Available online at www.sciencedirect.com
SCIENCE DIRECT®

Physica A 346 (2005) 34–43

PHYSICA A

www.elsevier.com/locate/physa

Characterization and modeling of weighted networks

Marc Barthélemy[a,*], Alain Barrat[b], Romualdo Pastor-Satorras[c], Alessandro Vespignani[b]

[a]CEA-Centre d'Etudes de Bruyères-le-Châtel, Département de Physique Théorique et Appliquée, BP12, 91680 Bruyères-Le-Châtel, France
[b]Laboratoire de Physique Théorique (UMR du CNRS 8627), Batiment 210, Université de Paris-Sud, 91405 Orsay, France
[c]Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, Mòdul B4, 08034 Barcelona, Spain

Available online 17 September 2004

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 representatives of critical infrastructure 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 review 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 new links are introduced in the system.

PACS: 89.75. – k; –87.23.Ge; 05.40. – a

Keywords: Disordered system; Networks

*Corresponding author.
E-mail address: Marc.Barthelemy@th.u-psud.fr (M. Barthélemy).

0378-4371/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.physa.2004.08.047

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 link. In ecology, the diversity of the predator–prey interaction is believed to be a critical ingredient of ecosystems stability [5] and in social systems, the weight of interactions is very important in the characterization of the corresponding networks [6]. Similarly, the Internet traffic [3] or the number of passengers in the airline network [4,7,8] are crucial quantities in the study of these systems.

We first review here the appropriate metrics which combine weighted and topological observables that enable one to characterize the complex statistical properties of weighted networks. Specifically, we present results on the scientific collaboration network and the worldwide air transportation network [8] which show that weights cannot be overlooked in the description of these systems. Motivated by these observations, we review a model for weighted networks that we have recently proposed in Ref. [9] and which naturally produces topology-weight correlations and broad distributions.

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 $w_{ij} = w_{ji} \geqslant 0$.

Connectivity 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 $2 \leqslant \gamma \leqslant 3$. Similarly, a first characterization of weights is obtained by the distribution $P(w)$ that any given edge has weight w.

Weighted connectivity: strength: Along with the degree of a node, a very significative measure of the network properties in terms of the actual weights is obtained by looking at the vertex *strength* s_i defined as [8,20,28]

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

where the sum runs over the set $\mathscr{V}(i)$ of neighbors of i. The strength of a node integrates the information both with 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 $s \simeq \langle w \rangle k$ where $\langle w \rangle$ is the average weight. In the presence of correlations we obtain in general $s \simeq Ak^{\beta}$ with $\beta = 1$ and $A \neq \langle w \rangle$ or $\beta > 1$.

Weighted clustering: The topological clustering [10] does not take into account the fact that some neighbors are more important than others. We thus have to introduce a measure of clustering that combines the topological information with the weight distribution of the network. The *weighted clustering coefficient* is defined as [8]

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

This quantity $c^{w}(i)$ is the 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 $s_i(k_i - 1)$ ensures that $0 \leqslant c_i^{w} \leqslant 1$ and that c_i^{w} 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. The ratio C^{w}/C (and similarly $C^{w}(k)/C(k)$ which allows an analysis with respect to the degree k) indicates if the interconnected triples are more likely formed by the edges with larger weights.

Weighted assortativity: affinity: Along with the weighted clustering coefficient, we introduce the *weighted average nearest neighbors degree*, defined as

$$k_{nn,i}^{w} = \frac{1}{s_i} \sum_{j=1}^{N} a_{ij} w_{ij} k_j \ . \tag{3}$$

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. $k_{nn,i}^{w}$ 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 [11] considering the actual interactions among the system's elements.

Disparity: For a given node i with connectivity 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 [12,13];

$$Y_2(i) = \sum_{j \in \mathscr{V}(i)} \left[\frac{w_{ij}}{s_i} \right]^2 \ . \tag{4}$$

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 [14] 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 worldwide airport network (WAN), [7,8] and the scientific collaboration network (SCN) [15,16]. In the airport network each given weight w_{ij} is the number of available seats on direct flights connections between the airports i and j and for the SCN the nodes are identified with authors and we follow the definition of weight introduced in Ref. [15]: The intensity w_{ij} of the interaction between two collaborators i and j is defined as $w_{ij} = \sum \delta_i^p \delta_j^p/(n_p - 1)$, 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. For the WAN, we analyze the International Air Transportation Association (IATA)[1] database for the year 2002 and for the SCN we consider the network of scientists who have authored manuscripts submitted to the e-print archive relative to condensed matter physics (http://xxx.lanl.gov/archive/cond-mat) between 1995 and 1998.

3.2. Empirical 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. [15] and we report on Fig. 1(A) the connectivity distribution showing a relatively broad law. As shown in Fig. 1(B), the topology of the WAN exhibits both small-world and scale-free properties as already observed in different dataset analyses [7,17]. In particular, the average shortest path length shows the value $\langle \ell \rangle = 4.37$, very small compared to the network size $N \sim 10^4$. 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 [4,7].

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). 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 from the 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 redistribution of the actual

[1] http://www.iata.org.

