



Scale-free, small-world networks et géographie

Laurent Beauguitte, César Ducruet

► To cite this version:

Laurent Beauguitte, César Ducruet. Scale-free, small-world networks et géographie. 2011. halshs-00601211

HAL Id: halshs-00601211

<https://shs.hal.science/halshs-00601211>

Preprint submitted on 17 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Scale-free, small-world networks et géographie

Laurent Beauguitte (bl<at>parisgeo.cnrs.fr)

César Ducruet (cdu<at>parisgeo.cnrs.fr)

CNRS, UMR Géographie-cités

Version 1 - Juin 2011



Introduction

Jusqu'à la fin des années 1990, la répartition des tâches concernant l'étude des réseaux suivait, en France comme dans les pays anglophones, une division disciplinaire nette en science sociale : les géographes étudiaient les réseaux planaires, les sociologues les réseaux non planaires. L'arrivée des physiciens¹ a quelque peu perturbé cette division du travail en proposant deux nouveaux modèles de graphes qui ont connu une popularité quasi instantanée : les réseaux petit-monde (*small world networks*) et les réseaux invariants d'échelle (*scale-free networks*).

Ce document présente de façon aussi claire que possible les principes de ces deux types de réseaux et quelques unes des mesures proposées (sur ce sujet, voir aussi [9] et [8]). L'utilisation (possible ou effective) de ces modèles en géographie est évoquée dans la dernière partie.

1 Le modèle du *small-world network*

Le concept (ou notion) de *small-world* est issu des travaux bien connus de Stanley Milgram dans les années 60. En deux mots, l'expérience consistait à demander à des habitants du Middle West de faire parvenir une lettre à un destinataire de la Côte Ouest en utilisant comme intermédiaire des personnes qu'ils appelaient par leur prénom. Milgram eu la surprise de constater que la moyenne des chaînes parvenues au destinataire n'était que de 5.6[14]. Cette expérience a donné lieu à un mythe devenu dans sa version populaire les

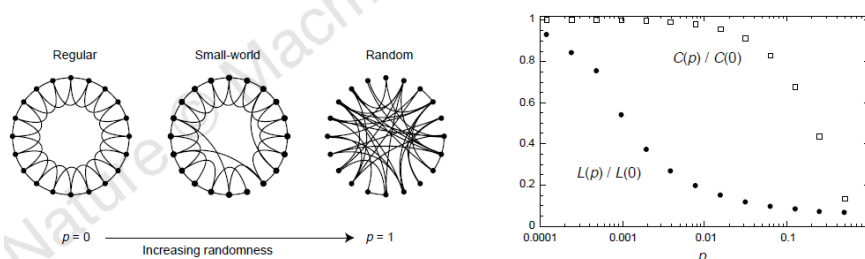
1. Ce label est quelque peu réducteur dans la mesure où il agrège physiciens, mathématiciens (Watts) et informaticiens.

six degrés de séparation : en clair, seules cinq personnes nous séparent de n'importe quelle autre personne dans le monde.

Watts et Strogatz ([22]) publient un article dans *Nature* en 1998 qui propose un modèle de graphe petit-monde (*small world network*) qui diffère du modèle dominant de graphe utilisé auparavant (graphe aléatoire ou *random graph*²).

La figure 1 résume le type théorique de graphe proposé et sa méthode de construction. Le graphe de départ est k -régulier (tous les sommets ont le même degré) et chaque sommet est lié à ses voisins proches. De façon aléatoire, un lien est supprimé et un lien ajouté au graphe de départ. Les auteurs mesurent deux paramètres : la moyenne des plus courts chemins entre chaque paire de sommets (le graphe est non orienté) et ce qu'ils appellent le *clustering coefficient*. Cette mesure n'est pas innovante dans la mesure où il s'agit, avec un nom différent, de la transitivité (*transitivity*) déjà utilisée en sociologie (nombre de triades fermées - ou triangles - divisé par le nombre de triades connexes).

FIGURE 1 – Le modèle du *small-world network*



Les graphes de gauche montrent comment le graphe k -régulier est peu à peu transformé en graphe aléatoire par suppression - ajout de liens. La figure de droite mérite plus d'attention : en abscisse, $L(p)$ désigne la longueur moyenne du plus court chemin entre les sommets et $C(p)$ le *clustering coefficient*. Les deux mesures sont normalisées en étant divisées par les mesures équivalentes d'un graphe k -régulier ($L(0)$ et $C(0)$). L'utilisation d'une échelle logarithmique permet de mettre en évidence la chute de $L(p)/L(0)$ quand $C(p)/C(0)$ reste élevé.

