

Weighted networks: empirical results and models

Marc Barthélemy¹, Alain Barrat², Alessandro Vespignani¹

*¹ School of Informatics and Biocomplexity Center, Indiana University,
Eigenmann Hall, 1900 East Tenth Street, Bloomington, IN 47406, USA*

*² Laboratoire de Physique Théorique (UMR du CNRS 8627), Batiment 210,
Université de Paris-Sud 91405 Orsay, France*

Abstract

We review the main tools which allow for the statistical characterization of weighted networks. We then present two case studies, the airline connection network and the scientific collaboration network, which are representative of critical infrastructures and social system, respectively. The main empirical results are (i) the broad distributions of various quantities and (ii) the existence of weight-topology correlations. These measurements show that weights are relevant and that in general the modeling of complex networks must go beyond topology. We propose a model which provides an explanation for the features observed in several real-world networks. This model of weighted network formation relies on the dynamical coupling between topology and weights, considering the rearrangement of weights when new links are introduced in the system. Finally, we discuss the effects of spatial constraints on the evolution of weighted networks.

Keywords: Complex Networks, Traffic.

1 Introduction

Networked structures arise in a wide array of different contexts such as technological and transportation infrastructures, social phenomena, and biological systems. These highly interconnected systems have recently been the focus of a great deal of attention that has uncovered and characterized their topological complexity [1–4]. Along with a complex topological structure, real networks display a large heterogeneity in the capacity and intensity of the connections—the weight of the links. For example, in ecology the diversity of the predator-prey interaction is believed to be a critical ingredient of ecosystems stability [5,6], and in social systems, the weight of interactions is very important in the characterization of the corresponding networks [7]. Similarly, the Internet traffic [3] or the number of passengers in the airline network [4,8,9] are crucial quantities in the study of these systems.

In this paper we review a set of metrics combining weighted and topological observables that allows to characterize the complex statistical properties of the strength of edges and vertices and to investigate the correlations among weighted quantities and the underlying topological structure. Specifically, we present results on the scientific collaboration network and the world-wide air-transportation network, which are representative examples of social and large infrastructure systems, respectively. The measures on weighted networks [9–12] have shown that they can exhibit additional complex properties such as broad distributions and non-trivial correlations of weights that do not find an explanation just in terms of the underlying topological structure. The heterogeneity in the intensity of connections may thus be very important in real-world systems and cannot be overlooked in their description. Motivated by these observations, we review also a model for weighted networks we have recently proposed [13], which naturally produces topology-weight correlations and broad distributions for various quantities. Finally, we present an analysis of the effect of spatial constraints on the evolution of weighted networks.

2 Tools for the characterization of weighted networks

We briefly review the different tools which allow for a first statistical characterization of weighted complex networks.

- *Weights*

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 $i, j = 1, \dots, N$, 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 $w_{ij} = w_{ji} \geq 0$.

- *Degree and weight distributions*

The standard topological characterization of networks is obtained by the analysis of the probability distribution $P(k)$ that a vertex has degree k . Complex networks often exhibit a power-law degree distribution $P(k) \sim k^{-\gamma}$ with $2 \leq \gamma \leq 3$. Similarly, a first characterization of weights is obtained by the distribution $P(w)$ that any given edge has weight w .

- *Weighted degree: Strength*

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 defined as [9]

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

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 degree and the importance of the weights of its links, and can be considered as the natural generalization of the degree. When the weights are independent from the topology, we obtain that the strength of the vertices of degree k is $s(k) \simeq \langle w \rangle k$ where $\langle w \rangle$ is the average weight. In the presence of correlations we obtain in general $s(k) \simeq A k^\beta$ with $\beta = 1$ and $A \neq \langle w \rangle$ or $\beta > 1$.

- *Weighted clustering*

The clustering coefficient c_i measures the local cohesiveness and is defined for any vertex i as the fraction of connected neighbors of i [14]. The average clustering coefficient $C = N^{-1} \sum_i c_i$ thus expresses the statistical level of cohesiveness measuring the global density of interconnected vertex triplets in the network. Further information can be gathered by inspecting the average clustering coefficient $C(k)$ restricted to the class of vertices with degree k . The topological clustering, however, does not take into account the fact that some neighbors are more important than others. In order to solve this incongruity we introduce a measure of the clustering that combines the topological information with the weight distribution of the network. The *weighted clustering coefficient* is defined as [9]

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

This quantity $c^w(i)$ counts for each triple formed in the neighborhood of the vertex i the weight of the two participating edges of the vertex i . In this way we are not just considering the number of closed triangles in the neighborhood of a vertex but also their total relative weight with respect to the vertex' strength. The factor $s_i(k_i - 1)$ is a normalization factor and

ensures that $0 \leq c_i^w \leq 1$. Consistently, the c_i^w definition recovers the topological clustering coefficient in the case that $w_{ij} = \text{const}$. 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. In the case of a large randomized network (lack of correlations) it is easy to see that $C^w = C$ and $C^w(k) = C(k)$. In real weighted networks, however, we can face two opposite cases. If $C^w > C$, we are in presence of a network in which the interconnected triples are more likely formed by the edges with larger weights. On the contrary, $C^w < C$ signals a network in which the topological clustering is generated by edges with low weight. In this case it is obvious that the clustering has a minor effect in the organization of the network since the largest part of the interactions (traffic, frequency of the relations, etc.) is occurring on edges not belonging to interconnected triples. The same may happen for $C^w(k)$, for which it is also possible to analyze the variations with respect to the degree class k .

