The role of geography and traffic in the structure of complex networks

Marc Barthelemy, Alain Barrat, Alessandro Vespignani

The role of geography and traffic
in the structure of complex networks

MARC BARTHELEMY*

School of Informatics and Biocomplexity Center
Indiana University, Bloomington, IN 47406, USA
mbarthel@indiana.edu

ALAIN BARRAT

Laboratoire de Physique Théorique (UMR du CNRS 8627)
Bâtiment 210, Université de Paris-Sud 91405 Orsay, France
alain.barrat@th.u-psud.fr

ALESSANDRO VESPIGNANI

School of Informatics and Biocomplexity Center
Indiana University, Bloomington, IN 47406, USA
alexv@indiana.edu

We report a study of the correlations among topological, weighted and spatial properties of large infrastructure networks. We review the empirical results obtained for the air transportation infrastructure that motivates a network modeling approach which integrates the various attributes of this network. In particular we describe a class of models which include a weight-topology coupling and the introduction of geographical attributes during the network evolution. The inclusion of spatial features is able to capture the appearance of non-trivial correlations between the traffic flows, the connectivity pattern and the actual distances of vertices. The anomalous fluctuations in the betweenness-degree correlation function observed in empirical studies are also recovered in the model. The presented results suggest that the interplay between topology, weights and geographical constraints is a key ingredient in order to understand the structure and evolution of many real-world networks.

Keywords: Networks; Traffic; Transportation systems.

1. Introduction

The empirical evidence coming from studies on systems belonging to areas as diverse as social sciences, biology and computer science have shown that in a wide range

---

*On leave of absence from CEA-Centre d'Etudes de Bruyères-le-Châtel, Département de Physique Théorique et Appliquée BP12, 91680 Bruyères-Le-Châtel, France.

of networks the occurrence of vertices with a very large degree (number of links to other vertices) is very common [2, 17, 4, 43]. The presence of these "hubs" often goes along with very large degree fluctuations. The large topological heterogeneity associated to these features is statistically expressed by the presence of heavy-tailed degree distributions with diverging variance that have a very strong impact on the networks' physical properties such as resilience and vulnerability, or the propagation of pathogen agents [15, 38, 1, 42].

The purely topological definition of networks, however, misses important attributes which are frequently encountered in real-world networks. In the first instance, networks are far from a binary structure and are better represented as weighted graphs with the intensity of links that may vary over many orders of magnitude. Indeed, in many graphs ranging from food-webs to metabolic networks, large variations of the link intensities are empirically observed [24, 19, 31, 32, 29, 7, 3]. Notably, the statistical properties of weights indicate non-trivial correlations and association with topological quantities [7].

Another important element of many real networks is their embedding in the real space. For instance, most people have their friends and relatives in their neighborhood, transportation networks depend on distance, and many communication networks include devices with definite transmission ranges [27, 37, 23, 20, 21]. A particularly important example of such a "spatial" network is the Internet which is a set or routers linked by physical cables with different lengths and latency times [30, 43]. An analogous situation is faced in the air transportation network with routes covering very different distances. The length of the link is a very important quantity usually associated with an intrinsic cost in the establishment of the connection. If the cost of a long-range link is high, most of the connections starting from a given node will go to the closest neighbors in the embedding space. Long-range links, on the other hand, correspond usually to connections towards already well-connected nodes (hubs). This seems natural in the case of the air transportation network for instance: short connections go to small airports while long distance flights are directed preferentially towards large airports (*i.e.* well connected nodes). It is therefore natural to find that spatial constraints can have important consequences on the topology of the resulting network [13, 25, 34]. This issue is particularly important in spatial economics where the evolution of an economic system depends strongly on the geographical distribution of the economic activity [5].

In this article, we discuss the interplay of the three aforementioned ingredients (heterogeneous topology, weights and spatial constraints) in a model of growing network combining these ingredients at once. The proposed model is obtained as the embedding of the weighted growing network introduced in [8] in a two-dimensional geometrical space. Spatial constraints are translated into a preference for short links, and combined with the coupling between the evolution of the network and the dynamical rearrangement of the weights. This mechanism naturally leads to the appearance of many features observed in real-world networks, in particular the

non-linear correlations between weights and topology, and the large fluctuations of the betweenness centrality. Let us note that although we were motivated by the airport network, we don't intend here to present a specific model for this network but rather to investigate the effects of different ingredients.

The paper is organized as follows. In section 2, we briefly review some important empirical results of the North-American airline network, highlighting the most salient geographical effects. Sections 3 and 4 are devoted to the presentation and to the numerical study of the spatial weighted model, stressing the effect of the spatial embedding and constraints on the properties of the resulting network. In section 5, we present a summary of the results and conclusions about large network modeling.

2. A case study: Space, topology and traffic in the North American airline network

2.1. *Topological characterization*

The characteristics of the world-wide air-transportation network (WAN) using the International Air Transportation Association (IATA) database [28] have been presented in [7]. The network is made of $N = 3880$ vertices and $E = 18810$ edges and shows both small-world and scale-free properties as also confirmed in different datasets and analyses [31, 32, 26, 25]. In particular, the average shortest path length, measured as the average number of edges separating any two nodes in the network shows the value $\langle \ell \rangle \simeq 4.4$, very small compared to the network size N. The degree distribution, takes the form $P(k) = k^{-\gamma} f(k/k_x)$, where $\gamma \simeq 2.0$ and $f(k/k_x)$ is an exponential cut-off function. The degree distribution is therefore heavy-tailed with a cut-off that finds its origin in the physical constraints on the maximum number of connections that a single airport can handle [26, 25, 4]. The airport connection graph is therefore a clear example of small-world network showing a heavy-tailed degree distribution and heterogeneous topological properties. This heavy-tailed distribution is the signature of very large degree fluctuations [a] which will have a crucial impact on all dynamical processes which take place on this type of networks.

The world-wide airline network necessarily mixes different effects. In particular there are clearly two different spatial scales, global (intercontinental) and domestic. The intercontinental scale defines two different groups of travel distances and for the statistical consistency we eliminate this specific geographical constraint by focusing on a single continental case. Namely, in the following we will essentially consider the North-American network constituted of $N = 935$ vertices with an average degree $\langle k \rangle \approx 8.4$ and an average shortest path $\langle \ell \rangle \simeq 3.9$. The statistical topological properties of the North American network are consistent with the WAN. In particular, the North American network presents a degree distribution statistically consistent with the world-wide airline network. Indeed, we observe (Fig. 1) also in this case a

---

[a]This can also be observed by a very large value of the variance or equivalently by the fact that $\langle k^2 \rangle \gg \langle k \rangle^2$.