La définition donnée est extensive : est considéré comme graphe petit-monde tout réseau possédant une transitivité supérieure et un diamètre inférieur à un graphe aléatoire de même ordre et de même taille.

Un *small-world network* est donc *in fine* un réseau connexe où le degré

2. Modèle de graphe simple (ni boucle ni liens multiples) étudié par une série d'article d'Erdős et Rényi à la fin des années 50 où les liens entre sommets sont créés de façon aléatoire.

de *cliquishness* est élevé - beaucoup de sous-graphes fortement connexes - et où la distance moyenne entre paire de sommets est faible. Le fait d'utiliser l'expression de « réseau petit monde » dès que l'on étudie les relations au sein d'un petit groupe peut donc être considéré comme un léger abus de langage.

Si les mesures utilisées sont connues, l'intérêt principal de cette approche est de proposer un type théorique de graphe, ce dont se souciaient peu les sociologues. L'inconvénient est que tous les graphes ou presque issus de données empiriques, et notamment les graphes non planaires, correspondent à la définition du *small-world network*. Il faut également ajouter que le modèle initial ne prend en compte ni la direction des liens ni leur intensité.

2 Les *scale-free networks*

Moins d'un an plus tard, deux physiciens, Barabási et Albert, publient dans *Science* un article où ils proposent un autre modèle de graphe et proposent le terme devenu classique depuis de *scale-free networks*[2]. L'approche et la méthodologie sont radicalement différentes. Les auteurs partent de l'étude de très gros graphes dont ils cherchent à modéliser la structure et les règles de construction. Contrairement à ce qui était communément admis, les graphes qu'ils étudient ne présentent pas une distribution gaussienne des degrés mais au contraire une distribution de type loi de puissance. En effet, si on représente en abscisse le degré et en ordonnée la fréquence sur un diagramme bi-logarithmique, on obtient la figure 2.

Le réseau est dit invariant d'échelle si, dans l'équation $P(k) \sim k^{-\gamma}$, où $P(k)$ désigne la fréquence³ et k le degré, γ est supérieur à 2.

Pour expliquer cette distribution très inégalitaire des degrés, les auteurs émettent et testent des hypothèses concernant la formation et l'évolution du réseau. Les graphes empiriques étudiés mettent en évidence, non des liens aléatoires, mais des liens dûs à un processus d'attachement préférentiel (*preferential attachment*). L'exemple le plus convaincant est celui d'internet : tout créateur d'un site sur un sujet précis crée une rubrique liens qui envoie de façon préférentielle vers les sites de référence concernant ce sujet, et la centralité de ces sites de référence tend à augmenter au cours du temps. Dans une thèse récente[23], Zaidi a montré que les propriétés de ce type de graphes étaient bien connues dans la littérature, que ce soit en économie, en sociométrie ou dans les études urbaines.

Si les principes de base des deux modèles sont différents (structure égalitaire pour les *small-world*, inégalitaire pour les *scale-free*), un graphe peut être à la fois l'un et l'autre[15].

3. Les auteurs évoquent la probabilité que l'acteur k interagisse avec les autres sommets.

FIGURE 2 – Distribution des degrés dans les réseaux invariants d'échelle

Version scientifique, 1999, *Nature*, p. 510.

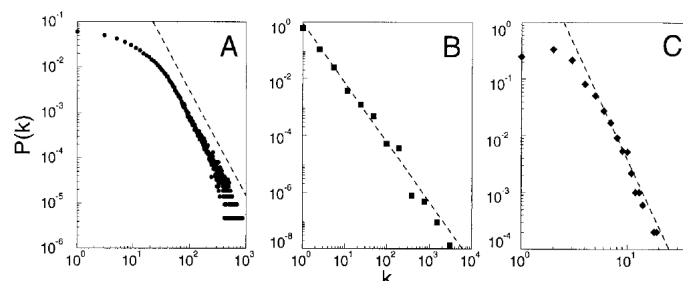
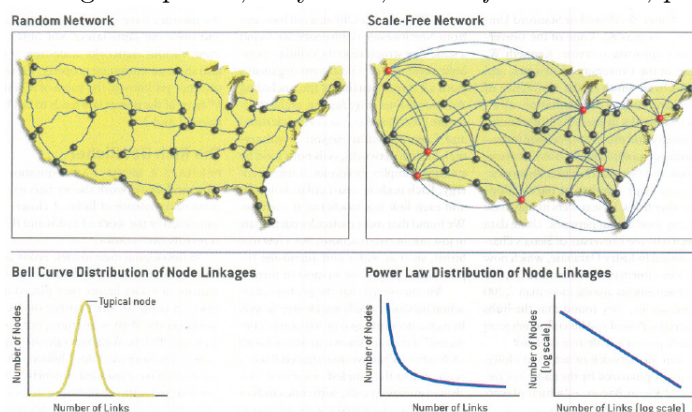


