

## CARACTÉRISATION ET MODÉLISATION DES RÉSEAUX VALUÉS

MARC  
BARTHÉLEMY

DÉPARTEMENT DE PHYSIQUE  
THÉORIQUE ET APPLIQUÉE  
CEA-DIF

ALAIN  
BARRAT

SCHOOL OF INFORMATICS  
CENTER FOR BIOCOMPLEXITY  
& DEPARTMENT OF PHYSICS  
INDIANA UNIVERSITY

ALESSANDRO  
VESPIGNANI

LABORATOIRE DE  
PHYSIQUE THÉORIQUE  
UMR 8627  
UNIV. DE PARIS-SUD

### 1. INTRODUCTION: RÉSEAUX, POIDS ET ESPACE

Des études empiriques sur des systèmes appartenant à des domaines aussi divers que les sciences sociales, la biologie et l'informatique ont montré que le paradigme habituel du réseau aléatoire (défini la première fois par ERDŐS et RÉNYI, 1959) n'est pas toujours bien adapté à la description de réseaux réels (AMARAL et al., 2000 ; ALBERT, BARABÁSI, 2002 ; DOROGOVTSSEV, MENDES, 2002 ; PASTOR-SATORRAS, VESPIGNANI, 2004). En particulier, un grand nombre de réseaux réels a une probabilité non négligeable d'avoir des nœuds de degré (nombre de liens vers d'autres nœuds du réseau) très grand. La présence de ces « *hubs* » s'accompagne souvent de larges fluctuations dans les degrés. Cette forte hétérogénéité topologique est en général exprimée par une distri-

bution de probabilité des degrés qui décroît très lentement quand le degré augmente. Ces distributions dont la variance est grande (et peut diverger avec le nombre de nœuds) entraînent des conséquences drastiques sur les propriétés du réseau comme la résilience, la vulnérabilité ou la propagation de maladies infectieuses (AMARAL et al., 2000 ; COHEN et al., 2000 ; CALLAWAY et al., 2000 ; ALBERT et al., 2000). Il convient de préciser ici que pour les réseaux dont les contraintes physiques sont trop fortes (les réseaux routier et ferroviaire par exemple), on n'observe pas de « *hubs* » (AMARAL et al., 2000).

De plus, la description purement topologique des réseaux omet une composante importante que l'on rencontre généralement dans les réseaux réels. En effet, les réseaux réels sont loin d'être une structure booléenne et sont mieux décrits par des réseaux valués pour lesquels l'intensité des liens peut varier sur plusieurs ordres de grandeur, comme cela a été observé sur plusieurs graphes allant des réseaux trophiques aux réseaux métaboliques (PASTOR-SATORRAS, VESPIGNANI, 2001 ; GRANOVETTER, 1973 ; GARLASCHELLI et al., 2005 ; LI, CAI, 2004 ; KRAUSE et al., 2003 ; BARRAT et al., 2004b ; ALMAAS et al., 2004). Il est par ailleurs intéressant de noter que même si un réseau ne présente pas de *hubs* et a donc une topologie « régulière », le trafic peut être quand même très fortement hétérogène (DE MONTIS et al., 2007).

Par ailleurs, les propriétés statistiques des poids indiquent une corrélation non-triviale avec des quantités topologiques (BARRAT et al., 2004b). Finalement, la corrélation entre les poids des différents liens est à l'origine de l'existence de chemins préférentiels ou « *pathways* » qui ont un rôle particulièrement important en épidémiologie ou dans les réseaux métaboliques (ALMAAS et al., 2004).

Dans cet article, nous montrons comment une approche issue de la physique statistique des réseaux permet de jeter quelques lumières sur le réseau de transport des lignes aériennes. Nous commencerons par introduire les outils récemment développés dans le cadre des réseaux complexes et qui permettent de caractériser un certain nombre de propriétés statistiques importantes des grands réseaux. Ces quantités sont très générales et peuvent s'appliquer à tous les réseaux. En particulier, nous nous attacherons à décrire l'interaction entre la topologie et le trafic. Ensuite, dans une démarche typique de la physique statistique, nous étudions l'interaction des différents ingrédients (topologie fortement hétérogène, trafic) en présentant un modèle simplifié de croissance de réseaux combinant tous ces ingrédients. En particulier, nous montrons à l'aide de modèles qu'un ingrédient fondamental est donné par le couplage entre l'évolution de la topologie et la dynamique de réarrangement des poids (BARRAT et al., 2004a). Ce mécanisme mène naturellement à l'apparition de plusieurs caractéristiques observées dans les réseaux réels, en particulier les corrélations entre poids et topologie.

Cet article est organisé de la manière suivante. Dans la partie 2, nous passons en revue les faits empiriques marquants du réseau des lignes aériennes de

l'Amérique du Nord. La partie 3 présente un modèle de réseau valué. Dans la partie 4, nous présentons un sommaire des différents résultats et donnons quelques perspectives à propos de la modélisation des grands réseaux.

## 2. ÉTUDE D'UN CAS PARTICULIER: ESPACE, TOPOLOGIE ET TRAFIC DANS LE RÉSEAU DES LIGNES AÉRIENNES D'AMÉRIQUE DU NORD

### 2.1. CARACTÉRISATION TOPOLOGIQUE

Les caractéristiques du réseau mondial des lignes aériennes (world-wide air – transportation network, ou WAN en abrégé), déduites de la base de données IATA (*International Air Transportation Association*, <http://www.iata.org>), ont été étudiées en détail dans BARRAT et al. (2004b). Ce réseau est constitué de  $N=3\,880$  aéroports et  $E=18\,810$  liens représentant les vols directs entre ces aéroports (pour l'année 2002). Ce réseau a des propriétés de « petit-monde » avec une distance topologique moyenne d'environ 4,4 (qui correspond au nombre moyen de vols directs nécessaires pour aller d'un aéroport à un autre), chiffre très petit devant  $N$ . On observe aussi des propriétés typiques des réseaux « sans-échelle » pour lesquels la distribution de probabilité des degrés (le degré est ici le nombre de connexions d'un aéroport) décroît très lentement avec le degré -typiquement comme une loi de puissance, comme cela a été confirmé par plusieurs études utilisant différentes bases de données du réseau mondial (GARLASCHELLI et al., 2005 ; GUIMERA, AMARAL, 2004 ; GUIMERA et al., 2005). La distribution des degrés est bien décrite par une

