Eur. Phys. J. B **50**, 249–254 (2006)
DOI: 10.1140/epjb/e2006-00057-5

THE EUROPEAN PHYSICAL JOURNAL B

Modeling the Internet

M. Ángeles Serrano[1,a], M. Boguñá[2], and A. Díaz-Guilera[2]

[1] School of Informatics, Indiana University, Eigenmann Hall, 1900 East Tenth Street, Bloomington, IN 47406, USA
[2] Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

Received 11 October 2005 / Received in final form 11 November 2005
Published online 16 February 2006 –

Abstract. We model the Internet as a network of interconnected Autonomous Systems which self-organize under an absolute lack of centralized control. Our aim is to capture how the Internet evolves by reproducing the assembly that has led to its actual structure and, to this end, we propose a growing weighted network model driven by competition for resources and adaptation to maintain functionality in a demand and supply balance. On the demand side, we consider the environment, a pool of users which need to transfer information and ask for service. On the supply side, ASs compete to gain users, but to be able to provide service efficiently, they must adapt their bandwidth as a function of their size. Hence, the Internet is not modeled as an isolated system but the environment, in the form of a pool of users, is also a fundamental part which must be taken into account. ASs compete for users and big and small come up, so that not all ASs are identical. New connections between ASs are made or old ones are reinforced according to the adaptation needs. Thus, the evolution of the Internet can not be fully understood if just described as a technological isolated system. A socio-economic perspective must also be considered.

PACS. 89.20.Hh World Wide Web, Internet – 05.70.Ln Nonequilibrium and irreversible thermodynamics – 87.23.Ge Dynamics of social systems – 89.75.Hc Networks and genealogical trees

1 Introduction

In an attempt to bring nearer theory and reality, many researchers working on complex networks [1] have recently shifted focus from unweighted graphs to weighted networks. Weight is just one of the relevant ingredients. Others come from the fact that real systems are not static but evolve. As broadly recognized, growth and preferential attachment are also key issues at the core of a set of recent network models focusing on evolution under an statistical physics approach [2–7]. These models have been able to approximate some topological features observed in many real networks –specifically the small-world property or a power-law degree distribution. Although a great step forward in the understanding of the laws that shape network evolution, these new degree driven models cannot describe other empirical properties. Further on, in order to achieve representations that closely match reality, it is necessary to uncover new mechanisms. Following this motivation, we believe that the general view of networks as isolated systems, although possibly appropriate in some cases, must be changed if we want to describe in a proper way complex systems which do not generate spontaneously but self-organize within a medium in order to perform a function. Many networks evolve in an environment to which they interact and which usually provides the clues to understand functionality. When analyzing the dynamics of network assembly, the interlock of its constituents with the environment cannot be systematically obviated.

In this work, we blend all ideas above to present a growing network model in which both, nodes and links, are weighted [8]. The dynamical evolution is driven by exponential growth, competition for resources, and adaptation to maintain functionality in a demand and supply balance, key mechanisms which may be relevant in a wide range of self-organizing systems, in particular those where functionality is tied to communication or traffic. The medium in which the network grows and to with it interacts is here represented by a pool of elements which, at the same time, provide resources to the constituents of the network and demand functionality, say for instance users in the case of the Internet [9] or passengers in the case of the world-wide airport network [10]. Competition is here understood as a struggle between network nodes for new resources and is modeled as a rich get richer (preferential attachment) process. For their part, these captured elements demand functionality so that nodes must adapt in order to perform efficiently. This adaptation translates into the creation of weighted links between nodes.

We apply those ideas to the Internet at the Autonomous System (AS) level [9]. ASs are defined as independently administered domains which autonomously

[a] e-mail: mdserran@indiana.edu

determine internal communications and routing policies and, as a first approximation, they can be identified with Internet Service Providers (ISP). This level of description means a coarse grained representation of the Internet, in contrast to a more detailed router description. The Internet is a paradigmatic example of a self-organizing complex system and significant efforts has been devoted to the development of models which reproduce its topological properties. Candidates run from topology generators to degree driven growing networks models or Highly Optimized Tolerance (HOT) models, see [9] and references therein. Some of them reproduce heavy-tailed degree distributions and small-world properties, but perform poorly when estimating correlations or other characteristic properties, such as the k-core structure [11]. By contrast, we will show that our model nicely reproduces an overwhelming number of observed topological features: the small-world property, the scale-free degree distribution $P(k)$, high clustering coefficient c_k that shows a hierarchical structure, disassortative degree-degree correlations $\bar{k}_{nn}(k)$ [12], the scaling of the higher order loop structure [13], the distributions of the betweenness centrality, $P(b)$, and triangles passing through a node, $P(T)$, and, finally, the k-core decomposition uncovering its hierarchical organization.