Fig. 1. The distribution function of connectivities for various large networks. **(A)** Actor collaboration graph with $N = 212,250$ vertices and average connectivity $\langle k \rangle = 28.78$. **(B)** WWW, $N = 325,729$, $\langle k \rangle = 5.46$. **(C)** Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{\text{actor}} = 2.3$, (B) $\gamma_{\text{www}} = 2.1$ and (C) $\gamma_{\text{power}} = 4$.

Version grand public, May 2003, *Scientific American*, p. 53.



Si l'idée simple de Barabási a eu un tel écho, c'est aussi qu'il a su, mieux que d'autres, s'adresser simultanément à des publics très variés... Le fait que toutes les publications ou presque des physiciens soient immédiatement accessibles gratuitement en ligne a également aidé à la diffusion de ces méthodes et de ce vocabulaire.

3 Des modèles aux mesures

Ces deux modèles théoriques de graphe ont entraîné la constitution d'un courant de recherche très actif concernant tant les physiciens que les informaticiens. De très nombreuses mesures ont été proposées et, si certaines ne font que reprendre sous un autre nom des mesures déjà connues, d'autres sont innovantes, même s'il n'est pas toujours facile de déterminer leur degré d'utilité. Cette section ne présente pas toutes les mesures proposées mais celles qui semblent les plus prometteuses (pour une revue extensive et quasi

exhaustive des indices, voir la synthèse déjà datée de Bocalleti *et al.*[5] et le récent manuel de Newman[16]).

Représenter la fréquence de la distribution des degrés donne une information utile sur le degré de hiérarchie présent dans le graphe. En cas de graphe orienté, représenter les *in* et les *out-degree* peut s'avérer intéressant. Le fait de représenter l'information sous forme de courbes permet également des comparaisons temporelles plus aisées.

La mesure de l'« assortativité » ou de la « disassortativité » d'un graphe se fait en calculant dans un premier temps le degré moyen des voisins des sommets de degré k , noté $k_{nn}(k)$. Si $k_{nn}(k)$ est une fonction croissante de k , le graphe est dit assortatif ; si c'est une fonction décroissante de k , le graphe est dit disassortatif. En clair, dans le premier cas, les sommets sont liés à des sommets de degré équivalent au leur ; dans le deuxième cas, les sommets se lient à des sommets ayant des degrés très différents des leurs.

Un peu plus complexe en apparence, le *rich-club coefficient* permet de vérifier si les sommets de degré élevé tendent à être fortement liés les uns aux autres. Il se calcule à l'aide de la formule :

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)}$$

où $E_{>k}$ est le nombre de liens parmi les $N_{>k}$ sommets dont le degré est supérieur à k , et $N_{>k}(N_{>k} - 1)/2$ le nombre maximal de liens entre ces $N_{>k}$ sommets⁴. Afin de limiter les effets de taille, l'indicateur normalisé est *rich-club phenomonom* :

$$\phi_{ran}(k) = \frac{\phi(k)}{\phi_{ran}(k)}$$

où $\phi_{ran}(k)$ est le *rich-club coefficient* d'un graphe aléatoire de même ordre que le graphe étudié. En cas de $\phi_{ran}(k) > 1$, on parlera d'effet *rich-club*.