4   *Barthélemy, Barrat, Vespignani*

power-law behavior on almost two orders of magnitude, followed by a cut-off indicating the maximum number of connections possible due to limited airport capacity and to the size of the network considered.
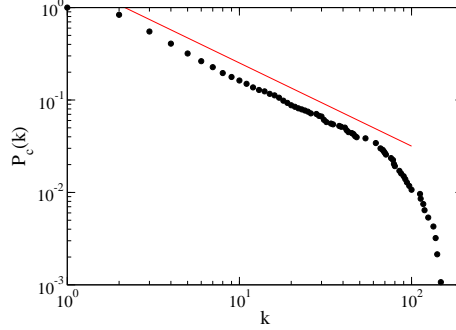

Fig. 1. Cumulative degree distribution $P_c(k)$ for the North-American network. The straight line indicates a power-law decay with exponent $\gamma - 1 = 0.9$.

A further characterization of the network is provided by considering quantities that takes into account the global topology of the network. For instance, the degree of a vertex is a local measure that gives a first indication of its centrality. However, a more global approach is needed in order to characterize the real importance of various nodes. Indeed, some particular low-degree vertices may be essential because they provide connections between otherwise separated parts of the network. In order to take properly into account such vertices, the betweenness centrality (BC) is commonly used [18, 39, 22, 14]. The betweenness centrality of a node v is defined as

$$g(v) = \sum_{s \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

where σ_{st} is the number of shortest paths going from s to t and $\sigma_{st}(v)$ is the number of shortest paths going from s to t and passing through v. This definition means that central nodes are part of more shortest paths within the network than peripheral nodes. Moreover, the betweenness centrality gives in transport networks an estimate of the traffic handled by the vertices, assuming that the number of shortest paths is a zero-th order approximation to the frequency of use of a given node. It is generally useful to represent the average betweenness centrality for vertices of the same degree

$$g(k) = \frac{1}{N(k)} \sum_{v/k_v = k} g(v) \ . \tag{2}$$

where $N(k)$ is the number of nodes of degree k. For most networks, $g(k)$ is strongly correlated with the degree k and in general, the largest the degree and the largest

the centrality. For scale-free networks it has been shown that the centrality scales with k as

$$g(k) \sim k^{\mu} \tag{3}$$

where μ depends on the network [39, 22, 14] and which value for the North American network is $\mu \approx 1.2$. For some networks however, the BC fluctuations around the behavior given by Eq. (3) can be very large and "anomalies" can occur, in the sense that the variation of the centrality versus degree is not a monotonous function. Guimerà and Amaral [25] have shown that this is indeed the case for the air-transportation network. This is a very relevant observation in that very central cities may have a relatively low degree and vice versa. In Fig. 2 we report the average behavior along with the scattered plot of the betweenness versus degree of all airports of the North American network. Also in this case we find very large fluctuations around the average with a behavior similar to those observed in Ref. [25]. Interestingly, Guimerà and Amaral have put forward a network model embedded in real space that considers geopolitical constraints. This model appears to reproduce the betweenness centrality features observed in the real network pointing out the importance of space as a relevant ingredient in the structure of networks. In the following we focus on the interplay between spatial embedding, topology and weights in a simple general model for weighted networks in order to provide a modeling framework considering these three aspects at once.
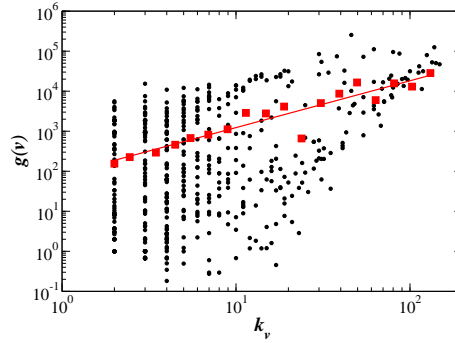

Fig. 2. Scatter-plot of the betweenness centrality versus degree for nodes of the North-American air-transportation network. The red squares correspond to the average BC versus degree. The straight line is a power law fit with exponent $\mu \approx 1.2$.

2.2. *Traffic properties*

The airline transportation infrastructure is a paramount example of large scale network which can be represented as a complex weighted graph: the airports are the vertices of the graph and the links represent the presence of direct flight connections

6   *Barthélemy, Barrat, Vespignani*

among them. The weight on each link is the number of maximum passengers on the corresponding connection.

More generally, the properties of a graph can be expressed via its adjacency matrix a_{ij}, whose elements take the value 1 if an edge connects the vertex i to the vertex j and 0 otherwise (with $i, j = 1, ..., N$ where N is the size of the network). Weighted networks are usually described by a matrix w_{ij} specifying the weight on the edge connecting the vertices i and j ($w_{ij} = 0$ if the nodes i and j are not connected). In the following we will consider only the case of symmetric positive weights $w_{ij} = w_{ji} \geq 0$.

Along with the degree of a node, a very significant measure of the network properties in terms of the actual weights is obtained by looking at the vertex *strength* s_i^w defined as [7]

$$s_i^w = \sum_{j \in \mathcal{V}(i)} w_{ij} \tag{4}$$

where the sum runs over the set $\mathcal{V}(i)$ of neighbors of i. The strength of a node integrates the information both about its connectivity and the importance of the weights of its links and can be considered as the natural generalization of the connectivity. When the weights are independent from the topology, we obtain $s^w \simeq \langle w \rangle k$ where $\langle w \rangle$ is the average weight. In the presence of correlations we obtain in general $s^w \simeq A k^{\beta_w}$ with $\beta_w = 1$ and $A \neq \langle w \rangle$ or $\beta_w > 1$.

For the North-American air transportation network we observe a non linear behavior with exponent $\beta_w \simeq 1.7$ (see Fig. 3) while for the global WAN network [7] we obtain $\beta_w(WAN) = 1.5 \pm 0.1$. These values of β_w larger than one imply that the strength of vertices grows faster than their degree, i.e. the weight of edges belonging to highly connected vertices tends to have a value higher than the one corresponding to a random assignment of weights (in the randomized version of the network where the weights are randomly re-distributed on the existing topology of the network, we checked that we obtain $\beta_w = 1$). There is thus a strong correlation between the weight and the topological properties which means that the larger is an airport, the more traffic it can handle. The fingerprint of these correlations is also observed in the behavior of the average weight as a function of the end points degrees $\langle w_{ij} \rangle \sim (k_i k_j)^\theta$ with an exponent $\theta_{NA} = 0.7 \pm 0.1$. for the North-American network and $\theta_{WAN} = 0.5 \pm 0.1$ for the WAN [7] (a simple argument presented in [7] suggests that in most cases, the exponents θ and β_w are related by $\beta_w = 1 + \theta$).

The topological clustering [45] is defined as the fraction of connected neighbors

$$C(i) = \frac{2E_i}{k_i(k_i - 1)} \tag{5}$$

where E_i is the number of links between the neighbors of node i and $k_i(k_i - 1)/2$ is the maximum number of such links (k_i is the degree of node i).