In the next sections we analyze the growth of the Internet over the last years. Then we present the model. Working in the continuum approximation, we find analytically the distribution of the sizes (in number of users) of ASs and the degree distribution. Then, we introduce an algorithm in order to simulate network assembly. At this stage, we also make a first attempt to the consideration of geographical constraints. Finally, the synthetic networks are compared to the real maps of the Internet through a variety of different measures.

2 The growth of the Internet

Let $W(t)$ be the total number of users at a given time t, measured as hosts. $N(t)$ and $E(t)$ stand for the number of ASs and edges among them in the network, respectively. Empirical measures for the growth in the number of users have been obtained from the Hobbes Internet Timeline (`www.zakon.org/robert/internet/timeline/`). The growth of the network is analyzed from AS maps collected by the *Oregon route-views* project, which has recorded data since November 1997 (`moat.nlanr.net/Routing/rawdata/`). According to those observations, see reference [8], we will assume exponential growths for these quantities, $W(t) \approx W_0 e^{\alpha t}$, $N(t) \approx N_0 e^{\beta t}$, and $E(t) \approx E_0 e^{\delta t}$. These exponential growths, in turn, determine the scaling relations with the system size: $W \propto N^{\alpha/\beta}$, $E \propto N^{\delta/\beta}$ and $\langle k \rangle \propto N^{\delta/\beta - 1}$. The rates of growth can be measured to be $\alpha = 0.036 \pm 0.001$, $\beta = 0.0304 \pm 0.0003$, and $\delta = 0.0330 \pm 0.0002$ (units are month^{-1}), where $\alpha \gtrsim \delta \gtrsim \beta$. These three rates are quite close to each other but they are not equal. In fact, the inequality $\alpha \gtrsim \beta$ must hold in order to preserve network functionality. When the number of users increases at a rate α, there are two mechanisms capable to compensate the demand they represent: the creation of new nodes and the creation of new connections by nodes already present in the network. When both mechanisms take place simultaneously, the rate of growth of new nodes, β, as well as the rate for the number of connections, δ, must necessarily be smaller than α. Any other situation would lead to an imbalance between demand and supply of service in the system. On the other hand, in a connected network, δ must be equal or greater than β. If δ equals β the average number of connections per node, or average degree, remains constant in time, whereas it increases when $\delta \gtrsim \beta$. This increase could correspond to a demand per user which is not constant but grows in time, probably due to the increase of the power of computers over time and, as a consequence, to the ability to transfer bigger files or to use more demanding applications.

3 The model

We define our model according to the following rules: (i) at rate $\alpha W(t)$, new users join the system and choose node i according to some preference function, $\Pi_i(\{\omega_j(t)\})$, where $\omega_j(t)$, $j = 1, \cdots, N(t)$, is the number of users already connected to node j at time t. The function $\Pi_i(\{\omega_j(t)\})$ is normalized so that $\sum_i \Pi_i(\{\omega_j(t)\}) = 1$ at any time. (ii) at rate $\beta N(t)$, new nodes join the network with an initial number of users, ω_0, randomly withdrawn from the pool of users already attached to existing nodes. Therefore, ω_0 can be understood as the minimum number of users required to keep nodes in business. (iii) at rate λ, each user changes his AS and chooses a new one using the same preference function $\Pi_i(\{\omega_j(t)\})$. Finally, (iv) each node tries to adapt its number of connections to other nodes according to its present number of users or size, in an attempt to provide them an adequate functionality. With all specifications above, we will work in the continuum approximation to find some analytic results, specifically the distribution of the sizes of ASs and the degree distribution.