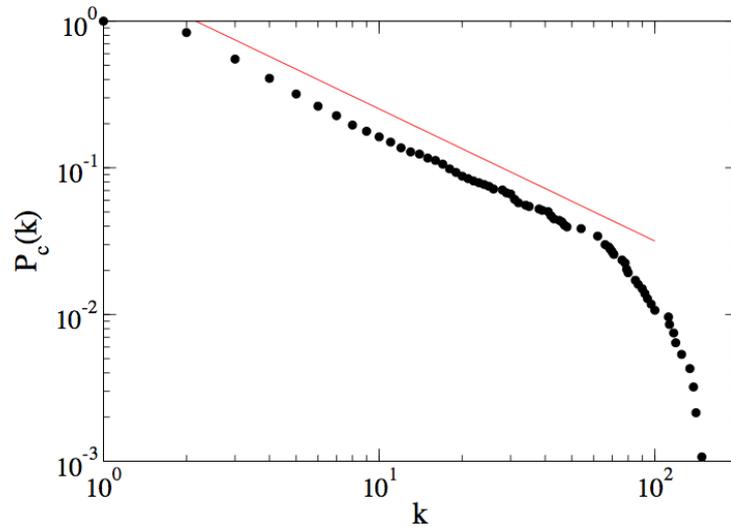
fonction de la forme  $P(k) = k^{-\gamma} f\left(\frac{k}{k_0}\right)$  où  $\gamma \approx 2$  et  $f(x)$  est une fonction de

coupure exponentielle. La distribution des degrés est donc une loi large et la coupure aux grands degrés vient des contraintes physiques, le nombre maximal de connexions qu'un aéroport peut avoir étant limité par sa taille (PASTOR-SATORRAS, VESPIGNANI, 2004 ; GUIMERA, AMARAL, 2004 ; GUIMERA et al., 2005). Le réseau des lignes aériennes est un exemple clair de réseau petit-monde ayant une distribution large de degrés. Ce résultat en particulier implique l'existence de nœuds avec des degrés très élevés, les *hubs*. Ces *hubs* jouent bien sûr un rôle privilégié et très important dans de tels réseaux (O'KELLY, MILLER, 1994). Bien que ces *hubs* et leur voisinage aient déjà été étudiés depuis longtemps par de nombreux auteurs, la nouveauté dans l'approche proposée dans ces récentes études est la caractérisation statistique des réseaux, un point de vue qui tire son origine dans la physique statistique habituée à caractériser l'émergence de phénomènes collectifs à grande échelle.

Le réseau de transport aérien, considéré au niveau mondial, combine plusieurs échelles d'espace, l'échelle globale (intercontinentale) et l'échelle nationale. Afin d'obtenir des résultats statistiques cohérents et pour ne pas mélanger des effets très différents, nous nous restreignons ici essentiellement au cas d'un seul continent. Nous allons donc considérer le cas du réseau

nord-américain constitué de 935 aéroports, avec un degré moyen de l'ordre de 8,4 et une distance topologique moyenne de l'ordre de 4. Les propriétés statistiques du réseau nord-américain sont cohérentes avec celles du réseau mondial. Ainsi la distribution des degrés est une loi de puissance sur presque deux ordres de grandeur (Figure 1) suivie par une décroissance exponentielle.

Figure 1. Distribution cumulative des degrés pour le réseau nord-américain (en échelle logarithmique double)



La ligne représente une décroissance en loi de puissance avec un exposant  $\gamma - 1 \approx 0.9$

La très faible valeur de l'exposant  $\gamma$  (autour de 2) indique que ce réseau a une distribution de degrés extrêmement hétérogène sans présenter toutefois un seul nœud dominant (qui correspondrait à  $\gamma < 2$ ). Dans la partie 3, nous proposons un modèle qui nous permet de discuter plus en détail la valeur de cet exposant.

## 2.2. PROPRIÉTÉS DU TRAFIC

Le réseau de lignes aériennes est un exemple emblématique de grands réseaux qui peuvent être représentés par un réseau valué. Puisque les liens représentent les vols directs entre les aéroports (nœuds du réseau), le poids sur chaque lien est naturellement défini comme la capacité en termes de nombre maximal de passagers. Plus précisément, IATA fournit le nombre de sièges sur une ligne, qui est donc proportionnel au nombre d'avions (en négligeant les variations du nombre de sièges disponibles d'un avion à un autre, ce qui est raisonnable pour les avions de ligne dont la capacité ne varie pas sur plusieurs ordres de grandeur). D'autre part, le taux de remplissage moyen est de 70 %, ce qui implique que le nombre de sièges est représentatif du nombre de passagers.

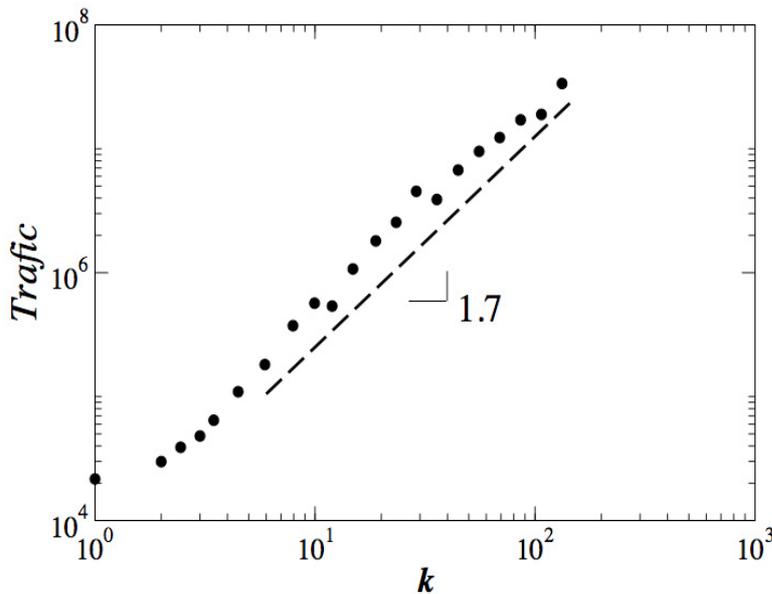
Plus généralement, les propriétés d'un graphe sont décrites par la matrice

d'adjacence  $a_{ij}$  qui prend la valeur 1 si  $i$  et  $j$  sont reliés et vaut zéro dans le cas contraire. Les réseaux valués ou pesés sont décrits par une matrice  $w_{ij}$  qui donne le poids du lien connectant  $i$  et  $j$  ( $w_{ij} = 0$  si les nœuds  $i$  et  $j$  ne sont pas connectés). La généralisation naturelle du degré d'un nœud est le trafic des aéroports (dans le cadre plus général de la théorie des réseaux, cette quantité est appelée « *strength* » en anglais) d'un nœud  $s_i^w$  défini par (BARRAT et al., 2004b)

$$s_i^w = \sum_{j \in V(i)} w_{ij}$$

où la somme porte sur l'ensemble  $V(i)$  des voisins de  $i$ . Le trafic d'un nœud intègre l'information relative à sa connectivité et à l'importance du trafic qui y arrive et qui en part (ces deux quantités étant approximativement égales comme le montrent les mesures empiriques, BARRAT et al., 2004b). Quand les poids sont indépendants d'un lien à un autre, on obtient  $s^w \approx \langle w \rangle k$  où  $\langle w \rangle$  est le poids moyen. En présence de corrélations entre les poids et la topologie, on observe en général un autre comportement de la forme  $s^w \approx Ak^\beta$  avec  $\beta \neq 1$  ou bien  $\beta = 1$  et  $A \neq \langle w \rangle$ . En particulier on observe pour le WAN une loi de puissance avec un exposant « anormal »  $\beta \approx 1.5$ . Pour le réseau nord-américain, on obtient une valeur légèrement différente de l'ordre de 1,7 (Figure 2).