The topological clustering however doesn't take into account the fact that some neighbors are more important than others. We thus have to introduce a measure of
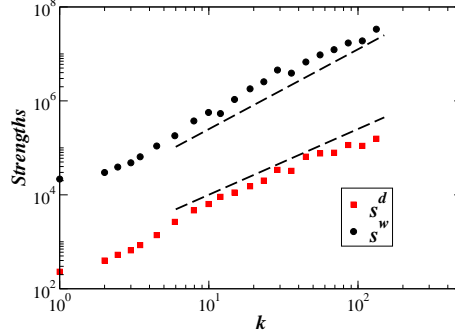
Fig. 3. Weight and distance strengths versus degree for the North-American network. The dashed lines correspond to the power-laws $\beta_d \simeq 1.4$ and $\beta_w \simeq 1.7$.

the clustering that combines the topological information with the weight distribution of the network. The *weighted clustering coefficient* is defined as [7]

$$C^w(i) = \frac{1}{s_i^w(k_i - 1)} \sum_{j,h} \frac{(w_{ij} + w_{ih})}{2} a_{ij} a_{ih} a_{jh}. \tag{6}$$

This quantity $C^w(i)$ is counting for each triple formed in the neighborhood of the vertex i the weight of the two participating edges of the vertex i. The normalization factor $s_i^w(k_i - 1)$ ensures that $0 \leq C_i^w \leq 1$ and that C_i^w recovers the topological clustering coefficient in the case that $w_{ij} = const$. It is customary to define C^w and $C^w(k)$ as the weighted clustering coefficient averaged over all vertices of the network and over all vertices with degree k, respectively. The ratio C^w/C (and similarly $C^w(k)/C(k)$ which allow an analysis with respect to the degree k) indicates if the interconnected triples are more likely formed by the edges with larger weights.

Another quantity used to probe the networks' architecture is the behavior of the average degree of nearest neighbors, $k_{nn}(k)$, for vertices of degree k [44]. This last quantity is related to the correlations between the degree of connected vertices since it can be expressed as $k_{nn}(k) = \sum_{k'} k' P(k'|k)$, where $P(k'|k)$ is the conditional probability that a given vertex with degree k is connected to a vertex of degree k'. In the absence of degree correlations, $P(k'|k)$ does not depend on k and neither does the average nearest neighbors' degree; i.e. $k_{nn}(k) = const.$ [44]. In the presence of correlations, the behavior of $k_{nn}(k)$ identifies two general classes of networks. If $k_{nn}(k)$ is an increasing function of k, vertices with high degree have a larger probability to be connected with large degree vertices. This property is referred in physics and social sciences as *assortative mixing* [40]. On the contrary, a decreasing behavior of $k_{nn}(k)$ defines *disassortative mixing*, in the sense that high degree vertices have a majority of neighbors with low degree, while the opposite holds for low degree vertices.

Along with the weighted clustering coefficient, we can measure the weighted average nearest neighbors degree, defined as

$$k_{nn}^w(i) = \frac{1}{s_i} \sum_{j=1}^{N} a_{ij} w_{ij} k_j. \tag{7}$$

This definition implies that $k_{nn}^w(i) > k_{nn}(i)$ if the edges with the larger weights are pointing to the neighbors with larger degree and $k_{nn}^w(i) < k_{nn}(i)$ in the opposite case. The $k_{nn}^w(i)$ thus measures the effective *affinity* to connect with high or low degree neighbors according to the magnitude of the actual interactions. As well, the behavior of the function $k_{nn}^w(k)$ (defined as the average of $k_{nn}^w(i)$ over all vertices i with degree k) marks the weighted assortative or disassortative properties considering the actual interactions among the system's elements. Figure 4 displays for the North-American airport network the behavior of these various quantities as a function of the degree. An essentially flat $k_{nn}^w(k)$ is obtained and a slight disassortative trend is observed at large k for the $k_{nn}(k)$, due to the fact that large airports have in fact many intercontinental connections to other hubs which are located outside of North America and are not considered in this "regional" network. The clustering is very large and slightly decreasing at large k. This behavior is often observed in complex networks and is here a direct consequence of the role of large airports that provide non-stop connections to different regions which are not interconnected among them. Moreover, weighted correlations are systematically larger than the topological ones, signaling that large weights are concentrated on links between large airports which form well inter-connected cliques (see also [7] for more details).
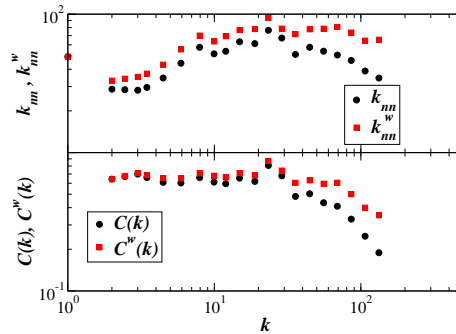

Fig. 4. Assortativity and clustering for the North-American network. Circles correspond to topological quantities while squares are for affinity and weighted clustering.

For a given node i with connectivity k_i and strength s_i^w different situations can arise. All weights w_{ij} can be of the same order s_i^w/k_i. In contrast, the most heterogeneous situation is obtained when one weight dominates over all the others.

A simple way to measure this "disparity" is given by the quantity Y_2 introduced in other context [16, 12]

$$Y_2(i) = \sum_{j \in \mathcal{V}(i)} \left[\frac{w_{ij}}{s_i} \right]^2 \tag{8}$$

If all weights are of the same order then $Y_2 \sim 1/k_i$ (for $k_i \gg 1$) and if a small number of weights dominate then Y_2 is of the order $1/n$ with n of order unity. This quantity was recently used for metabolic networks [3] which showed that for these networks one can identify dominant reactions. For the WAN and the North American network, we observe a behavior consistent with a decay for large degrees of the form $Y_2 \sim 1/k$ which indicates that for large airports, all connections carry essentially the same number of passengers (even if there are obviously differences among these connections).

2.3. *Spatial analysis*

The spatial attributes of the North American airport network are embodied in the physical spatial distance, measured in kilometers or miles, characterizing each connection. Fig. 5 displays the probability distribution $P(d_{ij})$ of the distances d_{ij} of the direct flights. These distances correspond to Euclidean measures of the links between airports and clearly show a fast decaying behavior reasonably fitted by an exponential. The exponential fit gives a value for a typical scale of the order 1000 kms. The origin of the finite scale can be traced back to the existence of physical and economical restrictions on airline planning in a continental setting.
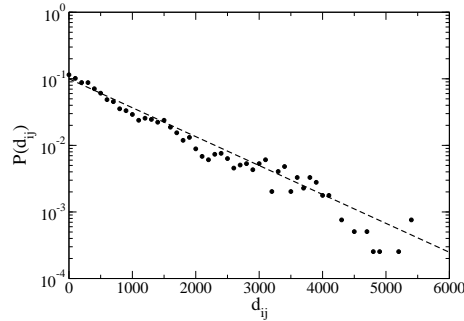

Fig. 5. Distribution of distances (in kms) between airports linked by a direct connection for the North-American network. The straight line indicates an exponential decay with scale of order 1000 km.