4 Intérêts et limites pour la géographie

Le vocabulaire employé paraît séduisant : petit-monde et invariance d'échelle peuvent sembler des notions facilement importables en géographie. En ce qui concerne le modèle réseau petit-monde, il est devenu une expression fourre-tout employée à tort et à travers dès lors qu'un petit groupe de personnes est étudié. Le principal obstacle à son importation en géographie est son côté trivial : en dehors des graphes k -réguliers (tous les sommets sont de degré k) et des graphes aléatoires, tous les graphes ou presque sont des

4. Cela équivaut à un simple calcul de densité parmi les sommets ayant un degré supérieur à un seuil donné.

graphes petit-monde. ... En ce qui concerne les réseaux invariants d'échelle, la boîte à outils des physiciens s'avère sans doute plus utile.

L'intégration des « réseaux complexes » en géographie reste encore très limitée comme le prouve le faible nombre de travaux recensés sur ce thème, même en élargissant le champ à la science régionale. D'après Waters[21], ceci s'expliquerait par la stagnation de l'analyse de réseau au sens large en géographie depuis les années 1960, et ceci malgré les progrès réalisés dans le domaine des SIG.

Parmi les travaux pionniers, on peut citer celui de Jiang et Claramunt [13] sur la recherche de propriétés *small-world* dans la voirie urbaine, ou Gorman *et al.*[12] sur les réseaux Internet. Fruit d'une collaboration entre géographes et informaticiens, l'analyse multi-niveaux du réseau aérien mondial par Amiel *et al.* [1] a été l'un des premiers articles français sur la question. Si Batty [4] est reconnu par certains comme le pionnier de la recherche en géographie sur les « réseaux complexes », son travail reste essentiellement théorique et peu cité par les recherches empiriques.

La plupart des géographes ont par la suite appliqué des mesures de centralité à différents réseaux, comme ceux des multinationales [18], des collaborations scientifiques[6], du réseau aérien chinois [20], ou de la combinaison entre les réseaux mondiaux aérien et maritime[10]. Tandis que les physiciens incorporent de façon croissante la dimension spatiale dans leurs recherches, créant la catégorie des *spatial networks*, les géographes tiennent étrangement peu compte de la distance comme paramètre déterminant dans l'organisation des réseaux.

Certains travaux utilisent les réseaux complexes comme complément à d'autres méthodes d'analyse, comme Patuelli *et al.* [17] avec les *Spatial Interaction Models* (SIM), Gorman and Kulkarni [11] avec l'équivalence structurelle, et Ducruet *et al.* [10] avec les corrélations de matrices (*Quadratic Assignment Procedure*, QAP) et des éléments plus classiques de la théorie des graphes plus classique (flux majeurs). Seuls Vinciguerra et Frenken [19] s'attachent à la dimension dynamique des réseaux à travers l'application du modèle de Barabási et Albert sur l'attachement préférentiel. Mais les géographes comme les physiciens se contentent la plupart du temps d'une description statique des propriétés topologiques du réseau observé.

La date de parution de la plupart des travaux cités semble cependant montrer une intégration croissante de ces méthodes dans les travaux géographiques sur les réseaux.

Conclusion

En une décennie, les notions de *small-world* et de *scale-free networks* sont devenues très répandues, tout comme celle de *complex networks*. Or, le caractère innovant de deux premières est très relatif : si le réseau petit-monde

FIGURE 3 – Quand les physiciens s’attaquent à la morphologie urbaine

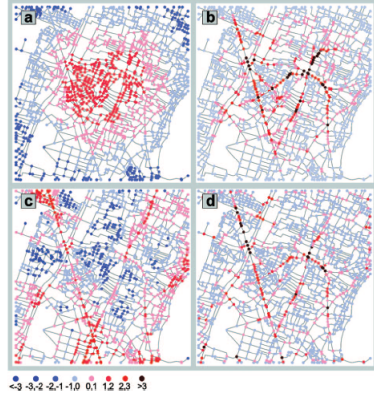


FIG. 4. (Color) Thematic color map representing the spatial distributions of centrality in Cairo, an example of a largely self-organized city. The four indices of node centrality, (a) closeness C^C , (b) betweenness C^B , (c) straightness C^S , and (d) information C^I , used in the MCA, are visually compared over the primal graph. Different colors represent classes of nodes with different values of the centrality index. The classes are defined in terms of multiples of standard deviations from the average, as reported in the color legend.