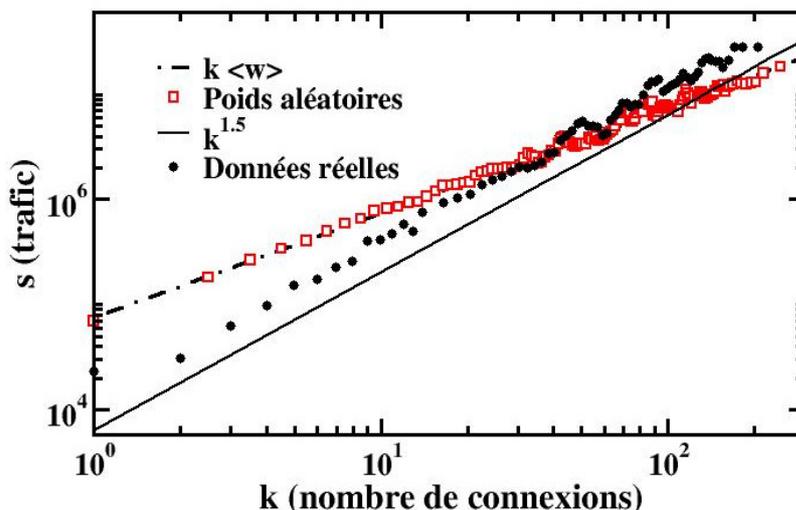
Figure 2. Trafic des aéroports versus le degré pour le réseau nord-américain



La ligne en pointillés correspond à une loi de puissance avec un exposant d'une valeur 1,7.

Une valeur de cet exposant  $\beta$  plus grande que 1 implique que le nombre de passagers par connexion n'est pas constant mais croît avec le nombre de connexions. Autrement dit, les aéroports avec plus de connexions ont aussi un flux de passagers plus grand sur ces connexions. On a donc ici une forte corrélation entre la topologie du réseau et le trafic. En effet, en distribuant au hasard le trafic sur les liens d'un réseau fixé, nous obtiendrions une valeur  $\beta=1$  avec un préfacteur égal à  $\langle w \rangle$  (Figure 3). Dans ce cas en effet, même si le trafic d'un aéroport varie proportionnellement avec le degré, les processus de construction du réseau et d'assignation du trafic sont indépendants statistiquement.

Figure 3. Trafic  $s^w$  versus le degré pour le réseau global WAN (en Log-Log)



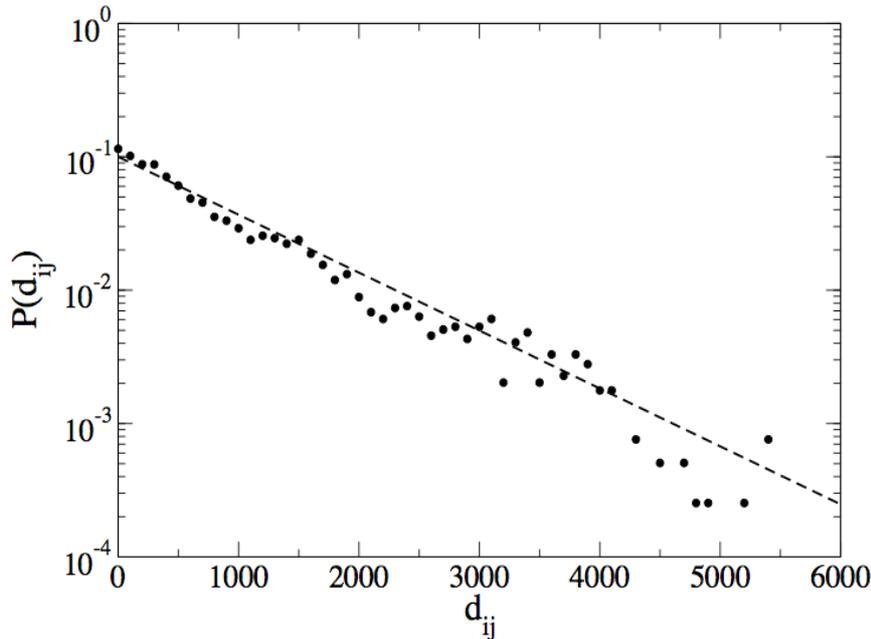
Les points représentent les données réelles, la ligne continue une loi de puissance avec un exposant égal à 1,5. Quand on redistribue les poids aléatoirement (symboles carrés), on obtient bien un comportement linéaire (ligne en pointillés).

### 2.3. ANALYSE SPATIALE DU RÉSEAU NORD AMÉRICAIN

Les attributs spatiaux du réseau Nord Américain sont contenus dans la distribution de la distance des connexions. En effet, cette distribution intègre la répartition des différentes villes et les choix fixés par des contraintes économiques. En particulier depuis la dérégulation de 1978, l'optimisation des coûts a imposé la présence de *hubs* qui sont situés près de villes importantes. Le réseau des lignes aériennes résulte donc de processus complexes mêlant géographie et économie. La Figure 4 représente la distribution de la distance parcourue (en kms) sur les vols directs. Cette Figure montre clairement une décroissance exponentielle signalant l'existence d'une échelle de longueur typique de l'ordre de 1 000 kms. L'origine de cette échelle

typique découle probablement de considérations économiques des compagnies aériennes; il est apparemment plus économique d'effectuer plusieurs segments de longueurs moindres, en passant par des *hubs* du réseau.

Figure 4. Distribution de probabilité des distances (en kms) entre les aéroports reliés par un vol direct pour le réseau nord-américain



La ligne en pointillé indique un fit par une exponentielle et qui donne une échelle de longueur de l'ordre de l'ordre de 1 000 kms.

