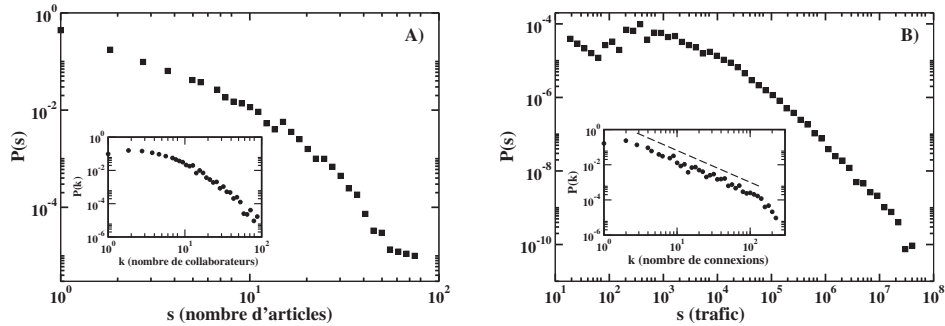


Figure 4.—A) Distributions de connectivité k (insert) et de connectivité pondérée s pour les collaborations scientifiques ; k représente le nombre de co-auteurs, et s le nombre de publications. B) Cas du réseau des connexions aériennes. k est le nombre de connexions vers d'autres aéroports tandis que s correspond au nombre total de passagers passant par un aéroport.

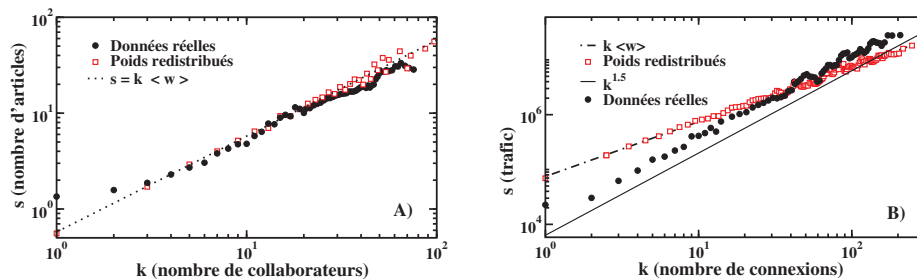


Figure 5.—Connectivité pondérée moyenne $s(k)$ en fonction du degré k . A) Pour le réseau des collaborations scientifiques les données sont similaires au cas du réseau avec poids aléatoirement redistribués. B) Pour le réseau des connexions aériennes $s(k)$ est en loi de puissance avec exposant $\beta \approx 1.5$, ce qui montre l'existence de corrélations fortes entre topologie et trafic.

rations, on obtient dans les deux cas $s(k) \approx \langle w \rangle k$. Par contre, pour le réseau du transport aérien, un exposant «anormal» $\beta = 1.5$ est obtenu : la connectivité pondérée croît donc plus vite que la connectivité, ce qui signifie que les intensités des liens appartenant aux sommets les plus connectés sont plus grandes que ce qui serait obtenu avec des poids aléatoires. On observe donc une très forte corrélation entre topologie et trafic dans ce réseau.

► Etude du clustering

— Collaborations scientifiques

Le spectre de clustering $C(k)$ présente une décroissance monotone. Cette structure montre que les auteurs ayant peu de collaborateurs (faible degré) font habituellement partie d'un groupe de recherche bien défini dans lequel tous les chercheurs travaillent ensemble. Par contre, les auteurs à grand degré collaborent avec différents groupes ou communautés qui n'ont que peu de collaborations, d'où un plus faible nombre de triangles. L'inspection de $C^w(k)$ montre cependant que pour $k \geq 10$ le clustering pondéré est nettement supérieur au clustering

topologique. Ceci implique que les auteurs ayant de nombreuses collaborations ont tendance à publier plus d'articles avec des « cliques » interconnectées, et constitue donc un signe que les chercheurs ayant beaucoup d'influence forment des groupes de recherche stables où se situe la plus grande part de leur production.

— Réseau de transport aérien

Dans ce cas, l'analyse des corrélations pondérées nous offre un complément important d'informations. En effet, la décroissance du clustering topologique $C(k)$, conséquence du rôle des grands aéroports qui fournissent des connexions vers des destinations lointaines qui ne sont pas directement connectées entre elles, est beaucoup plus prononcée que celle du clustering pondéré. Ceci montre que les aéroports de fort degré ont une tendance à former des groupes interconnectés avec un fort trafic. De plus, les aéroports ayant de nombreuses connexions ont une plus grande affinité pour les autres grands aéroports, et la plus grande part du trafic réside sur ces liens.