Analytic results. The resource dynamics of single nodes is described by the stochastic differential equation

$$\frac{d\omega_i}{dt} = A(\omega_i, t) + [D(\omega_i, t)]^{1/2} \xi(t), \tag{1}$$

where ω_i is the number of users attached to AS i at time t. The time dependent drift is $A(\omega_i, t) = (\alpha + \lambda)\, W(t)\Pi_i - \lambda\omega_i - \beta\omega_0$, and the diffusion term is $D(\omega_i, t) = (\alpha + \lambda)\, W(t)\Pi_i + \lambda\omega_i + \beta\omega_0 - 2\lambda\omega_i\Pi_i$. Application of the Central Limit Theorem guarantees the convergence of the noise $\xi(t)$ to a Gaussian white noise in the limit $W(t) \gg 1$. The first term in the expression for the drift is a creation term accounting for new and old users that choose node i. The second term represents those users who decide to change their node and, finally, the last term corresponds to the decrease of users due to introduction of newly created nodes. To proceed further, we need to specify the preference function $\Pi_i(\{\omega_j(t)\})$. We assume that nodes bigger in resources get users more easily than small ones. The simplest function satisfying this condition corresponds to the linear preference, that is, $\Pi_i(\{\omega_j(t)\}) = \omega_i/W(t)$, where $W(t) = \omega_0 N_0 \exp(\alpha t)$. In this case, the stochastic differential equation (1) reads

$$\frac{d\omega_i}{dt} = \alpha\omega_i - \beta\omega_0 + [(\alpha + 2\lambda)\omega_i + \beta\omega_0]^{1/2}\,\xi(t). \qquad (2)$$

Notice that reallocation of users (*i.e.* the λ-term) only increases the diffusive part in equation (2) but has no net effect in the drift term, which is, eventually, the leading term. The complete solution of this problem requires to solve the Fokker-Planck equation corresponding to equation (2) with a reflecting boundary condition at $\omega = \omega_0$ and initial conditions $p(\omega_i, t_i|\omega_0, t_i) = \delta(\omega_i - \omega_0)$ ($\delta(\cdot)$ stands for the Dirac delta function). Here $p(\omega_i, t|\omega_0, t_i)$ is the probability that node i has a number of users ω_i at time t given that it had ω_0 at time t_i. The choice of a reflecting boundary condition at $\omega = \omega_0$ is equivalent to assume that β is the overall growth rate of the number of nodes, that is, the composition of the birth and dead processes ruling the evolution of the number of nodes.

To solve the problem we can take advantage of the fact that, when $\alpha > \beta$, the average number of users of each node increases exponentially and, since $D(\omega_i, t) = \mathcal{O}(A(\omega_i, t))$, fluctuations vanishes in the long time limit. Under this zero noise approximation, the number of users connected to a node introduced at time t_i is $\omega_i(t|t_i) = \frac{\beta}{\alpha}\omega_0 + (1 - \frac{\beta}{\alpha})\omega_0 e^{\alpha(t-t_i)}$. The probability density function of ω can be found to be, in the long time limit,

$$p(\omega, t) = \frac{\tau(1-\tau)^\tau \omega_0^\tau}{(\omega - \tau\omega_0)^{1+\tau}}\Theta(\omega_c(t) - \omega), \qquad (3)$$

where we have defined $\tau \equiv \beta/\alpha$ and the cut-off is given by $\omega_c(t) \sim (1-\tau)\omega_0 e^{\alpha t} \sim W(t)$. Thus, $p(\omega, t)$ approaches a stationary distribution with an increasing cut-off that scales linearly with the total number of users. The characteristic exponent of the distribution $1 + \tau$ takes a value smaller but close to 2, since $\alpha \gtrsim \beta$.

The key point now is to relate the number of users attached to an AS with its degree and bandwidth, see [8] for a detailed discussion. Our first assumption is that each node adapts its total bandwidth proportionally to its number of users following the lineal relation

$$b_i(t|t_i) = 1 + a(t)\,(\omega_i(t|t_i) - \omega_0)\,, \qquad (4)$$

where $b_i(t|t_i)$ is the total bandwidth of a node at time t given that it was introduced at time t_i. Summing equation (4) for all nodes we get $a(t) = (2B(t) - N(t))/(W(t) - \omega_0 N(t)) \approx 2B(t)/W(t)$, where $B(t)$ is the total bandwidth of the network and we assume will grow according to $B(t) = B_0 e^{\delta' t}$, $\delta' \geq \alpha$ and, thus, $\delta' > \delta$. As a consequence, the topological degree of a node cannot be proportional to its bandwidth and we then propose the following scaling relation