Puisque l'espace est un paramètre important dans ce réseau, une autre quantité importante est la «portée totale» de l'aéroport définie pour chaque aéroport  $i$  par

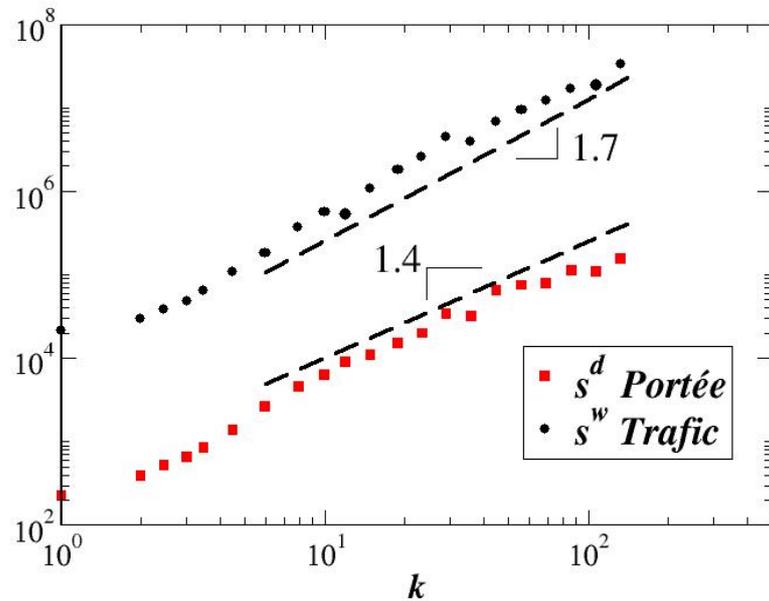
$$s_i^d = \sum_{j \in V(i)} d_{ij}$$

où  $d_{ij}$  est la distance euclidienne entre  $i$  et  $j$ . Cette portée donne les distances cumulées de toutes les connexions directes en partance de l'aéroport  $i$ . De manière analogue au trafic défini plus haut, on peut quantifier les corrélations entre les distances des vols et la topologie du réseau par la valeur de l'exposant décrivant le comportement de  $s^d(k)$  :  $s^d(k) \approx k^{\beta_d}$ .

Un comportement linéaire ( $\beta_d = 1$ ) indique l'absence de telles corrélations. Comme nous le montrons sur la Figure 5, nous obtenons pour valeur  $\beta_d \approx 1.4$ . Ce résultat démontre l'importance des corrélations entre la topologie et la géographie. En effet, une valeur de cet exposant plus grande que 1 montre que les aéroports qui ont un grand nombre de connexions ont (en plus

d'un plus grand nombre de passagers par connexion) une distance par connexion plus grande. Comme on pouvait s'y attendre intuitivement, le réseau aérien est donc un réseau très hétérogène où les *hubs* ont simultanément un plus grand degré, un plus grand trafic par connexion et une plus grande distance par connexion, toutes ces quantités étant reliées par des lois super-linéaires (BARRAT et al., 2004b ; BARRAT et al., 2005).

Figure 5. Trafic et portée totale versus le degré pour le réseau nord-américain



Les lignes en pointillés correspondent aux lois de puissance avec des exposants 1,4 et 1,7.

#### 2.4. DENSITÉ DE VOISINAGE ET CORRÉLATIONS

Les corrélations, dans un graphe, sont habituellement mesurées par deux quantités. D'une part, le coefficient de clustering topologique (WATTS, STROGATZ, 1998) mesure la connectivité locale, donnée par la fraction des voisins d'un nœud qui sont reliés entre eux,

$$C(i) = \frac{2E_i}{k_i(k_i + 1)},$$

où  $E_i$  désigne le nombre de liens entre les voisins de  $i$ . D'autre part, le degré moyen des voisins d'un site permet de savoir si ce site a tendance à avoir des liens avec des *hubs* ou avec des sites peu connectés. Ces mesures ne prennent cependant pas en compte le fait que certains voisins sont plus importants que d'autres.

Pour les réseaux valués, il faut donc introduire une mesure du clustering qui

combine l'information topologique et celle relative au trafic (plusieurs définitions sont possibles, voir par exemple AMIEL et al. 2005). Le coefficient de clustering valué est alors défini par BARRAT et al. (2004b)

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

Cette quantité compte le nombre de triplets formés par  $i$  et deux de ses voisins en pondérant chaque triplet avec les poids des liens vers ces voisins. La normalisation assure que ce coefficient est dans l'intervalle  $[0,1]$  et qu'il se réduit à sa contrepartie topologique  $C(i)$  lorsque les poids sont constants. On peut moyennner  $C(i)$  et  $C^w(i)$  sur tous les nœuds  $i$  de degré égal à  $k$  et l'on obtient les quantités notées  $C(k)$  et  $C^w(k)$ ,

$$C(k) = \frac{1}{N(k)} \sum_{i/k_i=k} C(i),$$

$$C^w(k) = \frac{1}{N(k)} \sum_{i/k_i=k} C^w(i)$$

où  $N(k)$  est le nombre de nœuds de degré  $k$ . Le rapport  $C^w(k)/C(k)$  permet une analyse en fonction du degré  $k$ , et d'identifier les triplets interconnectés qui sont formés par des liens possédant un fort trafic.

De façon similaire, on définit la moyenne des degrés des voisins, pondérée par le trafic sur les liens,

$$k^w_{nn,i} = \frac{1}{s_i} \sum_{j=1}^N a_{ij} w_{ij} k_j.$$

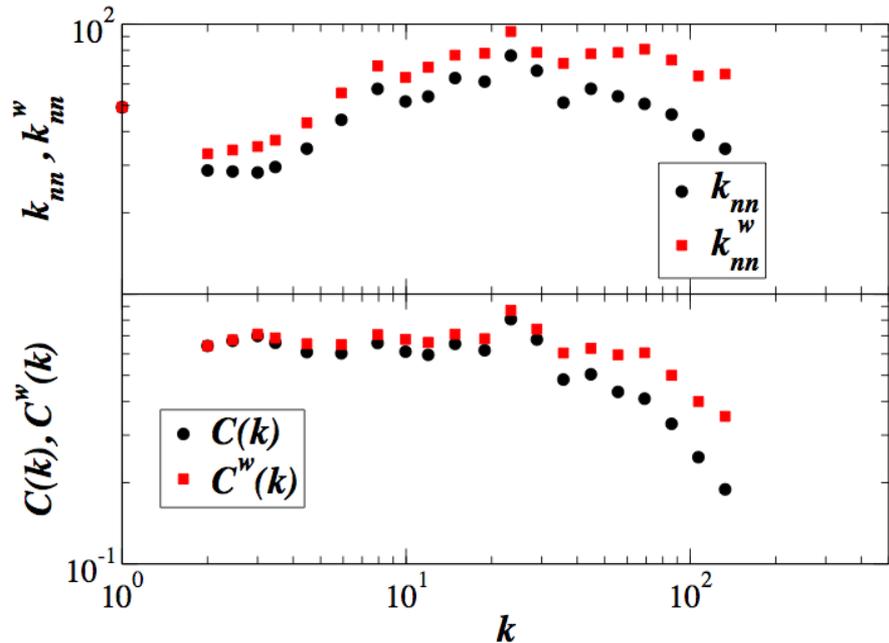
Cette définition implique que  $k^w_{nn,i}$  est supérieure à sa contrepartie topologique (obtenue pour  $w_{ij} = \text{const.}$ ) lorsque les poids les plus grands pointent vers les voisins de plus grand degré. La quantité  $k^w_{nn,i}$  mesure donc l'affinité effective de se connecter avec des voisins de fort ou faible degré selon l'ordre de grandeur du trafic sur les liens correspondants. En moyennant sur les nœuds de même degré  $k$  on obtient la quantité  $k^w_{nn}(k)$  qui permet de distinguer un comportement assortatif pour lequel les nœuds de degrés semblables sont fortement connectés d'un comportement disassortatif pour lequel des nœuds de degrés différents ont tendance à être connectés (NEWMAN, 2002).