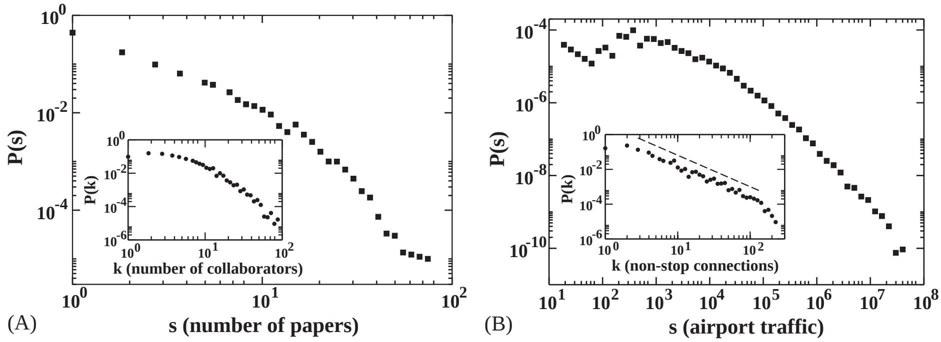

Fig. 1. (A) Degree and strength distribution in the scientific collaboration network. The degree k corresponds to the number of co-authors of each scientist and the strength represent its total number of publications. The distributions are heavy tailed. (B) The same distributions for the worldwide airport 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. A fit gives $P(k) \sim k^{-\gamma}$ with $\gamma = 1.8 \pm 0.2$. The strength distribution has a heavy tail extending over more than four orders of magnitude.

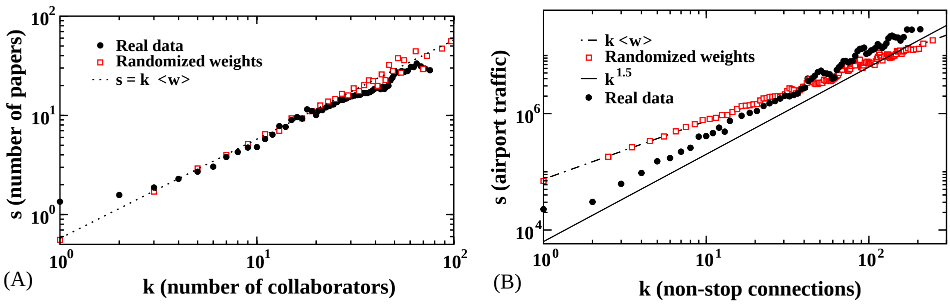

Fig. 2. Average strength $s(k)$ as function of the degree k of nodes. (A) In the scientific collaboration network, the real data are very similar to those obtained in a randomized weighted network. Only at very large k values it is possible to observe a slight departure from the expected linear behavior. (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.

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 and Fig. 2(B) 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 $\beta_{\mathrm{WAN}} = 1.5 \pm 0.1$. This implies that the strength of vertices grows faster than their degree, i.e., the weight of edges belonging to highly connected vertices tend 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 [8] 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 $\theta = 0.5 \pm 0.1$. In the SCN, instead, $\langle w_{ij}\rangle$ is almost constant leading to the value $\theta \simeq 0$ and confirming in this case a general lack of correlations between the weights and the vertices degree.

3.2.4. Weighted clustering and assortativity

We present the results [8] obtained for both the SCN and the WAN by comparing the regular topological quantities with the weighted ones introduced above.

3.2.4.1. Case of the SCN. We find that (i) the 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 collaborate with different groups and communities, which on their turn, do not have often collaborations (creating a lower clustering coefficient). (ii) 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. Finally, (iii) The SCN exhibits an assortative behavior in agreement with the general evidence that social networks are usually denoted by a strong assortative character [11].

3.2.4.2. Case of the WAN. A different picture is found in the WAN, where the weighted analysis provides a richer and somehow different scenario. (i) Large airports provide stop connections to very far destinations on an international and inter-continental scale. These destinations are usually not interconnected among them, giving rise to a low-clustering coefficient for the hubs. (ii) The traffic is accumulating on interconnected groups of vertices. (iii) High-degree airports have a progressive tendency to form interconnected groups with high traffic links. 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* [19]. (iv) The analysis of the weighted $k_{nn}^w(k)$ shows that high degree airports have a larger affinity for other large airports where the major part of the traffic is directed.

4. Modeling weighted networks

Previous approaches to the modeling of weighted networks focused on growing topologies where weights were assigned statically, i.e., once for ever, with different rules related to the underlying topology [20,21] (See also [29,30] for other recent models.) 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 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 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,

$$\Pi_{n\to i} = \frac{s_i}{\sum_j s_j} \, . \tag{5}$$

This rule of "busy get busier" relaxes the usual degree-preferential attachment [25], 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. (5)) appears to be a plausible mechanism in many networks [9].

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

$$w_{ij} \to w_{ij} + \Delta w_{ij}, \qquad \text{with} \quad \Delta w_{ij} = \delta \frac{w_{ij}}{s_i} \, . \tag{6}$$

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. 3), 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 [22–24] 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.

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.