- *Weighted assortativity: Affinity*

Another quantity used to probe the networks' architecture is the average degree of nearest neighbors of a vertex i

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in \mathcal{V}(i)} k_j, \quad (3)$$

where the sum runs on the nearest neighbors vertices of each vertex i . From this quantity a convenient measure to investigate the behavior of the degree correlation function is obtained by the average degree of the nearest neighbors, $k_{nn}(k)$, for vertices of degree k [15]. 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* [16]. 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.

In the case of weighted networks an appropriate characterization of the assortative behavior is obtained by the *weighted average nearest neighbors degree*, defined as

$$k_{nn,i}^w = \frac{1}{s_i} \sum_{j=1}^N a_{ij} w_{ij} k_j. \quad (4)$$

In this case, we perform a local weighted average of the nearest neighbor degree according to the normalized weight of the connecting edges, w_{ij}/s_i . This definition implies that $k_{nn,i}^w > k_{nn,i}$ if the edges with the larger weights are pointing to the neighbors with larger degree and $k_{nn,i}^w < k_{nn,i}$ in the

opposite case. The $k_{nn,i}^w$ function 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 $k_{nn,i}^w$ over all vertices with degree k), marks the weighted assortative or disassortative properties considering the actual interactions among the system's elements.

- *Disparity*

For a given node i with degree k_i and strength s_i different situations can arise. All weights w_{ij} can be of the same order s_i/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 [17,18]

$$Y_2(i) = \sum_{j \in \mathcal{V}(i)} \left[\frac{w_{ij}}{s_i} \right]^2 \quad (5)$$

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 [19] which showed that for these networks one can identify dominant reactions.

3 Empirical results

3.1 Weighted networks data

Prototypical examples of weighted networks can be found in the world-wide airport network (WAN) [8,9] and the scientific collaboration network (SCN) [20,21]. In the airport network each given weight w_{ij} is the number of available seats on direct flight connections between the airports i and j . For the WAN, we analyze the International Air Transportation Association (IATA) (www.iata.org) database containing the world list of airports pairs connected by direct flights and the number of available seats on any given connection for the year 2002. The resulting air-transportation graph comprises $N = 3880$ vertices denoting airports and $E = 18810$ edges accounting for the presence of a direct flight connection. The average degree of the network is $\langle k \rangle = 2E/N = 9.70$, while the maximal degree is 318.

In the SCN the nodes are identified with authors and the weight depends on the number of co-authored papers [20,9]. We consider the network of scientists who have authored manuscripts submitted to the e-print archive relative to condensed matter physics (xxx.lanl.gov/archive/cond-mat) between 1995 and 1998. Scientists are identified with nodes and an edge exists between two scientists if they have co-authored at least one paper. The resulting connected

network has $N = 12722$ nodes, with an average degree (i.e. average number of collaborators) $\langle k \rangle = 6.28$ and maximal degree 97. For the SCN we follow the definition of weight introduced in Ref. [20]: The intensity w_{ij} of the interaction between two collaborators i and j is defined as $w_{ij} = \sum_p \delta_i^p \delta_j^p / (n_p - 1)$ where the index p runs over all papers, n_p is the number of authors of the paper p , and δ_i^p is 1 if author i has contributed to paper p , and 0 otherwise. This definition seems to be rather objective and representative of the scientific interaction: It is large for collaborators having many papers in common but the contribution to the weight introduced by any given paper is inversely proportional to the number of authors.

3.2 Results

3.2.1 Topological properties

The topological properties of the SCN network and other similar networks of scientific collaborations have been studied in Ref. [20] and we report on Fig. (1A) the degree distribution showing a relatively broad law.

As shown in Fig. (1B), the topology of the WAN exhibits both small-world and scale-free properties as already observed in different dataset analyses [10,8]. In particular, the average shortest path length, measured as the average number of edges separating any two nodes in the network, shows the value $\langle \ell \rangle = 4.37$, very small compared to the network size N . The degree distribution, on the other hand, 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 that finds its origin in physical constraints on the maximum number of connections that a single airport can handle [4,8]. The airport connection graph is therefore a clear example of heterogeneous network showing scale-free properties on a definite range of degree values.