Nous montrons sur la Figure 6 le comportement de ces différentes quantités pour le réseau aérien nord-américain. Nous observons une valeur essentiellement constante pour  $k^w_{nn}(k)$  et un comportement légèrement disassortatif pour les grands degrés  $k$ . Ceci est probablement dû au fait que de nombreux aéroports importants ont un grand nombre de connexions intercontinentales qui ne sont pas prises en compte ici. Dans le cas du WAN,

on obtient bien en effet un comportement légèrement assortatif pour lequel les *hubs* ont tendance à échanger un gros volume de trafic avec les autres *hubs* (BARRAT et al., 2004b).

Le *clustering* est très élevé et décroît pour les grands degrés  $k$ . Ce comportement est souvent observé pour les réseaux complexes réels et est une conséquence directe du rôle des *hubs* qui sont reliés entre eux et qui redistribuent le trafic vers les plus petites échelles régionales. Ces *hubs* fortement connectés forment donc des cliques échangeant un fort trafic, un phénomène que l'on retrouve en sociologie sous le nom de « *Rich-club phenomenon* » (ZHOU, MONDEAGON, 2003).

Figure 6. Assortativité et clustering pour le réseau nord-américain



Les cercles correspondent aux quantités topologiques et les carrés correspondent aux mêmes quantités mais valuées.

Notons finalement que, pour un nœud  $i$  donné, de degré  $k_i$  et de trafic  $s_i$ , on peut faire face à des situations très différentes. Ainsi, les poids  $w_{ij}$  peuvent être du même ordre de grandeur ou bien au contraire un des poids peut dominer sur tous les autres. Un moyen simple de mesurer cette disparité est donné par la quantité  $Y_2$  introduite dans un autre contexte (DERRIDA, FLYVBERG, 1987 ; BARTHÉLEMY et al., 2003),

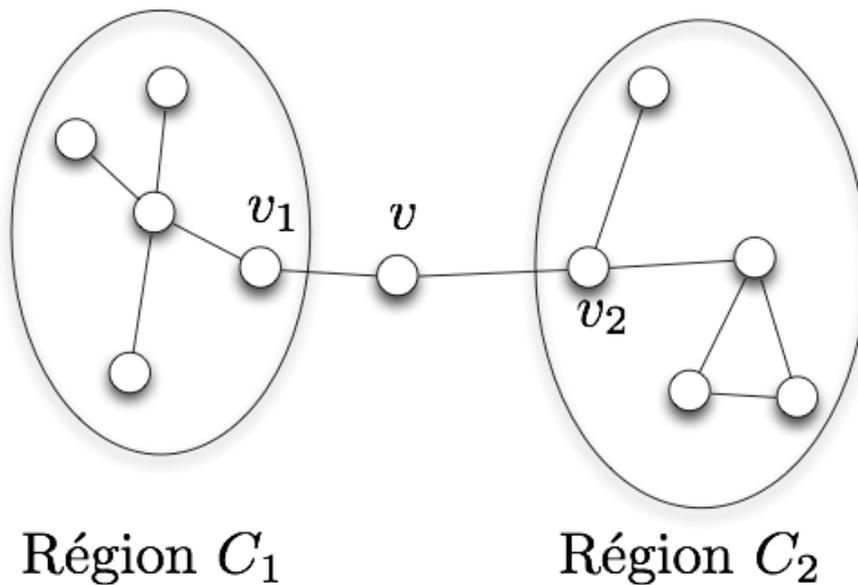
$$Y_2(i) = \sum_{j \in V(i)} \left[ \frac{w_{ij}}{s_i} \right]^2 .$$

Si tous les poids sont du même ordre de grandeur, alors  $Y_2 \approx \frac{1}{k_i}$  (pour  $k_i \gg 1$ ) et si au contraire un petit nombre de poids dominant alors  $Y_2 \approx 1/n$  (avec  $n$  de l'ordre de l'unité). Cette quantité a été récemment utilisée pour caractériser les réseaux métaboliques (ALMAAS et al., 2004) et a permis d'identifier les réactions dominantes. Pour le WAN et le réseau nord-américain, nous observons des comportements similaires pour lesquels la moyenne par degré varie comme  $Y_2 \approx 1/k$  ce qui signifie que pour les grands aéroports, toutes les connexions ont le même ordre de grandeur (même s'il y a évidemment des différences, parfois notables).

### 2.5. CENTRALITÉ (« BETWEENNESS CENTRALITY »)

Les quantités et corrélations discutées précédemment correspondent à une échelle locale sur le réseau (deux ou trois sites voisins). Une caractérisation plus complète est fournie par des quantités qui prennent en compte la topologie globale du réseau. Ainsi, le degré d'un nœud est bien entendu une première indication de la centralité d'un nœud. Cependant une approche plus globale est nécessaire car un nœud même de degré faible peut être très important dans la structure globale, comme montré en Figure 7 par un exemple.

Figure 7 Le nœud  $v$  a un très faible degré ( $k=2$ ) mais une très grande centralité (BC) car tous les chemins allant d'une région à une autre passent nécessairement par ce nœud



Une quantité qui prend en compte l'importance d'un nœud donné dans la

structure globale du réseau est la « *betweenness centrality* » (BC) (FREEMAN, 1977 ; GOH et al., 2001 ; BARTHÉLEMY, 2003). Cette centralité est définie pour chaque nœud  $v$  par

$$g(v) = \sum_{s \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

où  $\sigma_{st}$  est le nombre de plus courts chemins allant de  $s$  à  $t$  et  $\sigma_{st}(v)$  le nombre de plus courts chemins allant de  $s$  à  $t$  en passant par  $v$ . Cette définition signifie que les nœuds centraux participent à beaucoup plus de chemins que ne le font les nœuds à la périphérie du réseau. De plus la BC donne une estimation du trafic pour les réseaux de transports en supposant que le nombre de plus courts chemins est une approximation de la fréquence d'utilisation d'un nœud.