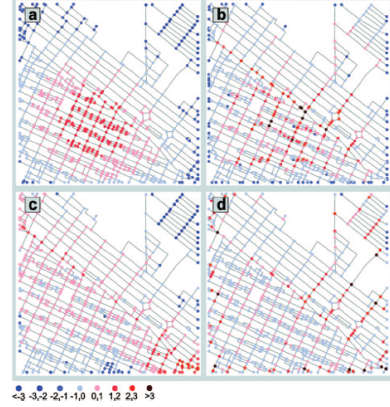


FIG. 5. (Color) Thematic color map representing the spatial distributions of the same four centrality indices of Fig. 4 in Richmond, an example of a mostly planned city.

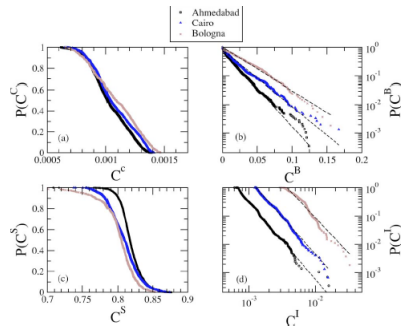


FIG. 6. Cumulative distributions of (a) closeness C^C , (b) betweenness C^B , (c) straightness C^S , and (d) information C^I for three planned cities, Ahmedabad, Cairo, and Bologna. The dashed lines in panels (b) are exponential fits to the betweenness distributions, while the dashed lines in panel (d) are power law fits to the information centrality.

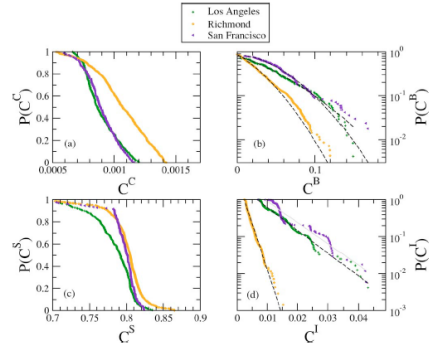


FIG. 7. Cumulative distributions of (a) closeness C^C , (b) betweenness C^B , (c) straightness C^S , and (d) information C^I for three planned cities, Los Angeles, Richmond, and San Francisco. The dashed lines in panels (b) are Gaussian fits to the betweenness distributions, while the dashed lines in panel (d) are exponential fits to the information centrality.

Extraits de Crucitti *et al.*[7]. Physiciens de formation, ils ont écrit une série d’articles qui propose une classification de villes (auto-organisées et planifiées) basée sur des mesures de centralité classiques (*closeness*, *betweenness*) analysées sous l’angle des « réseaux complexes » (courbes de fréquence cumulées, échelles logarithmiques).

est un modèle de graphe astucieux, il est également trivial. L’invariance d’échelle est quant à elle une propriété connue, répandue et étudiée depuis longtemps.

Quant aux devenus incontournables « réseaux complexes », il est utile de rappeler que les définitions données sont peu satisfaisantes. Barrat *et al.* ont beau, suivant la définition canonique des systèmes complexes, évoquer

en substance des réseaux auto-organisés aboutissant à des structures aux propriétés inattendues⁵, un panorama de la littérature sur le sujet permet de dégager deux sens différents majoritairement associés à cet adjectif. Un réseau complexe est d'abord et avant tout un gros réseau, comprenant au minimum plusieurs milliers de sommets et d'arêtes. Plus rarement, un réseau est dit complexe parce qu'il est étudié à plusieurs moments. Mais les propriétés émergentes sont beaucoup plus rarement au rendez-vous.

Ressources internet

Site de Barabási : <http://www.barabasilab.com/>

Quoi qu'on pense du personnage (certains livres et articles sont d'une prétention à toute épreuve), un site riche (liens, cours) et incontournable. Permet de récupérer tous ses articles en version pdf.

Center for the Study of Complex Systems : <http://www.cscs.umich.edu/>

Moins égocentré que le précédent, et plus fourni.

Interdisciplinary Center for Network Science & Applications :

<http://icensa.nd.edu/>

Institut des Systèmes Complexes : <http://www.iscpif.fr/>

Resources in complex networks (São Paulo) :

<http://cyvision.ifsc.usp.br/networks/>

Bonne collection de données téléchargeables.

Santa Fe institute : <http://www.santafe.edu/research/topic/information-computation/>

Complex network : <http://complexnetworks.fr/>.

Site d'informaticiens français. Articles, vidéos.