3.2.2 Strength distribution

The probability distribution $P(s)$ that a vertex has strength s is heavy tailed in both networks and the functional behavior exhibits similarities with the degree distribution $P(k)$ (see Fig. 1). A precise functional description of the heavy-tailed distributions may be very important in understanding the network evolution and will be deferred to future analysis. This behavior is not unexpected since it is plausible that the strength s_i increases with the vertex degree k_i , and thus the slow decaying tail of $P(s)$ stems directly from the very slow decay of the degree distribution.

3.2.3 Topology-weight correlations

In Fig. 2 we report the behavior obtained for both the real weighted networks and their randomized versions, generated by a random re-distribution

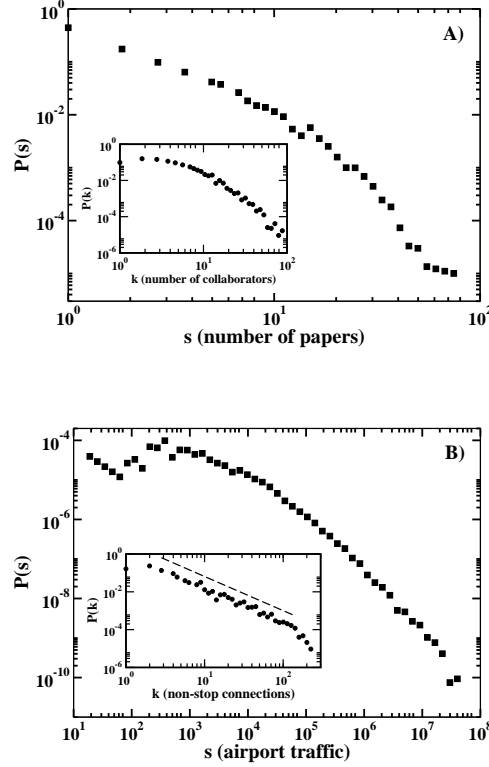


Fig. 1. **A** Degree and strength distribution in the scientific collaboration network. The degree k corresponds to the number of co-authors of each scientist and the strength represent its total number of publications. The distributions are heavy-tailed even if it is not possible to distinguish a definite functional form. **B** The same distributions for the world-wide airport network. The degree is the number of non-stop connections to other airports and the strength is the total number of passengers handled by any given airport. In this case, the degree distribution can be approximated by the power-law behavior $P(k) \sim k^{-\gamma}$ with $\gamma = 1.8 \pm 0.2$. The strength distribution has a heavy-tail extending over more than four orders of magnitude.

of the actual weights on the existing topology of the network. For the SCN the curves are very similar and well fitted by the uncorrelated approximation $s(k) = \langle w \rangle k$. Strikingly, this is not the case of the WAN. Fig. 2B clearly shows a very different behavior for the real data set and its randomized version. In particular, the power-law fit for the real data gives an “anomalous” exponent $\beta_{\text{WAN}} = 1.5 \pm 0.1$. This implies 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. This denotes a strong correlation between the weight and the topological properties in the WAN, where the larger is an airport, the more traffic it can handle.

The fingerprint of these correlations is also observed in the dependence of the weight w_{ij} on the degrees of the end point nodes k_i and k_j . For the WAN

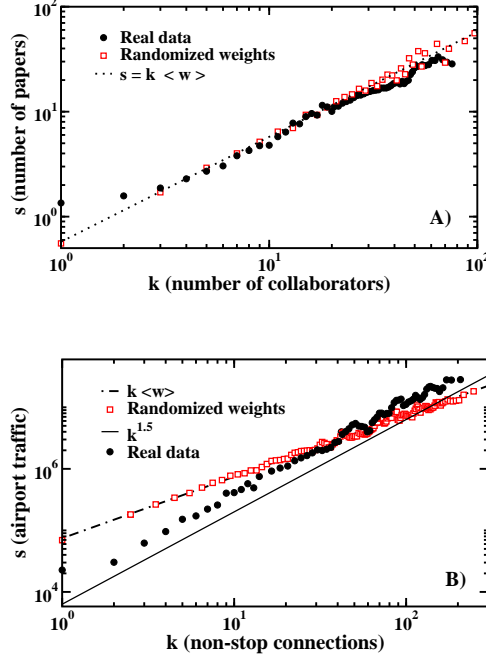


Fig. 2. Average strength $s(k)$ as function of the degree k of nodes. **A** In the scientific collaboration network the real data are very similar to those obtained in a randomized weighted network. Only at very large k values it is possible to observe a slight departure from the expected linear behavior. **B** In the world airport network real data follow a power-law behavior with exponent $\beta = 1.5 \pm 0.1$. This denotes anomalous correlations between the traffic handled by an airport and the number of its connections.

the behavior of the average weight as a function of the end points degrees can be well approximated by a power-law dependence $\langle w_{ij} \rangle \sim (k_i k_j)^\theta$ with an exponent $\theta = 0.5 \pm 0.1$. This exponent can be related to the β exponent by noticing that $s(k) \sim k(k k_j)^\theta$, resulting in $\beta = 1 + \theta$, if the topological correlations between the degree of connected vertices can be neglected. This is indeed the case of the WAN where the above scaling relation is well satisfied by the numerical values provided by the independent measurements of the exponents. In the SCN, instead, $\langle w_{ij} \rangle$ is almost constant for over two decades confirming a general lack of correlations between the weights and the vertices degree. In this case $\theta = 0$ and the relation $\beta = 1 + \theta$ also holds.