$$k(t|t_i) = [b(t|t_i)]^{\beta/\delta'}\,. \qquad (5)$$

All four growth rates in the model are not independent. Summing equation (5) for all vertices, the scaling of the total number of connections is $E(t) \propto N(t)^{2-\alpha/\delta'}$, which leads to $\delta' = \alpha\beta/(2\beta - \delta)$. Combining this with equations (3–5), the degree distribution reads

$$P(k) \approx \frac{\tau(1-\tau)^\tau\,[\omega_0 a(t)]^\tau}{\mu}\frac{1}{k^\gamma}\Theta(k_c(t) - k) \qquad (6)$$

for $k \gg 1$, where the exponent γ takes the value $\gamma = 1 + 1/(2 - \delta/\beta)$. Strikingly, the exponent γ has lost any direct dependence on α becoming a function of the ratio δ/β. Using the empirical values for β and δ, the predicted exponent is $\gamma = 2.2 \pm 0.1$, in excellent agreement with the values reported in the literature [12,14]. Of course, this does not mean that the exponent γ is independent of α, since both, β and δ, may depend on the growth of the number of users. Anyway, our model turns out to depend on just two independent parameters which can be expressed as ratios of the rates of growth, β/α and δ/β.

Simulations. Our algorithm, following the lines of the model, works in four steps: (i) at iteration t, $\Delta W(t) = \omega_0 N_0\,(e^{\alpha t} - e^{\alpha(t-1)})$ users join the network and choose provider among the existing nodes using the linear preference rule. (ii) $\Delta N(t) = N_0(e^{\beta t} - e^{\beta(t-1)})$ new ASs are introduced with ω_0 users each, those being randomly withdrawn from already existing ASs. Newly created ASs are located in a two dimensional plane following a fractal set of dimension $D_f = 1.5$ [7]. (iii) Each AS evaluate its increase of bandwidth, $\Delta b_i(t|t_i)$, according to equation (4). (iv) a pair of nodes, (i, j), is chosen with probability proportional to $\Delta b_i(t|t_i)$ and $\Delta b_j(t|t_j)$. The reason is that those nodes that need a high bandwidth increase will be more active when looking for partners to whom form connections. Whenever they both need to increase their bandwidth, they form a connection with probability $D(d_{ij}, \omega_i, \omega_j) = e^{-d_{ij}/d_c(\omega_i,\omega_j)}$, where $d_c(\omega_i, \omega_j) = \omega_i\omega_j/\kappa W(t)$, and κ is a characteristic cost of number of users per unit distance. We assume that the physical distance among the ASs follows the same spatial distribution as the one measured for routers [7]. This distance function takes into consideration that, due to connection costs, physical links over long distances are unlikely to be created by small peers. Once the first connection has been formed, they create a new connection with probability r (trade-off connection costs versus diversification), whenever they still need to increase their bandwidth. This step is repeated until all nodes have the desired bandwidth.

All simulations are performed using $\omega_0 = 5000$, $N_0 = 2$, $B_0 = 1$, $\alpha = 0.035$, $\beta = 0.03$, and $\delta' = 0.04$. The final size of the networks is $N \approx 11\,000$, which approximately corresponds to the size of the actual maps for 2001 that we are considering in this work.

4 Testing the model

To test the model we construct synthetic networks from our algorithm with and without taking into consideration the geographical distribution of ASs, and we contrast several measures on those graphs to those of real