Il est utile en général de définir la BC moyennée sur tous les nœuds de même degré,

$$g(k) = \frac{1}{N(k)} \sum_{v/k_v=k} g(v),$$

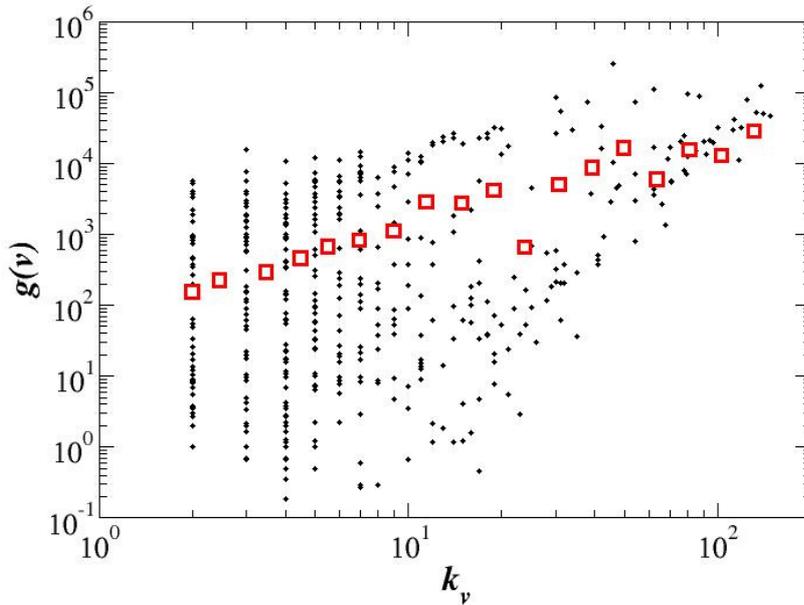
où  $N(k)$  est le nombre de sites de degré  $k$ . En général, plus le degré est élevé et plus la centralité est grande. Pour les réseaux sans-échelle, il a été montré que la centralité varie avec  $k$  comme

$$g(k) \approx k^\mu$$

où  $\mu$  dépend du réseau (BARTHÉLEMY, 2003). Pour certains réseaux, les fluctuations de la BC autour du comportement moyen donné par l'équation précédente peuvent être larges et des « anomalies » peuvent apparaître dans le sens où des nœuds de degré plus grand (resp. petit) peuvent avoir une centralité beaucoup plus petite (resp. grande) que la moyenne.

GUIMERA et AMARAL (2004) ont montré que ce phénomène apparaît en effet dans le réseau aérien. Cette remarque est très importante car des aéroports très centraux peuvent avoir un degré relativement faible et vice-versa. Nous reportons sur la Figure 8, le comportement moyen ainsi que tous les points  $(k_i, g(i))$  pour le réseau nord-américain: les larges fluctuations autour de la moyenne sont manifestes. Les modèles décrits dans GUIMERA et AMARAL (2003) et BARRAT et al. (2005) suggèrent que les fluctuations de la centralité par rapport au degré sont dues aux contraintes spatiales. Dans la suite, par manque de place, nous négligerons cet aspect et nous nous concentrerons sur le couplage topologie-traffic. Nous décrivons donc ici uniquement le modèle «non-spatial» et renvoyons le lecteur intéressé à l'inclusion des effets de contraintes spatiales dans le modèle présenté dans BARRAT et al. (2005).

Figure 8. Centralité (BC) versus le degré pour les aéroports du réseau nord-américain



Chaque point représente un aéroport, les carrés représentent la valeur moyennée (pour un degré donné).

### 3. MODÉLISATION DES RÉSEAUX VALUÉS

#### 3.1. LA NÉCESSITÉ D'UN COUPLAGE ENTRE LES POIDS ET LA TOPOLOGIE

Les premiers modèles de réseaux valués considéraient des évolutions indépendantes pour la topologie et pour les poids (YOOK et al., 2001) (ou bien des poids évoluant en fonction de la topologie, ZHENG et al., 2003). Ces modèles négligent donc le fait que les évolutions du trafic et de la topologie sont couplées. Nous pouvons illustrer ce point sur le réseau des lignes aériennes. Si une nouvelle connexion entre deux aéroports est créée, elle va généralement perturber le trafic existant dans les deux aéroports. En général, l'activité va augmenter et dépendra de la nature spécifique du réseau et de la dynamique locale. Dans la suite de ce paragraphe, nous allons décrire un modèle simplifié qui prend en compte ce couplage, en faisant évoluer les poids de façon corrélée à la topologie. Ce mécanisme très simple permet de mimer l'évolution et le renforcement des interactions dans les réseaux d'infrastructure. Il s'agit ici, bien entendu, de mettre en lumière les mécanismes principaux, mais ce modèle ne prétend en aucun cas décrire la formation historique du réseau des lignes aériennes. Notre démonstration porte sur le fait qu'un modèle réaliste devrait incorporer d'une façon ou d'une autre ces mécanismes élémentaires.

### 3.2. LE MODÈLE

Le modèle (BARRAT et al., 2004a) considère  $N_0$  nœuds initialement connectés par des liens avec un poids égal à  $w_0$ . A chaque pas de temps, un nouveau nœud est ajouté au système et se connecte à  $m$  différents nœuds existants. Chaque nouveau nœud introduit donc  $m$  nouveaux liens, avec chacun un poids initial  $w_0$ . Le nouveau nœud  $n$  se connecte à un nœud existant  $i$  avec une probabilité donnée par

$$\Pi_{n \rightarrow i} = \frac{S_i^w}{\sum_j S_j^w}.$$

Autrement dit, plus le trafic total d'un aéroport est grand et plus la probabilité qu'un nouvel aéroport s'y connecte est grande elle aussi. Cette règle de « *busy get busier* » semble raisonnable dans le cas du réseau aérien mais aussi dans le cas de nombreux réseaux tels que Internet où un nouveau routeur va préférentiellement se connecter à un routeur de grande capacité.

La présence d'une nouvelle connexion ( $n,i$ ) va perturber les poids existants dans le réseau. En particulier, nous considérons des réarrangements locaux entre le nœud  $i$  et ses voisins  $j$  décrits par la forme simple

$$w_{ij} \rightarrow w_{ij} + \Delta w_{ij}$$

$$\text{où } \Delta w_{ij} = \delta \frac{w_{ij}}{S_i^w}.$$

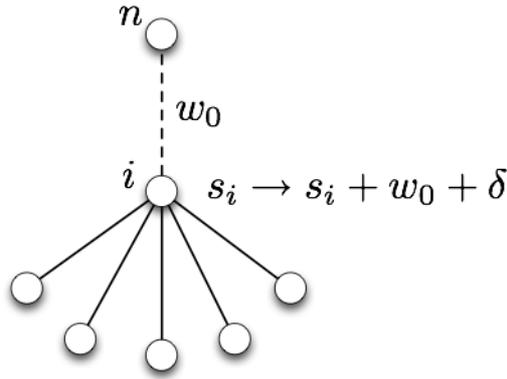
Cette équation exprime le fait qu'une nouvelle connexion avec le nœud  $i$  induit une augmentation totale de trafic égale à  $\delta$  qui est répartie proportionnellement sur les différentes connexions partant de  $i$  vers ses voisins (Figure 9). Le modèle le plus simple est obtenu pour  $\delta = \text{const.}$ , mais il est possible de considérer différents choix pour  $\Delta w_{ij}$  qui traduisent différentes dynamiques locales de réarrangement (ZHENG et al., 2003 ; PANDYA, 2004 ; WANG et al., 2005 ; BARRAT et al., 2004c).