Since space is an important parameter in this network, another interesting quan-

tity is the *distance strength* of i

$$s_i^d = \sum_{j \in \mathcal{V}(i)} d_{ij} \tag{9}$$

where d_{ij} are the *Euclidean* distances of the connections departing from the airport i. This quantity gives the cumulated distances of all the connections from (or to) the considered airport. Similarly to the usual weight strength, uncorrelated random connections would lead to a linear behavior of $s^d(k) \propto k$ while we observe in the North-American network a power law behavior

$$s^d(k) \sim k^{\beta_d} \tag{10}$$

with $\beta_d \simeq 1.4$ (Fig. 3). This result shows the presence of important correlations between topology and geography. Indeed, the fact that the exponents appearing in the relations (4) and (10) are larger than one have different meanings. While Eq. (4) means that larger airports have connections with larger traffic, (10) implies that they have also farther-reaching connections. In other terms, the traffic (and the distance) per connection is not constant but increases with k. As intuitively expected, the airline network is an example of a very heterogeneous network where the hubs have at the same time large connectivities, large weight (traffic) and long-distance connections [7], related by super-linear scaling relations.

3. The model

Early modeling of weighted networks just considered weight and topology as uncor-related quantities [48]. This is not the case in real world networks where a complex interplay between the evolution of weights and topological growth does exist as well as spatial constraints. In the following, we discuss the effects of these ingredients on a simple model of an evolving network.

3.1. *The weight-topology coupling ingredient*

Previous approaches to the modeling of weighted networks focused on growing topologies where weights were assigned statically, i.e. once for ever, with differ-ent rules related to the underlying topology [48, 50]. These mechanisms, however, overlook the dynamical evolution of weights according to the topological variations. We can illustrate this point in the case of the airline network. If a new airline con-nection is created between two airports it will generally provoke a modification of the existing traffic of both airports. In the following, we review a model that takes into account the coupled evolution in time of topology and weights. The model dy-namics starts from an initial seed of N_0 vertices connected by links with assigned weight w_0. At each time step, a new vertex n is added with m edges (with initial weight w_0) that are randomly attached to a previously existing vertex i according to the probability distribution

$$\Pi_{n \to i} = \frac{s_i}{\sum_j s_j}. \tag{11}$$

This rule of "busy get busier" relaxes the usual degree preferential attachment, focusing on a strength driven attachment in which new vertices connect more likely to vertices handling larger weights and which are more central in terms of the strength of interactions. This weight driven attachment (Eq. (11)) appears to be a plausible mechanism in many networks [8].

The presence of the new edge (n, i) introduces variations of the existing weights across the network. In particular, we consider the local rearrangements of weights between i and its neighbors $j \in \mathcal{V}(i)$ according to the simple rule

$$w_{ij} \to w_{ij} + \Delta w_{ij}, \quad \text{with} \quad \Delta w_{ij} = \delta \frac{w_{ij}}{s_i}. \tag{12}$$

This rule considers that the establishment of a new edge of weight w_0 with the vertex i induces a total increase of traffic δ that is proportionally distributed among the edges departing from the vertex according to their weights (see Fig. 6), yielding $s_i \to s_i + \delta + w_0$. We will focus on the simplest model with $\delta = const$, but one can consider different choices [9, 10, 41] of Δw_{ij} depending on the specific properties of each vertex (w_{ij}, k_i, s_i). After the weights have been updated the growth process is iterated by introducing a new vertex with the corresponding re-arrangement of weights.
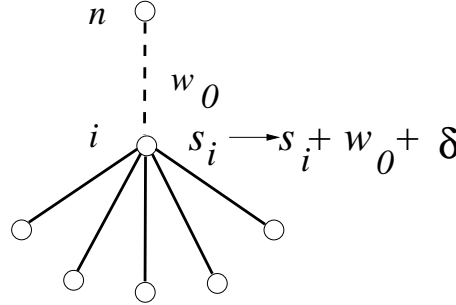

Fig. 6. Illustration of the construction rule. A new node n connects to a node i with probability proportional to $s_i / \sum_j s_j$. The weight of the new edge is w_0 and the total weight on the existing edges connected to i is modified by an amount equal to δ.

The model depends only on the dimensionless parameter δ (rescaled by w_0), that is the fraction of weight which is 'induced' by the new edge onto the others. If $\delta \approx 1$, the traffic generated by the new connection will be dispatched in the already existing connections. In the case of $\delta < 1$ we face situations where a new connection is not triggering a more intense activity on existing links. Finally, $\delta > 1$ is an extreme case in which a new edge generates a sort of multiplicative effect that is bursting the weight or traffic on neighbors.

The network's evolution can be inspected analytically by studying the time evolution of the average value of $s_i^w(t)$ and $k_i(t)$ of the i-th vertex at time t, and

12   *Barthélemy, Barrat, Vespignani*

by relying on the continuous approximation that treats k, s^w and the time t as continuous variables [2, 17, 7]. One obtains $s_i^w(t) = (2\delta+1)k_i(t)$ which implies $\beta_w = 1$ and a prefactor different from $< w >$ which indicates the existence of correlations between topology and weights. The fact that $s^w \propto k$ is also particularly relevant since it states that the weight-driven dynamics generates in Eq. (11) an effective degree preferential attachment that is parameter independent. This highlights an alternative microscopic mechanism accounting for the presence of the preferential attachment dynamics in growing networks. The behavior of the various statistical distribution can be easily computed and one obtains in the large time limit $P(k) \sim k^{-\gamma}$ and $P(s) \sim s^{-\gamma}$ with

$$\gamma = \frac{4\delta + 3}{2\delta + 1}. \tag{13}$$

This result shows that the obtained graph is a scale-free network described by an exponent $\gamma \in [2, 3]$ that depends on the value of the parameter δ. In particular, when the addition of a new edge doesn't affect the existing weights ($\delta = 0$), the model is topologically equivalent to the Barabasi-Albert model [6] and the value $\gamma = 3$ is recovered. It is also possible to show analytically [8] that $P(w) \sim w^{-\alpha}$ where $\alpha = 2 + 1/\delta$. The exponent α has large variations as a function of the parameter δ and this feature clearly shows that the weight distribution is extremely sensible to changes in the microscopic dynamics ruling the network's growth.

3.2. *The spatial ingredient*

Space is an important ingredient in many systems and it is crucial to understand its effect on the formation of networks. We consider this problem on a simple model of a growing weighted network whose nodes are embedded in a two-dimensional space. As in the weighted model above, it is reasonable to think that a newly created node n will establish links towards pre-existing nodes with heavy traffic or strength (hubs). Costs are however usually associated with distances and there is a trade-off between the need to reach a hub in a few hops and the connection costs. The cost naturally increases with the distance implying that the probability of establishing a connection between the new node n and a given vertex i decays as a function of the increasing Euclidean distance d_{ni}. As in the case of topological preferential attachment (i.e. connecting probability proportional to the degree [6]), this trade-off can be expressed in two different ways: the connecting probability can decrease either as a power-law of the distance [33, 47, 49] or as an exponential with a finite typical scale [13] as it seems more natural for networks such as transportation networks (see Fig. 5) or technological networks [46]. All the effects described here are obtained in the case of an exponential decay $\exp(-d_{ni}/r_c)$ but are also present in the case of a power-law d_{ni}^{-a} (the effect of a decreasing scale r_c is qualitatively the same as the effect of an increasing exponent a). Eventually, the creation of new edges will introduce new traffic which will trigger perturbations in the network. This model therefore consists of two combined mechanisms:

(1) *Growth.* We start with an initial seed of N_0 vertices randomly located (with uniform distribution) on a 2-dimensional disk (of radius L) and connected by links with assigned weight w_0. At each time step, a new vertex n is placed on the disk at a randomly assigned position $\mathbf{x}_n$ (still according to a uniform distribution). This new site is connected to m previously existing vertices, choosing preferentially nearest sites with the largest strength. More precisely, a node i is chosen according to the probability

$$\Pi_{n \to i} = \frac{s_i^w e^{-d_{ni}/r_c}}{\sum_j s_j^w e^{-d_{nj}/r_c}}, \tag{14}$$

where r_c is a typical scale and d_{ni} is the Euclidean distance between n and i. This rule of *strength driven preferential attachment with spatial selection*, generalizes the preferential attachment mechanism driven by the strength to spatial networks. Here, new vertices connect more likely to vertices which correspond to the best interplay between Euclidean distance and strength.

(2) *Weights dynamics.* The weight of each new edge (n, i) is fixed to a given value w_0 (this value sets a scale so we can take $w_0 = 1$). The creation of this edge will perturb the existing interactions and we consider *local* perturbations for which only the weights between i and its neighbors $j \in \mathcal{V}(i)$ are modified

$$w_{ij} \to w_{ij} + \delta \frac{w_{ij}}{s_i^w}. \tag{15}$$

After the weights have been updated, the growth process is iterated by introducing a new vertex, i.e. going back to step (1.) until the desired size of the network is reached.

The previous rules have simple physical and realistic interpretations. Equation (14) corresponds to the fact that new sites try to connect to existing vertices with the largest strength, with the constraint that the connection cannot be too costly. This adaptation of the rule "busy get busier" allows to take into account physical constraints. The weights' dynamics Eq. (15) expresses the perturbation created by the addition of the new node and link. It yields a global increase of $w_0 + \delta$ for the strength of i, which will therefore become even more attractive for future nodes.

The model contains two relevant parameters: the ratio between the typical scale and the size of the system $\eta = r_c/L$, and the ability to redistribute weights, δ. Depending on the value of η and δ we obtain different networks whose limiting cases are summarized in Figure 7. More precisely, we expect:

- For $\eta \gg 1$, the effect of distance is negligible and we recover the properties of the weighted model discussed above. The effect of the redistribution parameter δ is to broaden the various probability distributions, and to increase the correlations between topology and weights. Moreover, in this case, no correlations are introduced between the topology of the network and the underlying two-dimensional space, so that the distance strength s^d grows simply linearly with
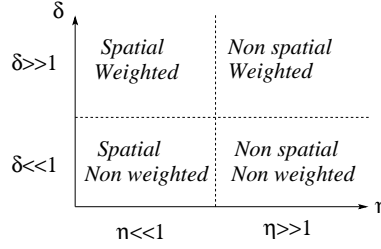
Fig. 7. Different limiting regimes of the model depending on the value of its parameters δ and η.

the degree $s^d \sim k$.

- When η decreases, additional constraints appear and have consequences that we will investigate numerically in the following. Unless otherwise specified, the simulations correspond to the parameters $m = 3$ (i.e. an average degree $\langle k \rangle = 6$), and $\delta = 1.0$. We consider networks of size up to $N = 10,000$, and the results are averaged over up to 100 realizations. All the observed dependences in η are essentially the same for other investigated values of δ.

4. Model's simulation

4.1. *Topology and weights*

At a purely topological level, the principal effect of a typical finite scale r_c in the creation of new connections is to introduce a cut-off in the scale-free degree distributions [13]. In Fig. 8 we report the degree distribution for a fixed value of δ and decreasing values η. A more pronounced cut-off appears at decreasing value of η signaling the onset of a trade-off between the number of connections and their cost in terms of Euclidean distance. The small-world properties of the network are as well modified [13]: on the one hand, the increasing tendency to establish connections in the geographical neighborhood favors the formation of cliques and leads to an increase in the clustering coefficient (see inset of Fig. 9). On the other hand, this same tendency leads to an increase in the diameter of the graph, measured as the average shortest path distances between pairs of nodes. The diameter however still increases logarithmically with the size of the graph, as shown in Fig. 9: the constructed networks do display the small-world property, even if strong geographical constraints are present.

The correlations appearing between space, traffic and topology of the network are affected by the presence of spatial constraints. Strikingly, the effect of the spatial constraint is to increase both exponents β_w and β_d to values larger than 1 and although the redistribution of the weights [Eq. (15)] is linear, non-linear relations $s^w(k)$ and $s^d(k)$ as a function of k appear. For the weight strength the effect is not very pronounced with an exponent of order $\beta_w \approx 1.1$ for $\eta = 0.01$, while for the distance strength the non-linearity has an exponent of order $\beta_d \approx 1.27$ for $\eta = 0.02$
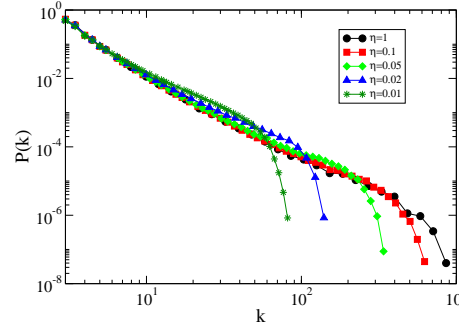
Fig. 8. Degree distribution $P(k)$ for different values of η and $\delta = 1$. The degree distribution is averaged over 50 networks of size $N = 10^4$ and minimum degree $m = 3$.
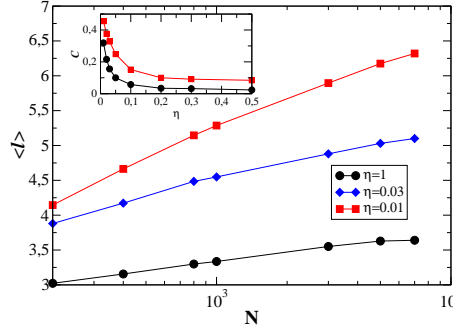

Fig. 9. Average shortest path distances as a function of the network size N for different values of η, with $m = 3$ and $\delta = 1$. Inset: clustering coefficient as a function of η for $N = 10^4$, $\delta = 1$ (circles) and $\delta = 5$ (squares). The data are averaged over 50 networks.

(the value of the exponents β_w and β_d depend on η; see also [36] for a spatial model with $\beta_d > 1$). We show on Fig. 10 the distance strength for two extreme situations for which spatial constraints are inexistent ($\eta = 10.0$) or on the contrary very strong ($\eta = 0.02$).