Ce genre de réseau est très fortement affecté par la disparition de quelques uns des sites à forte connexion : la possibilité de joindre deux sites quelconques par voie aérienne serait quasiment réduit à néant par la fermeture des quelques plus importants aéroports mondiaux.

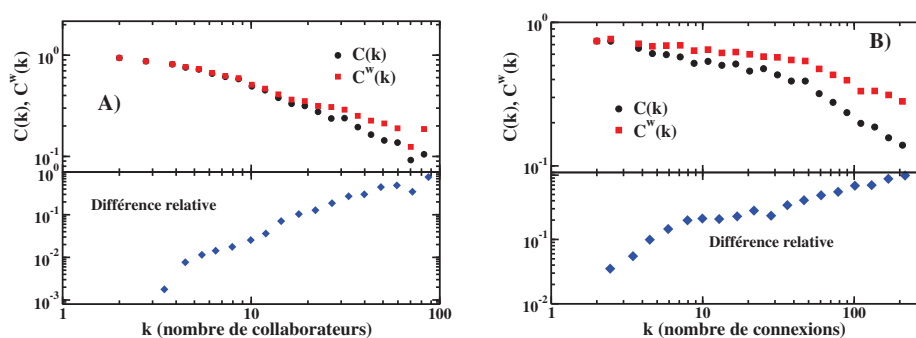


Figure 6.— Comparaison des corrélations topologiques et pondérées pour A) le réseau des collaborations scientifiques et B) le réseau des connexions aériennes. Le clustering pondéré est supérieur au clustering pour $k \geq 10$ en A) et sur toute la gamme des degrés pour B).

5. Conclusions et perspectives

La physique statistique fournit des outils particulièrement adaptés à la compréhension de l'émergence de propriétés complexes globales à partir de lois simples « microscopiques ». En interaction avec d'autres domaines scientifiques aussi divers que l'informatique, les sciences sociales ou l'épidémiologie, elle permet donc de mieux appréhender la complexité des nombreux systèmes, naturels ou artificiels, qui ont une structure de réseau.

Par ailleurs, la prise en compte de l'intensité des interactions entre les sites constituant les réseaux complexes permet d'en donner une vision plus complète. Ces interactions présentent elles-mêmes des comportements statistiques complexes, avec émergence de lois de puissance et de corrélations non-triviales avec la topologie. Ces corrélations

peuvent être quantifiées par la définition de quantités adaptées comme le clustering pondéré. L'analyse de réseaux réels montre la nécessité, pour une modélisation adéquate, de la prise en compte des poids et du couplage entre topologie et poids.

A partir de la caractérisation empirique de réseaux réels, il a été possible de proposer un mécanisme simple de modélisation de réseaux pondérés, en combinant croissance du réseau, attachement préférentiel et renforcement des liens préexistants lors de l'adjonction de nouveaux sites et liens [8]. Les modèles basés sur ce mécanisme reproduisent un certain nombre de caractéristiques des réseaux réels et peuvent servir de base pour des modélisations plus fines et détaillées, en incluant des règles éventuellement plus compliquées. De plus, ces modèles peuvent être utilisés pour la compréhension de l'effet des hétérogénéités des interactions sur des phénomènes de résistance aux pannes, ou sur la propagation d'épidémies sur de tels réseaux (un exemple frappant étant la propagation de l'épidémie de SRAS qui s'est faite le long du réseau de lignes aériennes).

Références

- [1] P. ERDÖS & P. RÉNYI, «On random graphs I», Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).
- [2] A.-L. BARABÁSI & R. ALBERT, Science **286**, 509 (1999).
- [3] R. ALBERT & A.-L. BARABÁSI, Rev. Mod. Phys. **74**, 47 (2002).
- [4] S.N. DOROGOVTSSEV & J.F.F. MENDES, *Evolution of Networks : From biological Nets to the Internet and WWW*, Oxford University Press, Oxford (2003).
- [5] R. PASTOR-SATORRAS & A. VESPIGNANI, *Evolution and Structure of the Internet : A statistical Physics Approach*, Cambridge University Press, Cambridge, 2004.
- [6] R. GUIMERÀ , S. MOSSA, A. TURTSCHI & L.A.N. AMARAL, Proc. Natl. Acad. Sci. USA **102**, 7794 (2005); R. GUIMERÀ & L.A.N. AMARAL, Eur. Phys. J. **38** 381 (2004).
- [7] A. BARRAT, M. BARTHÉLEMY R. PASTOR-SATORRAS & A. VESPIGNANI, Proc. Natl. Acad. Sci. USA **101**, 3747 (2004).
- [8] A. BARRAT, M. BARTHÉLEMY & A. VESPIGNANI, Phys. Rev. Lett. **92** (2004) 228701.