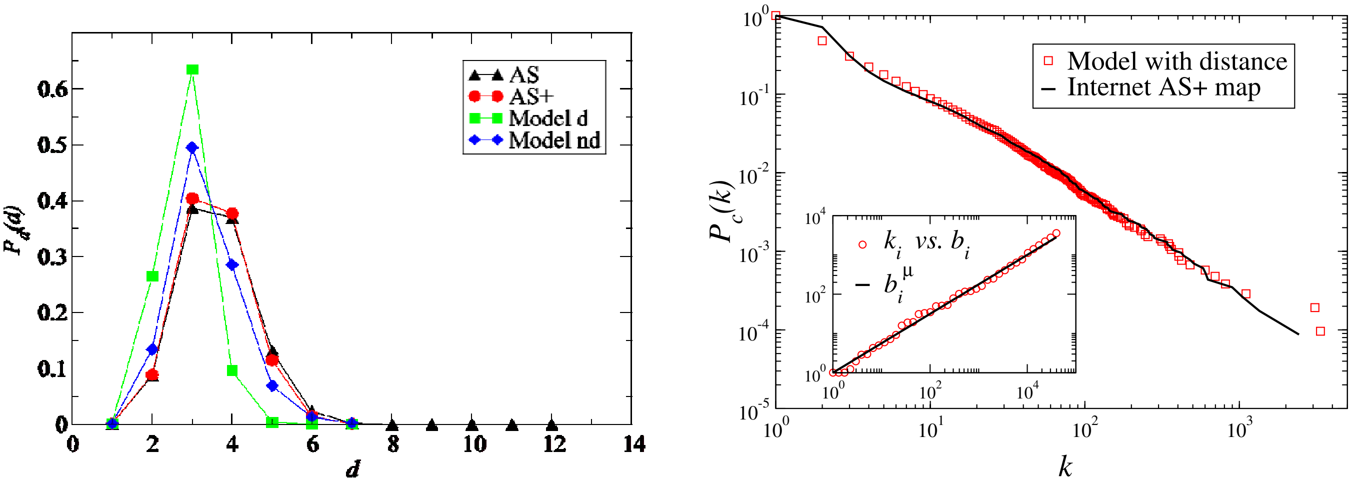

Fig. 1. Distribution of the shortest path lengths (left) and cumulative degree distribution ($P_c(k) = \sum_{k' \geq k} P(k')$) (right) for the extended AS map compared to simulations of the model, $r = 0.8$. Inset (right): Simulation results of the AS's degree as a function of AS's bandwidth. The solid line stands for the scaling relation equation (5) with $\mu = \beta/\delta' = 0.75$.

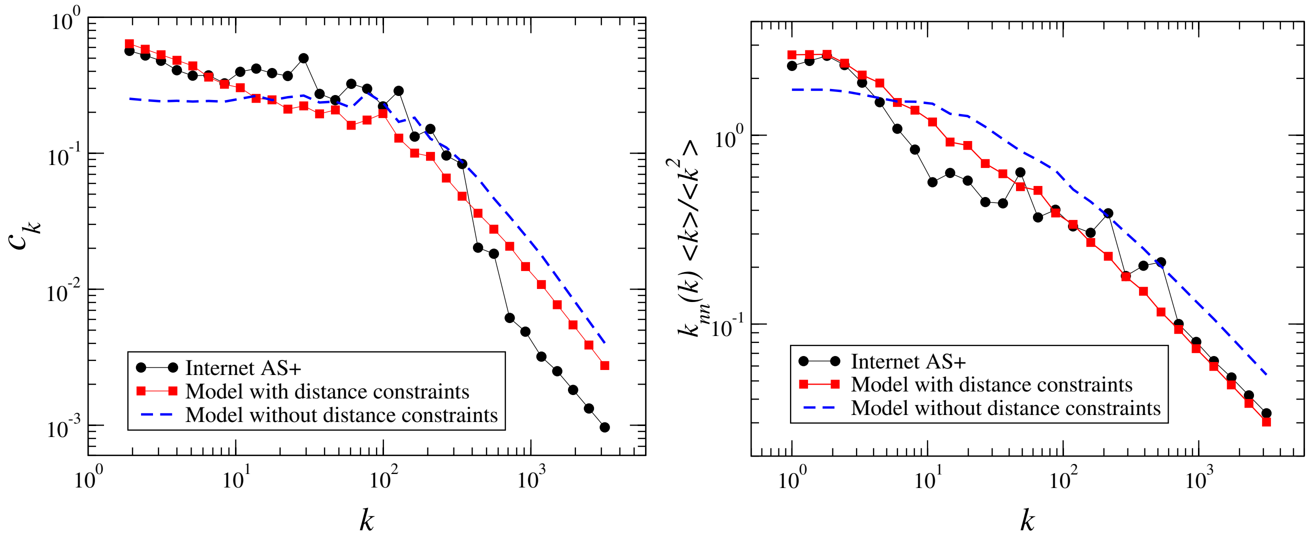

Fig. 2. Clustering coefficient, c_k, (left), and normalized average nearest neighbors degree, $\bar{k}_{nn}(k)\langle k\rangle/\langle k^2\rangle$, (right) for the AS+ map and for the model with and without distance constraints.