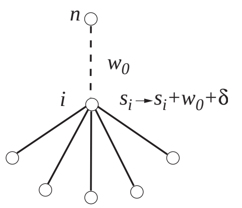

Fig. 3. Illustration of the construction rule. A new node n connects to a node i with probability proportional to $s_i / \sum_j s_j$. The weight of the new edge is w_0 and the total weight on the existing edges connected to i is modified by an amount equal to δ.

The 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 ith vertex at time t, and by relying on the continuous approximation that treats k, s and the time t as continuous variables [1,2,8]. One obtains $s_i(t) = (2\delta + 1)k_i(t)$ which implies $\beta = 1$ and a prefactor different from $\langle w \rangle$ which indicates the existence of correlations between topology and weights. The fact that $s \propto k$ is also particularly relevant since it states that the weight-driven dynamics generates in Eq. (5); 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

$$\gamma = \frac{4\delta + 3}{2\delta + 1} . \tag{7}$$

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 does not affect the existing weights ($\delta = 0$), the model is topologically equivalent to the Barabasi–Albert model [25] and the value $\gamma = 3$ is recovered. It is also possible to show analytically [9] 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 sensitive to changes in the microscopic dynamics ruling the network's growth.

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 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 empirical results—broad distributions and topology-weight correlations—show that purely topological models are inadequate and that there is a need for a model which goes 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 [22–24,31]. In particular, space should be included in the description of the airline network [26]. Finally, in this perspective the present model appears as a general starting point for the realistic modeling of complex weighted networks [18].

Acknowledgements

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 [27]). A.B., R.P.-S. and A.V. are partially funded by the European Commission—Fet Open Project COSIN IST-2001-33555. R.P.-S. acknowledges financial support from the Ministerio de Ciencia y Tecnología (Spain) and from the Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (Spain).

References

[1] R. Albert, A.-L. Barabási, Rev. Mod. Phys. 74 (2002) 47.
[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, 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, H.E. Stanley, Proc. Natl. Acad. Sci. USA 97 (2000) 11149.
[5] A.E. Krause, K.A. Frank, D.M. Mason, R.E. Ulanowicz, W.W. Taylor, Nature 426 (2003) 282.
[6] M. Granovetter, Am. J. Sociol. 78 (6) (1973) 1360–1380.
[7] R. Guimera, S. Mossa, A. Turtschi, L.A.N. Amaral, Preprint cond-mat0312535, 2003;
R. Guimerà, L.A.N. Amaral, Eur. Phys. J. 38 (2004) 381.
[8] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, Proc. Natl. Acad. Sci. USA 101 (2004) 3747.
[9] A. Barrat, M. Barthélemy, A. Vespignani, Phys. Rev. Lett. 92 (2004) 228701.
[10] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440.
[11] M.E.J. Newman, Phys. Rev. Lett. 89 (2002) 208701.
[12] B. Derrida, H. Flyvbjerg, J. Phys. A 20 (1987) 5273.
[13] M. Barthélemy, B. Gondran, E. Guichard, Physica A 319 (2003) 633.
[14] E. Almaas, B. Kovacs, T. Viscek, Z.N. Oltvai, A.-L. Barabási, Nature 427 (2004) 839.
[15] M.E.J. Newman, Phys. Rev. E 64 (2001) 016131;
M.E.J. Newman, Phys. Rev. E 64 (2001) 016132.
[16] A.-L. Barabasi, H. Jeong, R. Ravasz, Z. Neda, T. Vicsek, A. Schubert, Physica A 311 (2002) 590.
[17] W. Li, X. Cai, Phys. Rev. E 69 (2004) 046106.
C. Li, G. Chen, Preprint cond-mat/0311333, 2003.
[18] D. Garlaschelli, S. Battiston, M. Castri, V.D.P. Servedio, G. Caldarelli, Preprint cond-mat/0310503, 2003.
[19] S. Zhou, R.J. Mondragon, e-print cs.NI0303028, 2003.
[20] S.H. Yook, H. Jeong, A.-L. Barabasi, Y. Tu, Phys. Rev. Lett. 86 (2001) 5835.
[21] D. Zheng, S. Trimper, B. Zheng, P.M. Hui, Phys. Rev. E 67 (2003) 040102.
[22] A. Barrat, M. Barthélemy, A. Vespignani, Preprint cs.NI/0405070, 2004.
[23] A. Barrat, M. Barthélemy, A. Vespignani, submitted, Preprint condmat/0406238, 2004.
[24] R.V.R. Pandya, cond-mat/0406644, 2004.
[25] A.-L. Barabasi, R. Albert, Science 286 (1999) 509.

[26] M. Barthélemy, A. Barrat, A. Vespignani, in preparation (2004).
[27] http://www-personal.umich.edu/~mejn/collaboration/.
[28] J.P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A. Kanto, Phys. Rev. E 68 (2003) 056110.
[29] T. Antal, P.L. Krapisky, preprint cond-mat/0408285.
[30] S.N. Dorogovtsev, J.F.F. Mendes, preprint cond-mat/0408343.
[31] B. Hu, X.-Y. Jiang, J.-F. Ding, Y.-B. Xie, B.-H. Wang, Preprint cond-mat/0408125 (2004).