3.2.4 Weighted clustering and assortativity

We present the results [9] obtained for both the SCN (see Fig.3) and the WAN (see Fig.4) by comparing the regular topological quantities with the weighted ones introduced above. In the figures we report the relative difference of the values obtained for the topological and weighted quantities. It is striking to observe that for large degree values we observe up to 100% relative difference

signalling a strong difference in the clustering and correlation properties if we take into account the weighted nature of the networks.

- *SCN*

- (i) The measurements tell us that the SCN has a monotonously decaying spectrum $C(k)$. This implies that hubs present a much lower clustered neighborhood than low degree vertices which can be interpreted as the evidence that authors with few collaborators usually work within a well defined research group in which all the scientists collaborate together (high clustering). Authors with a large degree, however, collaborate with different groups and communities which on their turn do not have often collaborations, thus creating a lower clustering coefficient.
- (ii) The inspection of $C^w(k)$ shows generally that for $k \geq 10$ the weighted clustering coefficient is larger than the topological one. This implies that authors with many collaborators tend to publish more papers with interconnected groups of co-authors and is a signature of the fact that influential scientists form stable research groups where the largest part of their production is obtained.
- (iii) Furthermore, the SCN exhibits an assortative behavior in agreement with the general evidence that social networks are usually denoted by a strong assortative character [16]. Finally, the assortative properties find a clearcut confirmation in the weighted analysis with a $k_{nn}^w(k)$ strikingly growing as a power-law as a function of k .

- *WAN*

A different picture is found in the WAN, where the weighted analysis provides a richer and somehow different scenario.

- (i) This network also shows a decaying $C(k)$, consequence of the role of large airports that provide non-stop connections to very far destinations on an international and intercontinental scale. These destinations are usually not interconnected among them, giving rise to a low clustering coefficient for the hubs.
- (ii) We find, however, that $C^w/C \simeq 1.1$, indicating an accumulation of traffic on interconnected groups of vertices.
- (iii) The weighted clustering coefficient $C^w(k)$ has much more limited variation in the whole spectrum of k . This implies that high degree airports have a progressive tendency to form interconnected groups with high traffic links, thus balancing the reduced topological clustering. Since high traffic is associated to hubs, we have a network in which high degree nodes tend to form cliques with nodes with equal or higher degree, the so-called *rich-club phenomenon* [22].
- (iv) The topological $k_{nn}(k)$ does show an assortative behavior only at small degrees. For $k > 10$, $k_{nn}(k)$ approaches a constant value, a fact revealing an uncorrelated structure in which vertices with very different degrees have a very similar neighborhood. The analysis of the weighted $k_{nn}^w(k)$, however, exhibits a pronounced assortative behavior in the whole

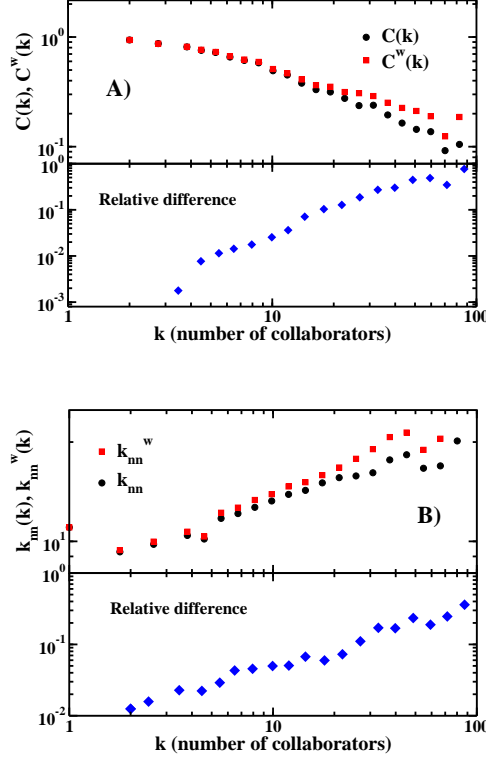


Fig. 3. Comparison of topological and weighted quantities for the SCN. **A** The weighted clustering separates from the topological one around $k \geq 10$. This marks a difference for authors with larger number of collaborators. **B** The assortative behavior is enhanced in the weighted definition of the average nearest neighbors degree. It is worth remarking the large relative difference (up to 50-100%) of the weighted and topological quantities (lower part of the figures).

k spectrum, providing a different picture in which high degree airports have a larger affinity for other large airports where the major part of the traffic is directed.