maps, more specifically, the AS map dated May 2001 from data collected by the Oregon Route Views Project (`moat.nlanr.net/Routing/rawdata/`), and the AS extended (AS+) map [15] which completes the previous one with data from other sources. Let us note that all the measures presented here are performed over the same synthetic networks. The parameters of the model are fixed once and for all before generating the networks so that they are not tuned in order to approach different properties.

First, we analyze the features of traditional interest when aiming to reproduce the Internet topology. The small world effect becomes clear when analyzing the distribution of the shortest path lengths, as seen in the left side graph of Figure 1, with an average value very close to the real one. The graph on the right of Figure 1 shows simulation results for the cumulative degree distribution, in nice agreement to that measured for the AS+ map. The inset exhibits simulation results of the AS's degree as a function of the AS's bandwidth, confirming the scaling *ansatz* equation (5). Clustering coefficient and average nearest neighbors degree are showed in Figure 2. Dashed lines result from the model without distance constraints, whereas squares correspond to the model with distance constraints. Interestingly, the high level of clustering coming out from the model arises as a consequence of the pattern followed to attach nodes, so that only those AS willing for new connections will link. Then, distance constraints introduce a disassortative component by inhibiting connections between small ASs so that the hierarchical structure of the real network is better reproduced. Now, we turn our attention to new measures, which run from the scaling of higher orders loops to the k-core structure. Not only two-point correlations are well approximated by

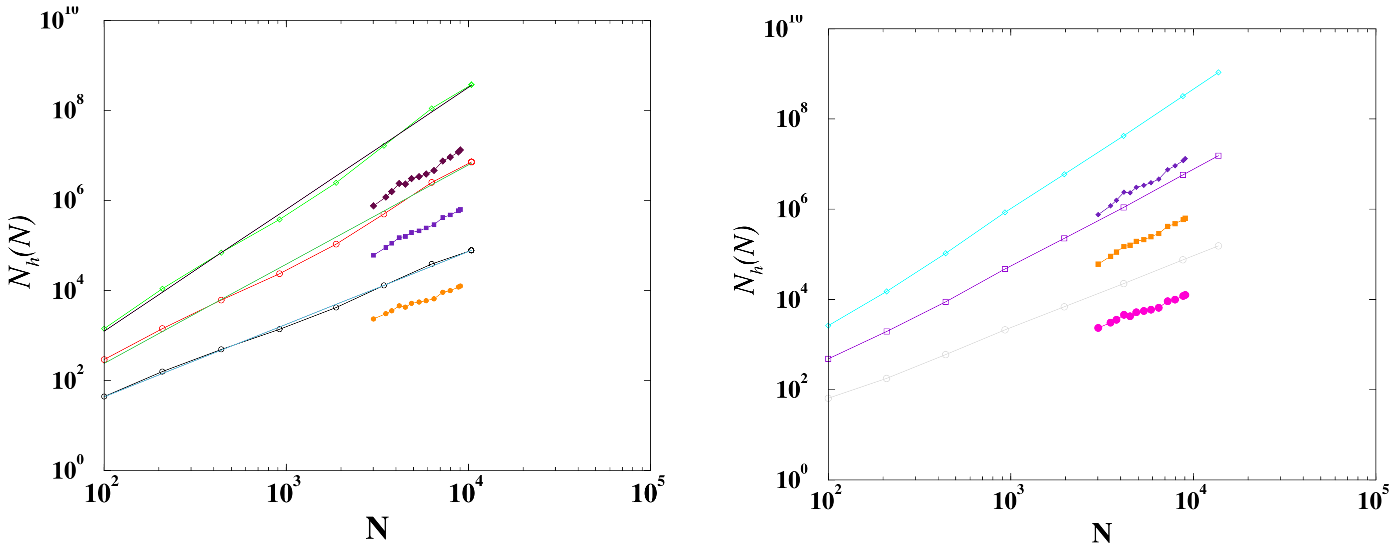

Fig. 3. Scaling of the number of loops of size 3, 4 and 5 (from bottom to top) for the model with and without distance constraints, on the left and on the right respectively. Short lines correspond to real measures.