Enfin, pour réaliser que la modélisation temporelle des graphes a été abordée depuis longtemps par des sociologues, voir la page personnelle de Tom Snijders : <http://stat.gamma.rug.nl/>

Références

- [1] M. AMIEL, G. MÉLANÇON et C. ROZENBLAT : Réseaux multi-niveaux : l'exemple des échanges aériens mondiaux de passagers. *Mappemonde*, (3):12, 2005.
- [2] A.L. BARABÁSI et R. ALBERT : Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [3] A. BARRAT, M. BARTHÉLÉMY et A. VESPIGNANI : *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

5. “All of these networks are self-organizing systems, which at the end of their evolution show an emergent architecture with unexpected properties and regularities”, [3], p. 47.

- [4] M. BATTY : Editorial ; Cities as small worlds. *Environment and Planning B : Planning and Design*, 28(5):637–638, 2001.
- [5] S. BOCCALETTI, V. LATORA, Y. MORENO, M. CHAVEZ et D.U. HWANG : Complex networks : Structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006.
- [6] M.N. COMIN : *Réseaux de villes et réseaux d’innovation en Europe : Structuration du système des villes européennes par les réseaux de recherches sur les technologies convergentes*. Thèse de doctorat, Université de Paris I Sorbonne, 2009.
- [7] P. CRUCITTI, V. LATORA et S. PORTA : Centrality in networks of urban streets. *Chaos : An Interdisciplinary Journal of Nonlinear Science*, 16: 015113, 2006.
- [8] C. DUCRUET : Les mesures locales d’un réseau. *Groupe fmr*, 10 p., 2010 (<http://halshs.archives-ouvertes.fr/FMR/fr/>).
- [9] C. DUCRUET : Les mesures globales d’un réseau. *Groupe fmr*, 8 p., 2010 (<http://halshs.archives-ouvertes.fr/FMR/fr/>).
- [10] C. DUCRUET, D. IETRI et C. ROZENBLAT : Cities in Worldwide Air and Sea Flows : A multiple networks analysis. *Cybergeo*, (528), 2011(<http://cybergeo.revues.org/23603>).
- [11] S.P. GORMAN et R. KULKARNI : Spatial Small Worlds : New Geographic Patterns for an Information Economy. *Environment and Planning B : Planning and Design*, forthcoming.
- [12] S.P. GORMAN, L. SCHINTLER, R. KULKARNI et R. STOUGH : The revenge of distance : Vulnerability analysis of critical information infrastructure. *Journal of Contingencies and Crisis Management*, 12(2):48–63, 2004.
- [13] B. JIANG et C. CLARAMUNT : Topological analysis of urban streets. *Environment and Planning B : Planning and Design*, 31(1):151–162, 2004.
- [14] S. MILGRAM et J. TRAVERS : An Experimental Study of the Small World Problem. *Sociometry*, 32(4):425–443, 1969.
- [15] M. NEWMAN, A.L. BARABÁSI et D.W. WATTS : *The Structure and Dynamics of Networks*. Princeton studies in complexity. Princeton University Press, Princeton, 2006.
- [16] M.E.J. NEWMAN : *Networks : An introduction*. Oxford University Press, 2010.
- [17] R. PATUELLI, A. REGGIANI, S. GORMAN, P. NIJKAMP et F. BADE : Network analysis of commuting flows : A comparative static approach to German data. *Networks and Spatial Economics*, 7(4):315–331, 2007.

- [18] C. ROZENBLAT : Opening the black box of agglomeration economies for measuring cities' competitiveness through international firm networks. *Urban Studies*, 47(13):2841–2865, 2010.
- [19] S. VINCIGUERRA, K. FRENKEN et M. VALENTE : The geography of internet infrastructure : an evolutionary simulation approach based on preferential attachment. *Urban Studies*, 47(9):1969–1984, 2010.
- [20] J. WANG, H. MO, F. WANG et F. JIN : Exploring the network structure and nodal centrality of China's air transport network : A complex network approach. *Journal of Transport Geography*, 2011.
- [21] N. WATERS : *Spatial dynamics, networks and modelling*, chapitre Network and Nodal Indices. Measures of Complexity and Redundancy : A Review, pages 13–33. Edward Elgar Publishing, 2006.
- [22] D.J. WATTS et S. STROGATZ : Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.
- [23] F. ZAIDI : *Analysis, structure and organization of complex networks*. Thèse de doctorat, LABRI, Université de Bordeaux I, 2010.