4 Modeling weighted networks

4.1 The need for weight-topology coupling

Previous approaches to the modeling of weighted networks focused on growing topologies where weights were assigned statically, i.e. once and for ever, with different rules related to the underlying topology [23,24]. 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 connection is created between two airports it

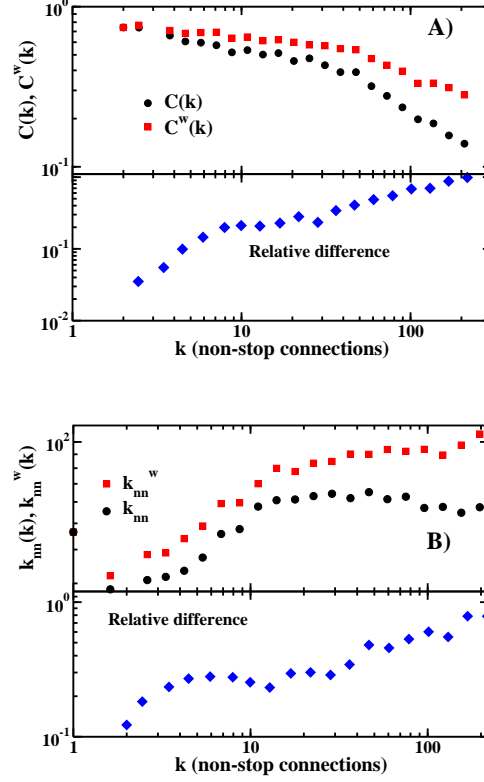


Fig. 4. Topological and weighted quantities for the WAN. **A** The weighted clustering coefficient is larger than the topological one in the whole degree spectrum. **B** $k_{nn}(k)$ is reaching a plateau for $k > 10$ denoting the absence of marked topological correlations. On the contrary $k_{nn}^w(k)$ exhibits a more definite assortative behavior. Also in this case the relative difference of the weighted and topological quantities are of the order of 100% (lower part of the figures).

will generally provoke a modification of the existing traffic of both airports. In general, it will increase the traffic activity depending on the specific nature of the network and on the local dynamics. In the following, we review a model that takes into account the coupled evolution in time of topology and weights. Instead of drawing randomly the weights, an alternative consists in coupling the evolution of the weights and of the topology and allowing the dynamical evolution of weights during the growth of the system. This mimics the evolution and reinforcements of interactions in natural and infrastructure networks.

The model dynamics 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

$$\Pi_{n \rightarrow i} = \frac{s_i}{\sum_j s_j}. \quad (6)$$

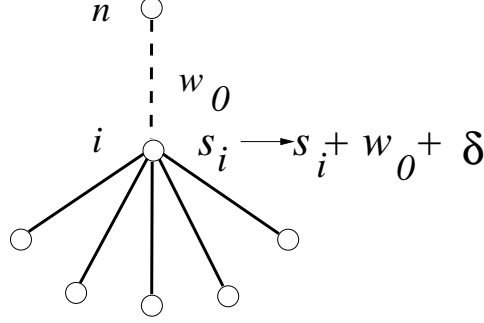


Fig. 5. Illustration of the construction rule. A new node n connects to a node i with probability proportional to $s_i / \sum_j s_j$. The weight of the new edge is w_0 and the total weight on the existing edges connected to i is modified by an amount equal to δ .

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. (6)) appears to be a plausible mechanism in many networks. In the Internet new routers connect to more central routers in terms of bandwidth and traffic handling capabilities and in the airport networks new connections are generally established to airports with a large passenger traffic. Even in the SCN this mechanism might play a role since an author with more co-authored papers is more visible and open to further collaborations.

The presence of the new edge (n, i) will introduce 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

$$w_{ij} \rightarrow w_{ij} + \Delta w_{ij}, \quad (7)$$

where

$$\Delta w_{ij} = \delta \frac{w_{ij}}{s_i}. \quad (8)$$

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. 5), yielding $s_i \rightarrow s_i + \delta + w_0$. At this stage, it is worth remarking that while we will focus on the simplest model with $\delta = \text{const}$, different choices of Δw_{ij} with heterogeneous δ_i or depending on the specific properties of each vertex (w_{ij}, k_i, s_i) can be considered [25,26,29]. Finally, after the weights have been updated the growth process is iterated by introducing a new vertex with the corresponding re-arrangement of weights.

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. According to the value of δ , different scenarios are possible. If the induced weight is $\delta \approx 1$ we mimic situations in which an appreciable fraction of traffic generated by the new connection will be dispatched in the already existing connections. This is plausible in the airport networks where the transit traffic is rather relevant in hubs. In the case of $\delta < 1$ we face situations such as the SCN where it is reasonable to consider that the birth of a new collaboration (co-authorship) is not triggering a more intense activity on previous collaborations. 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 $s_i(t)$ and $k_i(t)$ of the i -th vertex at time t , and by relying on the continuous approximation that treats k , s and the time t as continuous variables [1,2,13]. The behavior of the strength and the degree are easily obtained and one has in the long time limit