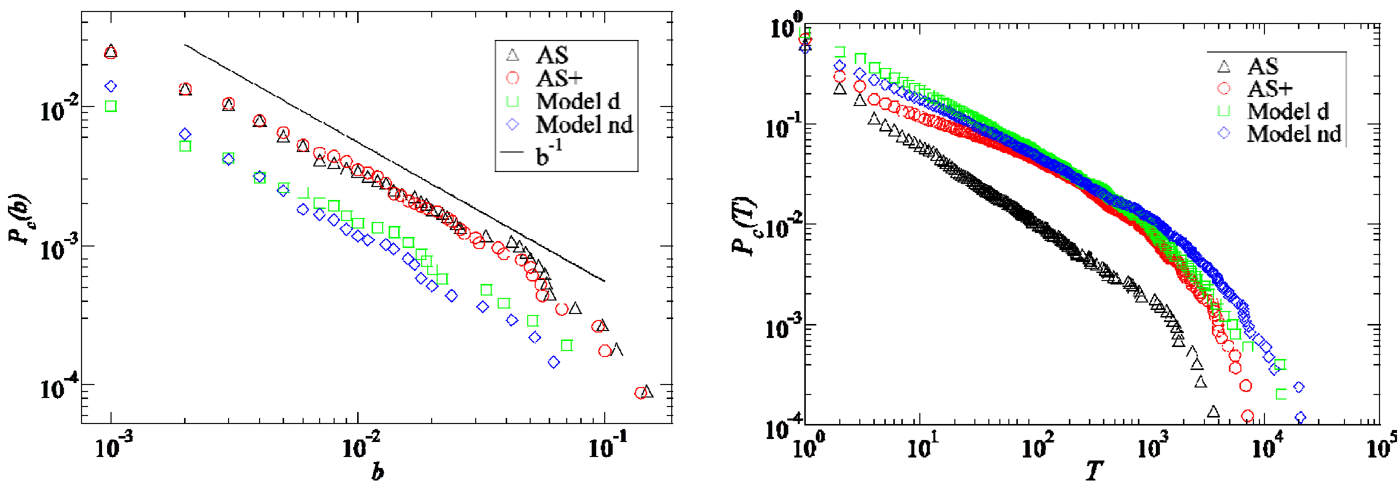

Fig. 4. Cumulative distributions of the betweenness centrality (left) and of the number of triangles passing by a node (right).

our model, but it is also able to reproduce the scaling behavior of the number of loops of size 3, 4 and 5. This has been recently measured for the Internet at the AS level in [13], and it is seen to follow a power of the system size of the form $N_h(N) \sim N^{\xi(h)}$, with exponents that are closely reproduced by our synthetic networks, see Figure 3 and Table 1. In Figure 4, we observe on the left the cumulative distribution of betweenness centrality as proposed by Freeman [16], a measure of the varying importance of the vertices in a network. On the right, the cumulative distribution of triangles passing by a node.

Finally, we also show the k-core decomposition of the actual and the synthetic maps [11]. The k-core decomposition is a recursive reduction of the network as a function of the degree, which allows the recognition of hierarchical structure and more central nodes. A very good agreement between real measures and our models can be appreciated in Figure 5. In the case of the model with distance constraints, even the coreness, the maximum number of layers in the k-core decomposition, is almost the same as in the Internet map.

Table 1. Values for the exponents $\xi(h)$ for $h = 3$, 4, and 5 for the Internet and the models with and without distance constraints (after Bianconi et al. [13]).