The nonlinearity induced by the spatial structure can be explained by the following mechanism affecting the network growth. The increase of spatial constraints affects the trend to form global hubs, since long distance connections are less probable, and drives the topology towards the existence of "regional" hubs of smaller degree. The total traffic however is not changed with respect to the case $\eta = \infty$, and is in fact directed towards these "regional" hubs. These medium-large degree vertices therefore carry a much larger traffic than they would do if global "hubs" were available, leading to a faster increase of the traffic as a function of the degree, eventually resulting in a super-linear behavior. Moreover, as previously mentioned, the
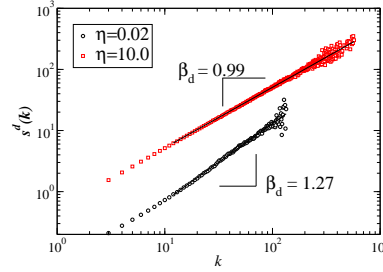
Fig. 10. Distance strength versus k for $\eta = 0.02$ and $\eta = 10.0$ (the networks are obtained for $\delta = 1$, $N = 10^4$, $\langle k \rangle = 6$, and averaged over 100 configurations). When η is not small, space is irrelevant and there are no correlations between degree and space. When spatial effects are important ($\eta = 0.02 \ll 1$), non-linear correlations appear. We observe a crossover for $k \simeq 10 - 20$ to a power-law behavior and the power-law fit over this range of values of k is shown (full lines).

increase in distance costs implies that long range connections can be established only towards the hubs of the system: this effect naturally lead to a super-linear accumulation of $s^d(k)$ at larger degree values.

Spatial constraints have also a strong effect on the correlations between neighboring nodes (Fig. 11). At large η, a disassortative network is created, as is the case in most growing networks [11]; as η decreases, k_{nn} decreases, and an increasing range of flat $k_{nn}(k)$ appears: the tendency for small nodes to connect to hubs is contrasted by the need to use small-range links. For small enough η, a nearly neutral behavior more similar to what is actually observed in the airport network is reached. Moreover, the affinity of nodes to establish strong links to large nodes, measured by $k_{nn}^w(k)$, goes from a flat behavior at large η to a slightly assortative one at small η. In all cases, the weighted correlation $k_{nn}^w(k)$ remains clearly larger than the unweighted $k_{nn}(k)$, showing that links to busier nodes are typically stronger.

A non-trivial clustering hierarchy is already displayed by the model without spatial constraints. As previously mentioned, the decrease of η leads to an increase of clustering. Moreover, the weighted clustering is always significantly larger than the unweighted one, showing that the cliques carry typically an important traffic (see Fig. 11). These effects are a general signature of spatial constraints as also observed in a non weighted network [13].

4.2. *Spatial constraints and betweenness centrality*

The spatial constraints act at both local and global level of the network structure by introducing a distance cost in the establishment of connections. It is therefore important to look at the effect of space in global topological quantities such as the betweenness centrality. The betweenness centrality of a vertex is determined by its ability to provide a path between separated regions of the network. Hubs are natural crossroads for paths and it is natural to observe a marked correlation between g
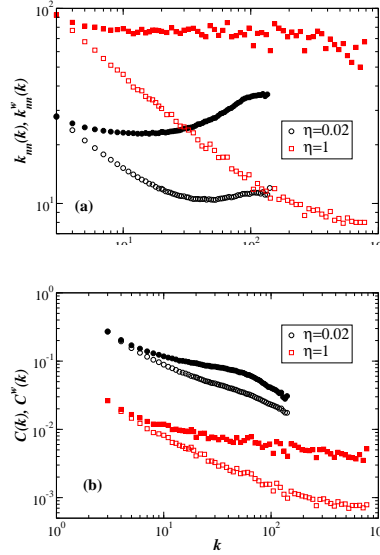
Fig. 11. (a) Assortativity and (b) clustering versus k obtained for the model for $\delta = 1$, $N = 10^4$, $\langle k \rangle = 6$, $\eta = 0.02$ (circles) and $\eta = 1$ (squares). Data are averaged over 100 configurations. Empty symbols refer to topological correlations while full symbols correspond to the weighted quantities k_{nn}^w and C^w.

and k as expressed in the general relation $g(k) \sim k^\mu$. The exponent μ depends on the characteristics of the network and we expect this relation to be altered when spatial constraints become important. In the present model, Fig. 12 clearly shows that this correlation in fact increases when spatial constraints become large (i.e. when η decreases). This can be understood simply by the fact that the probability to establish far-reaching short-cuts decreases exponentially in Eq. (14) and only the large traffic of hubs can compensate this decay. Far-away geographical regions can thus only be linked by edges connected to large degree vertices, which implies a more central role for these hubs.

In order to better understand the effect of space on the properties of betweenness centrality, we have to explicitly consider the geometry of the network along with the topology. In particular, we need to consider the role of the spatial position by introducing the spatial barycenter of the network. Indeed, in the presence of a spatial structure, the centrality of nodes is correlated with their position with respect to the barycenter G, whose location is given by $\mathbf{x}_G = \sum_i \mathbf{x}_i / N$. For a spatially ordered network—the simplest case being a lattice embedded in a one-dimensional space—the shortest path between two nodes is simply the Euclidean geodesic. In a limited region, for two points lying far away, the probability that the shortest path passes near the barycenter of all nodes is very large. In other words, this implies that the barycenter (and its neighbors) will have a large centrality. In
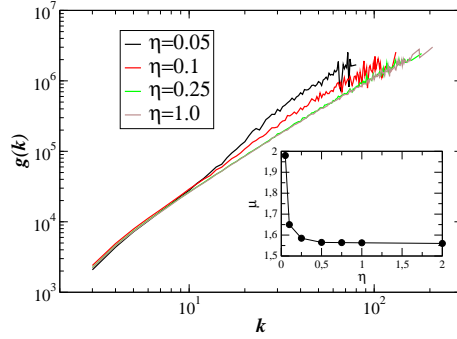
Fig. 12. Betweenness centrality versus degree for different values of η. Inset: Exponent μ (obtained by fitting the data for $k > 10$) of the betweenness centrality versus η (for $N = 5,000$, $m = 3$ and averaged over 50 configurations). For strong spatial constraints long-range shortcuts are very rare and hubs connect regions which are otherwise almost disconnected which in turn implies a larger centrality of the hubs.
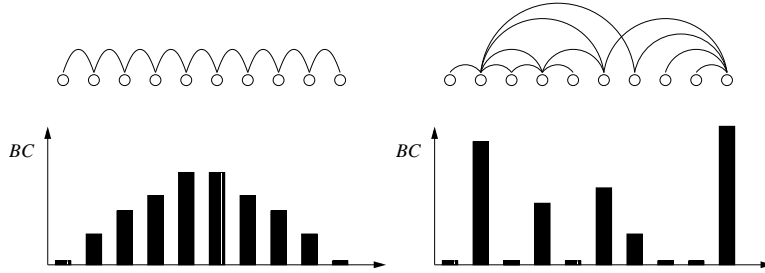

Fig. 13. (a) Betweenness centrality for the (one-dimensional) lattice case. The central nodes are close to the barycenter. (b) For a general graph, the central nodes are usually the ones with large degree.

a purely topological network with no underlying geography, this consideration does not apply anymore and the full randomness and the disordered small world structure are completely uncorrelated with the spatial position. It is worth remarking that the present argument applies in the absence of periodic boundary conditions that would destroy the geometrical ordering. This point is illustrated in Fig. 13 in the simple case of a one-dimensional lattice.