$$s_i(t) \simeq (2\delta + 1)k_i(t) \quad (9)$$

This proportionality relation $s \sim k$ implies $\beta = 1$ but the prefactor is different from $\langle w \rangle$ which indicates the existence of correlations between topology and weights. This relation (9) is particularly relevant since it states that the weight-driven dynamics generates in Eq. (6) 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

$$\gamma = \frac{4\delta + 3}{2\delta + 1}. \quad (10)$$

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 [30] and the value $\gamma = 3$ is recovered. For larger values of δ the distribution is progressively broader with $\gamma \rightarrow 2$ when $\delta \rightarrow \infty$. This indicates that the weight-driven growth generates scale-free networks with exponents varying in the range of values usually observed in the empirical analysis of networked structures [1–3]. Noticeably the exponents are non-universal and depend only on the parameter δ governing the microscopic dynamics of weights. The model therefore proposes a general mechanism for the occurrence of varying power-law behaviors without resorting on more complicate topological rules and variations of the basic preferential attachment mechanism.

Similarly to the previous quantities, it is possible to obtain analytical expressions for the evolution of weights and the relative statistical distribution [13]. The probability distribution $P(w)$ is in this case also a power-law $P(w) \sim w^{-\alpha}$ where $\alpha = 2 + \frac{1}{\delta}$. The exponent α has large variations as a function of the parameter δ and $P(w)$ moves from a delta function for $\delta = 0$ to a very slow decaying power-law with $\alpha = 2$ if $\delta \rightarrow \infty$. This feature clearly shows that the weight distribution is extremely sensible to changes in the microscopic dynamics ruling the network's growth.

In summary, the networks generated by the model display power-law behavior for the weight, degree and strength distributions with non-trivial exponents depending on the unique parameter defining the model's dynamics. These results suggest that the inclusion of weights in networks modeling naturally explains the diversity of scale-free behavior empirically observed in real networked structures. Strikingly, the weight-driven growth recovers an effective preferential attachment for the topological properties, providing a microscopic explanation for the ubiquitous presence of this mechanism.

4.2 *Effect of spatial constraints*

Another important aspect lies in the embedding of some networks in real space. For instance, most people have their friends and relatives in their neighborhood, transportation networks depend on distance, and many communication networks devices have short radio range [33–37]. A particularly important example of such a “spatial” network is the Internet which is a set of routers linked by physical cables with different lengths and latency times [38,39]. More generally, many networks can be seen as embedded in a given parameter space: each node carries a certain value of the parameter, and links are established with a certain cost depending on these values. One can for example think of social dimensions and social distances measured by salary, socio-professional category differences. 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). These spatial constraints can have important consequences on the topology of the resulting network [40].

So far, most models and simple mechanisms have focused on the strong heterogeneity of weights and topology, as well as the existence of some level of correlations between these aspects of complex networks. Not all features observed in real networks are however reproduced: in particular, the “traffic” as measured by the strength of a vertex grows only linearly with its degree,

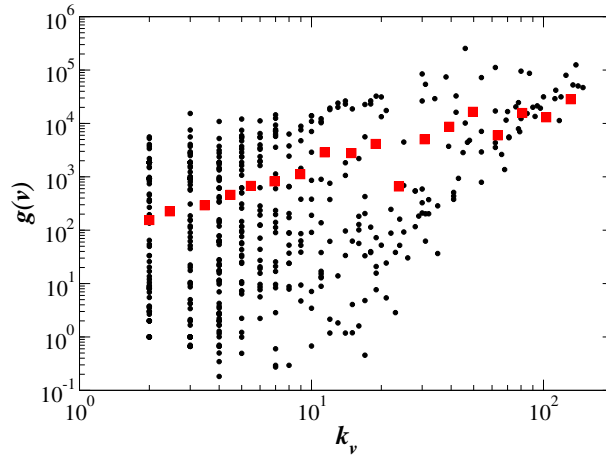


Fig. 6. Scatter-plot of the betweenness centrality versus degree for nodes of the North-American air-transportation network. The red symbols correspond to the average BC versus degree.

(except if non-linearities are introduced by hand [25]) while a non-linear behavior is observed in real-world [9]. The centrality fluctuations can also become anomalously large as it was observed in the airport network [8] (figure 6). One may therefore wonder if these interesting and non-trivial features can result from the combination of weight, topology and space. To this end, we define a simple model of growing network combining these ingredients which can be seen as an investigating tool. This simple model is obtained as the embedding of the weighted growing network introduced above in a two-dimensional 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. We thus start with an initial seed of N_0 vertices randomly (with uniform distribution) located on a 2d disk (of radius L) and connected by links with assigned weight w_0 . At each time step, a new vertex n is randomly placed on the disk. 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

$$\Pi_{n \rightarrow i} = \frac{s_i e^{-d_{ni}/r_c}}{\sum_j s_j e^{-d_{nj}/r_c}}, \quad (11)$$

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. We then update the weight according to the same rule as for the model without spatial effect $w_{ij} \rightarrow w_{ij} + \delta \frac{w_{ij}}{s_i}$.