System	$\xi(3)$	$\xi(4)$	$\xi(5)$
Internet AS map	1.45 ± 0.07	2.07 ± 0.01	2.45 ± 0.08
Model with distance	1.60 ± 0.01	2.20 ± 0.03	2.70 ± 0.03
Model without distance	1.59 ± 0.03	2.11 ± 0.03	2.64 ± 0.03

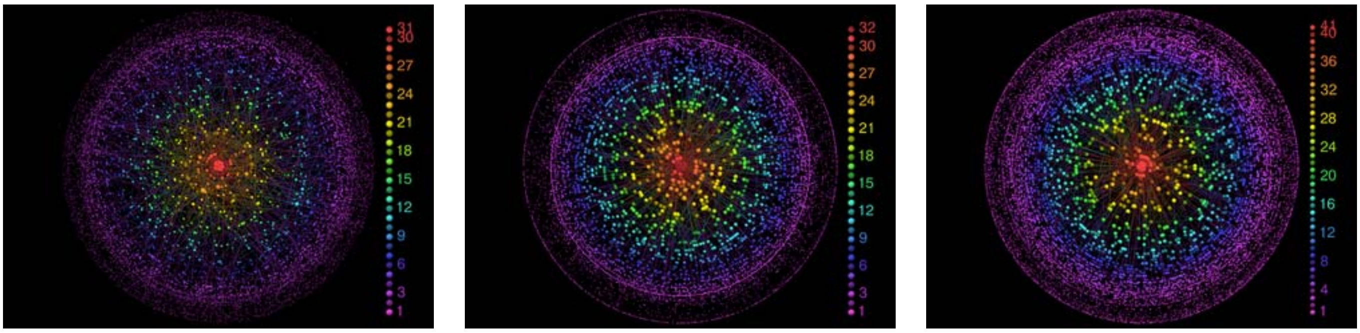

Fig. 5. k-cores for the AS+ Internet map (left) and for the maps generated from our model with and without distance (center and right). Visualizations produced by the tool LANET-VI [11].

5 Conclusions

We have presented a simple weighted growing network model for the Internet, based on evolution, environmental interaction and heterogeneity. The dynamics is driven by two key mechanisms, competition and adaptation, which may be relevant in other self-organizing systems. Beyond technical details, many empirical features are nicely reproduced but open questions remain, perhaps the more important one being whether the general ideas and mechanisms exposed in this work could help us to better understand other complex systems.

We acknowledge G. Bianconi for kindly providing us with Figure 3, and Ignacio Alvarez-Hamelin and Alessandro Vespignani for valuable comments on the k-core decomposition. This work has been partially supported by DGES of the Spanish government, Grant No. FIS2004-05923-CO2-02 and Grant No. BFM-2003-08258, and EC-FET Open project COSIN IST-2001-33555. M.B. acknowledges financial support from the MCyT (Spain).

References

1. S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW* (Oxford University Press, Oxford, 2003)
2. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)
3. B.A. Huberman, L.A. Adamic, Nature (London) **401**, 131 (1999)
4. K.-I. Goh, B. Kahng, D. Kim, Phys. Rev. Lett. **88**, 108701 (2002)
5. A. Capocci, G. Caldarelli, R. Marchetti, L. Pietronero, Phys. Rev. E **64**, 035105 (2001)
6. M. Fayed, P. Krapivsky, J.W. Byers, M. Crovella, D. Finkel, S. Redner, Comput. Commun. Rev. **33**, 41 (2003)
7. S.H. Yook, H. Jeong, A.-L. Barabási, Proc. Nat. Acad. Sci. USA **99**, 13382 (2002)
8. M.A. Serrano, M. Boguñá, A. Díaz-Guilera, Phys. Rev. Lett. **94**, 038701 (2005)
9. R. Pastor-Satorras, A. Vespignani, *Evolution and Structure of the Internet. A Statistical Physics Approach* (Cambridge University Press, Cambridge, 2004)
10. A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, Proc. Natl. Acad. Sci. **101**, 3747 (2004).
11. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, A. Vespignani, cs.NI/0504107, `http://xavier.informatics.indiana.edu/lanet-vi/`
12. R. Pastor-Satorras, A. Vázquez, A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001); A. Vázquez, R. Pastor-Satorras, A. Vespignani, Phys. Rev. E **65**, 066130 (2002)
13. G. Bianconi, G. Caldarelli, A. Capocci, Phys. Rev. E **71**, 066116 (2005)
14. M. Faloutsos, P. Faloutsos, C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999)
15. C. Qian, H. Chang, R. Govindan, S. Jamin, S. Shenker, W. Willinger, Proceedings IEEE, IEEE Computer Society Press **2**, 608 (2002)
16. L.C. Freeman, Sociometry **40**, 35 (1977)