The present model defines an intermediate situation in that we have a random network with space constraints that introduces a local structure since short distance connections are favored. Shortcuts and long distance hops are present along with a spatial local structure that clusters spatially neighboring vertices. In Fig. 14 we plot the average distance $d(G, C)$ between the barycenter G and the 10 most central nodes. As expected, as spatial constraints become more important, the most central nodes get closer to the spatial barycenter of the network.
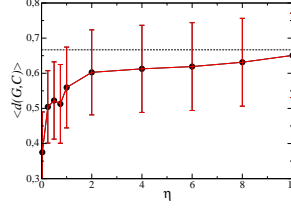
Fig. 14. Average Euclidean distance between the barycenter G of all nodes and the 10 most central nodes (C) versus the parameter η (Here $\delta = 0$, $N = 5,000$ and the results are averaged over 50 configurations). When space is important (ie. small η), the central nodes are closer to the gravity center. For large η, space is irrelevant and the average distance tends to the value corresponding to a uniform distribution $\langle r \rangle_{unif} = 2/3$ (dotted line).

Another effect observed when the spatial constraints become important are the large fluctuations of the BC. Fig. 15a displays the relative fluctuation

$$\delta g(k) = \frac{\sqrt{\langle \delta g^2(k) \rangle}}{\langle g(k) \rangle} \, , \qquad (16)$$

where $\langle \delta g^2(k) \rangle$ is the variance of the BC and $\langle g(k) \rangle$ its average (computed for each value of k). The value of η modifies the degree cut-off and in order to be able to compare the results for different values of η we rescale the abscissa by its maximum value k_{max}. This plot (Fig. 15a) clearly shows that the BC relative fluctuations increase as η decreases and become quite large. This means that nodes with small degree may have a relatively large BC (or the opposite), as observed in the air-transportation network (see Fig. 2 and [25]). In order to quantify these "anomalies" we compute the fluctuations of the betweenness centrality $\Delta_{RN}(k)$ for a randomized network with the same degree distribution than the original network and constructed with the Molloy-Reed algorithm [35]. We consider a node i as being "anomalous" if its betweenness centrality $g(i)$ lies outside the interval $[\langle g(k) \rangle - \alpha \Delta_{RN}(k), \langle g(k) \rangle + \alpha \Delta_{RN}(k)]$, where we choose $\alpha \simeq 1.952$ so that the considered interval would represent 95% of the nodes in the case of Gaussian distributed centralities around the average.

In Fig. 15b, we show the relative number of anomalies versus k/k_{kmax} for different values of η. This plot shows that the relative number of anomalies $N_a(k)/N(k)$ increases when the degree increases and more interestingly strongly increases when η decreases. Note that since for increasing k the number of nodes $N(k)$ is getting small, the results become more noisy.

The results of Figs. (12-15) can be summarized as follows. In a purely topological growing network, centrality is strongly correlated with degree since hubs have a natural ability to provide connections between otherwise separated regions or neighborhoods [14]. As spatial constraints appear and become more important, two factors compete in determining the most central nodes: (i) on the one hand hubs become even more important in terms of centrality since only a large traffic
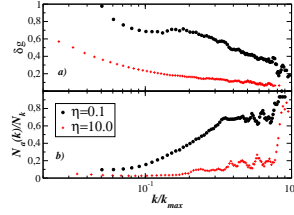
20   *Barthélemy, Barrat, Vespignani*

Fig. 15. (a) Relative fluctuations of the betweenness centrality versus k/k_{max} for two values of η ($N = 5,000$ and the results are averaged over 50 configurations and binned). The fluctuations increase when η decreases (i.e. when spatial constraints increase). (b) Number of anomalies $N_a(k)$ rescaled by the number $N(k)$ of nodes of degree k versus k/k_{max} for different values of η. The relative number of anomalies is larger when spatial constraints are large, especially for large k.

can compensate for the cost of long-range connections which implies that the correlations between degree and centrality become thus even stronger; (ii) on the other hand, many paths go through the neighborhood of the barycenter, reinforcing the centrality of less-connected nodes that happen to be in the right place; this yields larger fluctuations of g and a larger number of "anomalies".

We finally note that these effects are not qualitatively affected by the weight structure and we observe the same behavior for $\delta = 0$ or $\delta \neq 0$.

5. Conclusions

In this article we have reviewed the main empirical characteristics of the airline network. In particular, we have shown that weight-topology correlations as well as geographical effects are essential features of this system. We also have reviewed a simple model of growing weighted networks which introduces the effect of space and geometry in the establishment of new connections. When spatial constraints are important, the effects on the network structure can be summarized as follows:

- *Effect of spatial embedding on topology-traffic correlations*
  Spatial constraints induce strong nonlinear correlations between topology and traffic. The reason for this behavior is that spatial constraints favor the formation of regional hubs and reinforces locally the preferential attachment, leading for a given degree to a larger strength than the one observed without spatial constraints. Moreover, long-distance links can connect only to hubs, which yields a value $\beta_d > 1$ for small enough η. The existence of constraints such as spatial distance selection induces some strong correlations between topology (degree) and non-topological quantities such as weights or distances.
- *Effect of space embedding on centrality*
  Spatial constraints also induce large betweenness centrality fluctuations. While hubs are usually very central, when space is important central nodes tend to get closer to the gravity center of all points. Correlations between spatial position and centrality compete with the usual correlations between degree and centrality,

leading to the observed large fluctuations of centrality at fixed degree.

- *Effect of space embedding on clustering and assortativity*
  Spatial constraints implies that the tendency to connect to hubs is limited by the need to use small-range links. This explains the almost flat behavior observed for the assortativity. Connection costs also favor the formation of cliques between spatially close nodes and thus increase the clustering coefficient.

Including spatial effects in a simple model of weighted networks thus yields a large variety of behavior and interesting effects. This study sheds some light on the importance and effect of different ingredients such as the geographical distribution of nodes or the diversity of interaction weights in the structure of large complex networks. We believe that this attempt of a network typology could be useful in the understanding and modeling of real-world networks. We finally note that the distance is here a simple measure of the cost of a connection but that the model we proposed could be easily generalized to more complex cost functions that can be found in economic systems.

Acknowledgments

References

[1] R. Albert, H. Jeong, and A.-L. Barabási, Error and attack tolerance of complex networks, Nature **406**, 378 (2000).

[2] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. **74**, 47 (2002).

[3] E. Almaas, B. Kovacs, T. Viscek, Z.N. Oltvai, and A.-L. Barabási, Global organization of metabolic fluxes in the bacterium, Escherichia coli, Nature **427**, 839 (2004).

[4] L.A.N. Amaral, A. Scala, M. Barthélemy, and H.E. Stanley, Classes of small-world networks, Proc. Natl. Acad. Sci. (USA) **97**, 11149 (2000).