The model thus 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 . When $\eta \gg 1$ space is irrelevant and in the opposite case a small value $\eta \ll 1$ indicates that long links are costly. When spatial constraints exist, the network properties are modified and the effects are summarized in the following.

- *Effect of spatial embedding on topology-traffic correlations*

Spatial constraints induce strong nonlinear correlations ($\beta > 1$) between topology and traffic as measured by the degree and the strength, respectively. 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. 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 [28], leading to the observed large fluctuations of centrality at fixed degree. Generally speaking when spatial constraints exist central nodes are correlated with barycentric considerations. This phenomenon can be understood with the following argument. For a spatial network—the extreme case being a lattice—the shortest path between two nodes is simply the Euclidean geodesic. If the two endpoints are far away, the probability that the shortest path passes near the barycenter of all nodes is very high. In other words, this implies that the barycenter (and its neighbors) will have a large centrality. This point is illustrated in Fig. 7 in the simple case of a one-dimensional lattice. This general consideration applies also in the case of more complex networks such as the spatial network considered here and indeed as shown in Figure 8 the average distance $\langle d(G, C) \rangle$ between the barycenter G and the most central nodes decreases when spatial constraints become important. As expected, as spatial constraints become more important, the most central nodes get closer to the barycenter of the network.

- *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 leads to an almost flat behavior for the assortativity. Connection costs also favor the formation of cliques between spatially close nodes and thus increase the clustering coefficient.

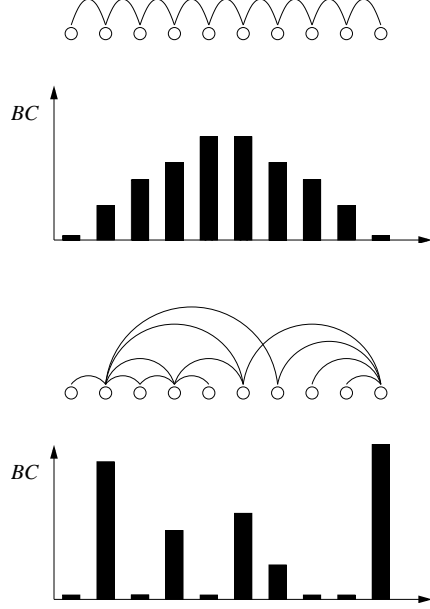


Fig. 7. **A** Betweenness centrality for the (one-dimensional) lattice case. The central nodes are close to the barycenter. **B** For a general graph, the central nodes are usually the ones with largest degree.

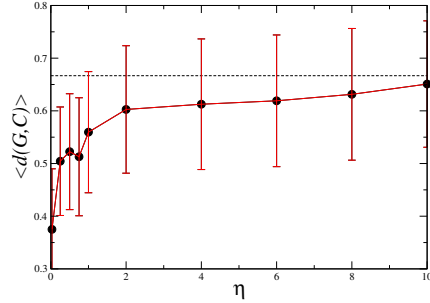


Fig. 8. Average distance between the barycenter G of all nodes and the 10 most central nodes (C) versus the parameter η (Here $\delta = 0$, $N = 5,000$ and the results are averaged over 50 configurations). When space is important (ie. small η), the central nodes are closer to the gravity center. For large η , space is irrelevant and the average distance tends to its value computed for a uniform distribution $\langle r \rangle_{unif} = 2/3$ (dotted line).

5 Conclusions and perspectives

A more complete view of complex networks is thus provided by the study of the interactions defining the links of these systems. The weights characterizing the various connections exhibit complex statistical features with highly varying distributions and power-law behavior. In particular we have considered the specific examples of the scientific collaboration and world-wide airport networks where it is possible to appreciate the importance of the correlations between weights and topology in the characterization of real networks prop-

erties. Indeed, the analysis of the weighted quantities and the study of the correlations between weights and topology provide a complementary perspective on the structural organization of the network that might be undetected by quantities based only on topological information. The weighted quantities thus offer a quantitative and general approach to understanding the complex architecture of real weighted networks. The empirical results shows that purely topological models are inadequate and that there is a need for models going beyond pure topology. The model we have presented is possibly the simplest one in the class of weight-driven growing networks. A novel feature in the model is the weight dynamical evolution occurring when new vertices and edges are introduced in the system. This simple mechanism produces a wide variety of complex and scale-free behavior depending on the physical parameter δ that controls the local microscopic dynamics. While a constant parameter δ is enough to produce a wealth of interesting network properties, a natural generalization of the model consists in considering δ as a function of the vertices degree or strength. Similarly, more complicated variations of the microscopic rules may be implemented to mimic in a detailed fashion particular networked systems [25,26,29,31]. In particular, we have shown that including spatial effects in this simple model of weighted networks leads to a large variety of behavior and interesting effects such as the appearance of large centrality fluctuations. This study sheds some light on the importance and effect of different ingredients such as spatial embedding or diversity of interaction weights in the structure of large complex networks and we believe that this attempt of a network typology could be useful in the understanding and modeling of real-world networks.