Finalement, une fois que les poids ont été modifiés, le processus de croissance du réseau est itéré en introduisant un nouveau nœud jusqu'à obtenir un réseau avec le nombre de nœuds désiré.

Ce modèle ne dépend donc que du paramètre  $\delta$  qui est la fraction du poids induite par le nouveau lien sur les liens préexistants. Selon la valeur de  $\delta$  plusieurs scénarios sont possibles. Si  $\delta \approx 1$ , le modèle décrit une situation où le trafic du nouveau lien est distribué dans les liens existants de manière conservative. Ce scénario est plausible pour les aéroports où le trafic de transit est important pour les *hubs*. Dans le cas où  $\delta < 1$ , une nouvelle connexion crée un faible surplus d'activité sur les connexions existantes. Enfin, si  $\delta > 1$ , une nouvelle connexion engendre une sorte d'effet multi-

plicatif qui démultiplie le trafic sur les voisins.

Figure 9 : Illustration de la règle de construction



Un nouveau nœud  $n$  se connecte au nœud  $i$  avec une probabilité proportionnelle à  $s_i^w / \sum_l s_l^w$ . Le poids du nouveau lien est donné par  $w_0$  et le trafic total sur les liens pré-existants est augmenté d'un montant total égal à  $\delta$ .

L'évolution du réseau peut être décrite analytiquement en écrivant les équations d'évolution des quantités  $s_i^w(N)$  et du nœud  $i$  en fonction du nombre de sites  $N$  dans le réseau (ALBERT, BARABÁSI, 2002 ; BARRAT et al., 2004a). Dans la limite des grands  $N$ , on obtient alors

$$s_i^w(N) \approx (2\delta + 1)k_i(N).$$

Cette relation de proportionnalité implique la valeur  $\beta = 1$ , mais le préfacteur est différent de  $\langle w \rangle$  ce qui indique l'existence de corrélations entre la topologie et les poids. On peut montrer que l'inclusion de l'espace modifie cette relation ce qui suggère qu'un modèle réaliste du réseau des lignes aériennes (décrit par  $\beta \approx 1.5$ ) doit nécessairement tenir compte des effets spatiaux.

Les distributions de différentes quantités peuvent aussi être calculées et l'on obtient dans la limite des grandes tailles des lois de puissance pour le degré  $k$  et de trafic  $s$  avec le même exposant donné par

$$\gamma = \frac{4\delta + 3}{2\delta + 1}.$$

Ce résultat montre que le graphe obtenu est un réseau sans-échelle décrit par un exposant  $\gamma \in [2,3]$  dont la valeur dépend de la dynamique de réarrangement par l'intermédiaire du paramètre  $\delta$ . En particulier, lorsque l'addition de nouveaux liens n'affecte pas le trafic existant ( $\delta = 0$ ), le modèle est équivalent à celui de BARABÁSI et ALBERT (1999) et l'on retrouve bien la valeur  $\gamma = 3$  pour ce modèle. Pour des valeurs plus grandes de  $\delta$ , l'exposant  $\gamma$  se

rapproche de la valeur 2. Ce dernier fait suggère que les réseaux dont la valeur de  $\gamma$  est proche de 2 (ce qui est le cas du réseau des lignes aériennes nord-américaines) correspondent à des réseaux très « réactifs » pour lesquels l'ajout d'une nouvelle connexion entraîne une très forte perturbation et une augmentation du trafic.

Ce résultat indique que la croissance d'un réseau pilotée par le trafic peut générer des réseaux sans-échelle dont l'exposant varie dans le domaine [2,3] généralement observé dans les réseaux réels. Ce résultat montre aussi que les exposants ne sont pas universels et dépendent de la dynamique microscopique, ce qui pourrait expliquer la grande variété d'exposants observée dans les réseaux réels.

#### 4. CONCLUSIONS ET PERSPECTIVES

Nous avons présenté ici les résultats obtenus à l'aide des outils développés dans le cadre de la physique statistique des grands réseaux complexes. Nous avons défini certains indicateurs importants qui caractérisent les propriétés statistiques des grands réseaux. Enfin, nous avons présenté un modèle simple de formation de réseaux valués qui possède l'avantage de mettre en évidence l'importance de certains ingrédients tel que le couplage entre la topologie et le trafic.

Plus précisément, nous avons montré qu'une vision plus complète des réseaux incorpore, en plus de la topologie, l'étude des interactions définissant les liens de ces systèmes. Les poids caractérisant les connexions peuvent exhiber des caractéristiques complexes avec des distributions de probabilité très larges. En particulier, nous avons illustré ce phénomène avec l'exemple du réseau des lignes aériennes de l'Amérique du Nord. Nous avons montré l'importance des corrélations entre la topologie et le trafic dans la caractérisation de ces réseaux. En effet, le trafic et la topologie offrent des perspectives complémentaires dans la description de la structure de ces réseaux. Ces corrélations restent indétectables si l'on se restreint à une étude de quantités purement topologiques. Les quantités valuées offrent donc une approche quantitative générale afin de comprendre l'architecture complexe des réseaux réels.

Les résultats empiriques démontrent que des modèles purement topologiques sont insuffisants et qu'il faut aller au-delà. Nous avons présenté ici un modèle -peut-être le plus simple possible dans la classe des modèles de croissance de réseaux pilotée par le trafic. L'ingrédient essentiel de ce modèle est la possibilité qu'a le trafic d'évoluer lorsque la topologie varie et de nouvelles connexions sont créées. Ce mécanisme permet de produire une grande variété de réseaux valués possédant des caractéristiques observées dans les réseaux réels. De nombreuses variations de ce modèle sont possibles comme l'inclusion de l'espace ou de dynamiques locales non-linéaires (en particulier,

l'espace joue un rôle très important dans ces réseaux). Dans cette perspective, ce modèle apparaît comme un bon point de départ pour une modélisation réaliste des réseaux valués.

## REMERCIEMENTS

Nous remercions IATA pour nous avoir procuré la base de données du transport aérien mondial.