[5] C. Andersson, A. Hellervik, K. Lindgren, A spatial network explanation for a hierarchy of urban power laws, Physica A, **227**, 345 (2005).

[6] A.-L. Barabasi and R. Albert, Emergence of scaling in random networks, Science **286**, 509 (1999).

[7] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, The architecture of complex weighted networks, Proc. Natl. Acad. Sci. (USA) **101** 3747 (2004).

[8] A. Barrat, M. Barthélemy, and A. Vespignani, Weighted evolving networks: coupling topology and weights dynamics, Phys. Rev. Lett. **92** 228701 (2004).

[9] A. Barrat, M. Barthélemy, and A. Vespignani, Modeling the evolution of weighted networks, Phys. Rev. E **70** (2004) 066149.

[10] A. Barrat, M. Barthélemy, and A. Vespignani, Traffic-driven model of the World Wide Web graph, LNCS **3243** (2004) 56.

[11] A. Barrat and R. Pastor-Satorras, Rate equation approach for correlations in growing network models, Phys. Rev. E **71**, 036127 (2005).

[12] M. Barthélemy, B. Gondran, and E. Guichard, Spatial structure of the internet traffic, Physica A **319** 633 (2003).

[13] M. Barthélemy, Crossover from Scale-Free to Spatial Networks, Europhys. Lett. **63**, 915 (2003).

[14] M. Barthélemy, Betweenness Centrality in Large Complex Networks, Eur. Phys. J. B **38**, 163 (2003).

[15] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, Resilience of the Internet to random breakdowns, Phys. Rev. Lett. **85**, 4626 (2000).

[16] B. Derrida and H. Flyvbjerg, Statistical properties of randomly broken objects and of multivalley structures in disordered systems, J. Phys. A **20** 5273 (1987).

[17] S.N. Dorogovtsev, & J.F.F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW*, Oxford University Press, Oxford (2003).

[18] L. C. Freeman, A Set of Measures of Centrality Based Upon Betweeness, Sociometry **40**, 35 (1977) .

[19] D. Garlaschelli, S. Battiston, M. Castri, V.D.P. Servedio, and G. Caldarelli, The scale-free topology of market investments, Physica A **350**, 491 (2005).

[20] M.T. Gastner and M.E.J. Newman, The spatial structure of networks, preprint: cond-mat/0407680.

[21] M.T. Gastner and M.E.J. Newman, Shape and efficiency in spatial distribution networks, preprint: condmat/0409702.

[22] K.-I. Goh, B. Kahng, and D. Kim, Universal behavior of load distribution in scale-free networks, Phys. Rev. Lett. **87**, 278701 (2001).

[23] S.P. Gorman and R. Kulkarni, Spatial Small Worlds: New Geographic Patterns for an Information Economy, submitted to Environment and Planning Journal B, (2003).

[24] M. Granovetter, The Strength of Weak Ties, American Journal of Sociology **78** (6) 1360-1380 (1973).

[25] R. Guimerà and L.A.N. Amaral, Modeling the world-wide airport network, Eur. Phys. J. B **38**, 381 (2004).

[26] R. Guimera, S. Mossa, A. Turtschi, and L.A.N. Amaral, The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles, Proc. Natl. Acad. Sci. (USA) **102** 7794 (2005).

[27] A. Helmy, Small Large-Scale Wireless Networks: Mobility-Assisted Resource Discovery, preprint: cs.NI/0207069.

[28] International Air Transport Association, http://www.iata.org.

[29] A.E. Krause, K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor, Compartments revealed in food-web structure, Nature **426**, 282 (2003).

[30] A. Lakhina, J.B. Byers, M. Crovella, and I. Matta, On the geographic location of Internet resources, technical report, online version available at : http://www.cs.bu.edu/techreports/pdf/2002-015-internet-geography.pdf

[31] W. Li and X. Cai, Statistical Analysis of Airport Network of China, Phys. Rev. E **69**, 046106 (2004).

[32] C. Li and G. Chen, Network connection strength: Another power-law ?, Preprint cond-mat/0311333 (2003).

[33] S.S. Manna and P. Sen, Modulated Scale-free Network in the Euclidean Space, Phys. Rev. E **66**, 066114 (2002).

[34] N. Masuda, H. Miwa and N. Konno, Geographical threshold graphs with small-world and scale-free properties, Phys. Rev. E **71**, 036108 (2005).

[35] M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence, Random Struct. Algorithms **6**, 161 (1995).

[36] G. Mukherjee and S. S. Manna, Growing spatial scale-free graphs by selecting local

edges, preprint: cond-mat/0503697.

[37] G. Nemeth and G. Vattay, Giant Clusters in Random Ad Hoc Networks, Phys. Rev. E **67**, 036110 (2003).

[38] D.S. Callaway, M.E.J. Newman, S.H. Strogatz,, and D.J. Watts, Network robustness and fragility: percolation on random graphs, Phys. Rev. Lett. **85**, 5468 (2000).

[39] M. E. J. Newman, Who is the best connected scientist ? A study of scientific coauthorship network I, Phys. Rev. E **64**, 016131 (2001); M. E. J. Newman, Who is the best connected scientist ? A study of scientific coauthorship network II, Phys. Rev. E **64**, 016132 (2001).

[40] M.E.J. Newman, Assortative mixing in networks, Phys. Rev. Lett. **89**, 208701 (2002).

[41] R.V.R. Pandya, A note on "Weighted Evolving Networks: Coupling Topology and Weight Dynamics", cond-mat/0406644 (2004).

[42] R. Pastor-Satorras and A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. **86**, 3200 (2001).

[43] R. Pastor-Satorras and A. Vespignani, *Evolution and structure of the Internet: A statistical physics approach* (Cambridge University Press, Cambridge, 2004).

[44] A. Vázquez, R. Pastor-Satorras and A. Vespignani, Large-scale topological and dynamical properties of Internet, Phys. Rev. E **65**, 066130 (2002).

[45] D.J. Watts and S.H. Strogatz, COllective dynamics of "small-world" networks, Nature **393**, 440 (1998).

[46] B.M. Waxman, Routing of multipoint connections, IEEE *J. Select. Areas. Commun.* **6**, 1617 (1988).

[47] R. Xulvi-Brunet and I.M. Sokolov, Evolving networks with disadvantaged long-range connections, Phys. Rev. E **66**, 026118 (2002).

[48] S.H. Yook, H. Jeong, A.-L. Barabasi, and Y. Tu, Weighted Evolving Networks, Phys. Rev. Lett. **86**, 5835 (2001).

[49] S.-H. Yook, H. Jeong, and A.-L. Barabasi, Modeling the Internet's large-scale topology, Proc. Natl. Acad. Sci. (USA) **99**, 13382 (2002).

[50] D. Zheng, S. Trimper, B. Zheng and P.M. Hui,Weighted Scale-Free Networks with Stochastic Weight Assignments, Phys. Rev. E **67**, 040102 (2003).