Acknowledgments. M.B. is on leave of absence from the CEA-Centre d'Etudes de Bruyères-Le-Chatel. We thank IATA for making the airline commercial flight database available to us. We also thank M.E.J. Newman for conceding us the possibility of using the scientific collaboration network data (see [32]). A.B. and A.V. are partially funded by the European Commission - Fet Open Project COSIN IST-2001-33555.

References

- [1] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [2] S.N. Dorogovtsev, & J.F.F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW*, Oxford University Press, Oxford (2003).
- [3] R. Pastor-Satorras and A. Vespignani, *Evolution and structure of the Internet: A statistical physics approach* (Cambridge University Press, Cambridge, 2004).
- [4] L.A.N. Amaral, A. Scala, M. Barthélemy, and H.E. Stanley, *Proc. Natl. Acad. Sci. (USA)* **97**, 11149 (2000).
- [5] S.L. Pimm, *Food Webs*, The University of Chicago Press (2nd edition, 2002).
- [6] A.E. Krause, K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor, *Nature* **426**, 282 (2003).
- [7] M. Granovetter, *American Journal of Sociology*, **78** (6) 1360-1380 (1973).

- [8] R. Guimerà , S. Mossa, A. Turtshi, and L.A.N. Amaral, Preprint cond-mat/0312535 (2003); R. Guimerà and L.A.N. Amaral, Eur. Phys. J. **38** 381 (2004).
- [9] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. (USA) **101** 3747 (2004).
- [10] C. Li and X. Cai, Preprint cond-mat/0309236 (2003).
- [11] C. Li and G. Chen, Preprint cond-mat/0311333 (2003).
- [12] D. Garlaschelli, S. Battiston, M. Castri, V.D.P. Servedio, and G. Caldarelli, Preprint cond-mat/0310503 (2003).
- [13] A. Barrat, M. Barthélemy, and A. Vespignani, Phys. Rev. Lett. **92** 228701 (2004).
- [14] D.J. Watts and S.H. Strogatz, Nature **393**, 440 (1998).
- [15] R. Pastor-Satorras, A. Vázquez and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
- [16] M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
- [17] B. Derrida and H. Flyvbjerg, J. Phys. A **20** 5273 (1987).
- [18] M. Barthélemy, B. Gondran, and E. Guichard, Physica A **319** 633 (2003).
- [19] E. Almaas, B. Kovacs, T. Vicsek, Z.N. Oltvai, and A.-L. Barabási, Nature **427**, 839 (2004).
- [20] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001); M. E. J. Newman, Phys. Rev. E **64**, 016132 (2001).
- [21] A.-L. Barabasi, H. Jeong, R. Ravasz, Z. Neda, T. Vicsek, and A. Schubert, Physica A **311**, 590 (2002).
- [22] S. Zhou & R.J. Mondragon, e-print cs.NI0303028 (2003).
- [23] S.H. Yook, H. Jeong, A.-L. Barabasi, and Y. Tu, Phys. Rev. Lett. **86**, 5835 (2001).
- [24] D. Zheng, S. Trimper, B. Zheng and P.M. Hui, Phys. Rev. E **67**, 040102 (2003).
- [25] A. Barrat, M. Barthélemy, and A. Vespignani, LNCS **3243** (2004) 56.
- [26] A. Barrat, M. Barthélemy, and A. Vespignani, Phys. Rev. E (in press, 2005). cond-mat/0406238.
- [27] M. Barthélemy, A. Barrat, and A. Vespignani, JSTAT in press (2005).
- [28] M. Barthélemy, Eur. Phys. J. B **38**, 163 (2003).
- [29] R.V.R. Pandya. Preprint cond-mat/0406644 (2004).
- [30] A.-L. Barabasi and R. Albert, Science **286**, 509 (1999).
- [31] Bo Hu, Xin-Yu Jiang, Jun-Feng Ding, Yan-Bo Xie, Bing-Hong Wang. Preprint cond-mat/0408125 (2004).
- [32] <http://www-personal.umich.edu/~mejn/collaboration/>
- [33] A. Helmy, online archive: cs.NI/0207069.
- [34] G. Nemeth and G. Vattay, cond-mat/0211325.
- [35] S.P. Gorman and R. Kulkarni, submitted to Environment and Planning Journal B, (2003).
- [36] M.T. Gastner and M.E.J. Newman, condmat/0407680
- [37] M.T. Gastner and M.E.J. Newman, condmat/0409702

- [38] A. Lakhina, J.B. Byers, M. Crovella, and I. Matta, technical report, online version available at : <http://www.cs.bu.edu/techreports/pdf/2002-015-internet-geography.pdf>
- [39] S.-H. Yook, H. Jeong, and A.-L. Barabasi, Proc. Natl. Acad. Sci. (USA) **99**, 13382 (2002).
- [40] M. Barthélemy, Europhys. Lett. **63**, 915 (2003).