## RÉFÉRENCES

- ALBERT R., JEONG H., BARABÁSI A.-L. (2000) Error and attack tolerance of complex networks. **Nature**, n° 406, pp. 378-382.
- ALBERT R., BARABÁSI A.-L. (2002) Statistical mechanics of complex networks. **Rev. Mod. Phys.** Vol. 74, pp. 47-97.
- ALMAAS E., KOVACS B., VISCEK T., OLTVAI Z.N., BARABÁSI A.-L. (2004) Global organization of metabolic fluxes in the bacterium *E. Coli*. **Nature**, Vol. 427, pp. 839-843.
- AMARAL L.A.N., SCALA A., BARTHÉLEMY M., STANLEY H.E. (2000) Classes of small-world networks. **Proc. Natl. Acad. Sci. (USA)**, 97, pp. 11149-11152.
- AMIEL M., MÉLANÇON G., ROZENBLAT C. (2005) Réseaux multi-niveaux : l'exemple des échanges aériens mondiaux. **M@ppemonde**, n° 79 (<http://mappemonde.mgm.fr/num7/articles/art05302.html>).
- BARABÁSI A.-L., ALBERT R. (1999) Emergence of scaling in random networks. **Science**, Vol. 286, pp. 509-512.
- BARRAT A., BARTHÉLEMY M., VESPIGNANI A. (2004a) Weighted evolving networks: coupling topology and weight dynamics. **Phys. Rev. Lett.**, Vol. 92, pp. 228701.1-228701.4.
- BARRAT A., BARTHÉLEMY M., PASTOR-SATORRAS R., VESPIGNANI A. (2004b) The architecture of complex weighted networks. **Proc. Natl. Acad. Sci. (USA)**, Vol. 101, pp. 3747-3752.
- BARRAT A., BARTHÉLEMY M., VESPIGNANI A. (2004c) Modeling the evolution of weighted networks. **Phys. Rev. E**, Vol. 70, pp. 066149.1-066149.12.
- BARRAT A., BARTHÉLEMY M., VESPIGNANI A. (2005) The effects of spatial constraints on the evolution of weighted complex networks. **J. Stat. Mech.**, P05003.
- BARTHÉLEMY M. (2003) Comment on 'Universal behavior of load distribution in scale-free networks'. **Phys. Rev. Lett.**, 91, 189803.
- BARTHÉLEMY M. (2004) Betweenness centrality in large complex networks. **Eur. Phys. J. B**, Vol. 38, pp. 163-168.

BARTHÉLEMY M., GONDRAN B., GUICHARD E. (2003) Spatial structure of the Internet traffic. **Physica A**, Vol. 319, pp. 633-642.

CALLAWAY D.S., NEWMAN M.E.J., STROGATZ S.H., WATTS D.J. (2000) Network robustness and fragility: percolation on random graphs. **Phys. Rev. Lett.**, Vol. 85, pp. 5468-5471.

COHEN R., EREZ K., BEN AVRAHAM D., HAVLIN S. (2000) Resilience of the Internet to random breakdowns. **Phys. Rev. Lett.**, Vol. 85, pp. 4626-4628.

DE MONTIS A., BARTHÉLEMY M., CHESSA A., VESPIGNANI A. (2007) The structure of inter-urban traffic: a weighted network analysis. **Environment and Planning B: Planning and Design**, Vol. 34, pp. 905-924.

DERRIDA B., FLYVBJERG, J. (1987) Statistical properties of randomly broken objects and of multivalley structures in disordered systems. **J. Phys. A**, Vol. 20, pp. 5273-5288.

DOROGOVTSSEV S.N., MENDES J.F.F. (2002) Evolution of networks. **Adv. Phys.**, Vol. 51, pp. 1079-1187.

ERDŐS P., RÉNYI A. (1959) On Random Graphs. **Publ. Math. Debrecen** 6, pp. 290-297.

FREEMAN L.C. (1977) A set of measures of centrality based on betweenness. **Sociometry**, Vol. 40, pp. 35-41.

GARLASCHELLI D., BATTISTON S., CASTRI M., SERVEDIO V.D.P., CALDARELLI G. (2005) The scale-free topology of market investments. **Physica A**, Vol. 350, pp. 491-499.

GOH K.-I., KAHNG B., KIM D. (2001) Universal behavior of load distribution in scale-free networks. **Phys. Rev. Lett.**, Vol. 87, 278701.

GRANOVETTER M. (1973) The strength of weak ties. **American Journal of Sociology**, Vol. 78, pp. 1360-1380.

GUIMERÀ R., AMARAL L.A.N. (2004) Modeling the world-wide airport network. **Eur. Phys. J. B**, Vol. 38, pp. 381-385.

GUIMERÀ R., MOSSA S., TURTSCHI A., AMARAL L.A.N. (2005) The worldwide air transportation network: anomalous centrality, community structure, and cities' global rules. **Proc. Natl. Acad. Sci. (USA)**, Vol. 102, pp. 7794-7799.

KRAUSE A.E., FRANK K.A., MASON D.M., ULANOWICZ R.E., TAYLOR W.W. (2003) Compartments revealed in food-web structure. **Nature**, Vol. 426, pp. 282-285.

LI W., CAI X. (2004) Statistical analysis of airport network China. **Phys. Rev. E**, Vol. 69, 046106.

NEWMAN M.E.J. (2002) Assortative mixing in networks. **Phys. Rev. Lett.**, Vol. 89, 208701.

O'KELLY M.E., MILLER H.J. (1994) The hub network design problem. **Journal of transport geography**, Vol. 2, pp. 31-40.

O'KELLY M.E. (1998) A geographer's analysis of hub-and-spoke networks. **Journal of Transport Geography**, Vol. 6, pp. 171-186.

PANDYA R.V.R. (2004) A note on 'Weighted evolving networks : coupling topology and weight dynamics'. <http://arxiv.org>, arXiv :cond-mat/0406644.

PASTOR-SATORRAS R., VESPIGNANI A. (2001) Epidemic spreading in scale-free networks. **Phys. Rev. Lett.**, Vol. 86, pp. 3200-3203.

PASTOR-SATORRAS R., VESPIGNANI A. (2004) **Evolution and structure of the Internet: A statistical physics approach**. Cambridge, Cambridge University Press, 284 p.

WANG W.-X., HU B., YAN G., OU Q., WANG B.-H. (2005) General dynamics of topology and traffic on weighted technological networks. **Phys. Rev. Lett.**, Vol. 94, 188702.

WATTS D.J., STROGATZ S.H. (1998) Collective dynamics of 'small-world' networks. **Nature**, Vol. 393, pp. 440-442.

YOOK S.H., JEONG H., BARABÁSI A.-L., TU Y. (2001) Weighted evolving networks. **Phys. Rev. Lett.**, Vol. 86, pp. 5835-5838.

ZHENG D., TRIMPER S., ZHENG B., HUI P.M. (2003) Weighted scale-free networks with stochastic weight assignments. **Phys. Rev. E**, Vol. 67, 040102.

ZHOU S., MONDRAGON R.J. (2003) The rich-club phenomenon in the Internet topology. **IEEE Comm. Lett.**, Vol. 8, pp. 180-182.