

## GROWING HIERARCHICAL SCALE-FREE NETWORKS BY MEANS OF NONHIERARCHICAL PROCESSES

S. BOCCALETTI and D.-U. HWANG

*CNR–Istituto dei Sistemi Complessi, Via Madonna del Piano,  
 10, 50019 Sesto Fiorentino (FI), Italy*

V. LATORA

*Dipartimento di Fisica e Astronomia,  
 Università di Catania, and INFN Sezione di Catania,  
 Via S. Sofia, 64, 95123 Catania, Italy*

Received July 20, 2006; Revised September 30, 2006

We introduce a fully nonhierarchical network growing mechanism, that furthermore does not impose explicit preferential attachment rules. The growing procedure produces a graph featuring power-law degree and clustering distributions, and manifesting slightly disassortative degree-degree correlations. The rigorous rate equations for the evolution of the degree distribution and for the conditional degree-degree probability are derived.

*Keywords:* Network models; growth process; scale-free distributions.

Complex networks are the prominent candidates to describe the topology of many real systems in social, technological and natural sciences [Albert & Barabási, 2002; Newman, 2003; Boccaletti *et al.*, 2006].

In particular, the massive and comparative analysis of the available data has uncovered four main properties common to most of the real-world networks, namely: (i) scale-free degree distributions, (ii) the small-world behavior, (iii) a hierarchical structure in the clustering features, and (iv) degree-degree correlations. The first property accounts for a degree distribution  $P_k$  (the probability that a node, chosen at random, has  $k$  edges) exhibiting a power law tail,  $P_k \sim k^{-\gamma}$  [Barabási & Albert, 1999]. The second indicates that the characteristic path length  $\langle l \rangle$  scales at most logarithmically with the number of nodes  $N$  in the graph, while the clustering coefficient  $C$  is of the order of one independent of  $N$  [Watts & Strogatz, 1998]. The third means that the clustering coefficient  $C_k$  of a

connectivity class  $k$  (the average clustering coefficient of all nodes with a given degree  $k$ ) does depend on  $k$  as  $C_k \sim k^{-\omega}$  [Ravasz *et al.*, 2002]. The fourth reveals that the probability that a node of degree  $k$  is connected to another node of degree  $k'$ , depends on both  $k$  and  $k'$ . This is reflected by a degree-degree correlation coefficient  $r \neq 0$  [Newman, 2002]. The specific values of  $\gamma$ ,  $\omega$  and  $r$  depend on the specific case under study. One usually finds  $2 \leq \gamma \leq 3$  [Albert & Barabási, 2002; Newman, 2003; Boccaletti *et al.*, 2006],  $\omega \simeq 1$  [Ravasz *et al.*, 2002], and  $r < 0$  ( $r > 0$ ) for technological and biological (social) networks [Newman, 2002].

Network models featuring a power-law degree distribution can be simply obtained as a particular case of random graphs with a given degree distribution [Bender & Canfield, 1978]. While these models are an appropriate representation of all those cases in which growth or aging do not play a dominant role in determining the structural properties of a network, the recent research on graph modeling

has instead concentrated on networks emerging from a growing process [Albert & Barabási, 2002; Newman, 2003; Boccaletti *et al.*, 2006; Barabási & Albert, 1999]. In this framework, it was initially postulated that a scale-free degree distribution is the direct consequence of a *preferential attachment* mechanism taking place in the growing process, through which the larger is the degree [Barabási *et al.*, 1999; Albert & Barabási, 2000] (or the intrinsic fitness [Bianconi & Barabási, 2001]) of an existing node, the higher the probability that this node will form connections with newly added ones. Only in the last few years some alternatives to the preferential attachment mechanism, like the so-called *vertex copying mechanisms* and *duplication models* [Kim *et al.*, 2002; Chung *et al.*, 2003; Vázquez *et al.*, 2003] have been introduced in order to generate power-law distributions.

Furthermore, the scaling of the clustering coefficient  $C_k \sim 1/k$  observed in real networks, can be attained in models of growing graphs by further introducing *ad hoc* mechanisms (as, for instance, by hierarchically repeating the same construction rules at different scales [Ravasz *et al.*, 2002], or by considering a set of local rules [Vázquez, 2003]).

Nevertheless, a direct preferential attachment mechanism manifests two main drawbacks. First of all, it is not plausible that a node adding to the network has full information on the degree sequence of the existing nodes, which in fact the preferential attachment rule would require to fix its connectivity properties. Secondly, a scale-free degree distribution comes out to be a singular limit corresponding to *linear* preferential attachment processes. Indeed, by considering attachment processes in which  $\Pi_{n \rightarrow j} \propto k_j^\alpha$  ( $\Pi_{n \rightarrow j}$  being the probability that a new node  $n$  attaches to the existing node  $j$ ), Krapivsky *et al.* [2000] has shown that, if  $\alpha < 1$ , the mechanism produces a stretched exponential degree distribution while, for  $\alpha > 1$ , a single site connects to nearly all other sites.

In this paper, we show how it is possible, instead, to grow a graph displaying power-law features in both degree and clustering distributions, by means of a fully nonhierarchical mechanism, that furthermore does not impose explicitly any preferential attachment rule.

The model to grow a network of  $N$  nodes starts from an initial ( $t = 0$ ) connected graph of  $N_0 \ll N$  nodes and  $K_0$  edges, such that no node has degree smaller than a given  $m$ . For simplicity, here, we fix  $K_0 = N_0(N_0 - 1)/2$ , i.e. we start from

a complete graph with  $N_0$  nodes. Time is discrete ( $t = 1, 2, \dots, N - N_0$ ), and at each time step a new node  $n$ , with  $m$  edges, is added to the graph. The new node is linked to  $m$  different nodes of the existing graph selected by the following three-steps rule:

- (A) The new node  $n$  selects randomly a node, say node  $j$ , in the graph with uniform probability among the set of  $N_0 + t$  existing nodes.
- (B) The new node considers the set  $\mathcal{S}_j(t)$  formed by  $k_j(t) + 1$  nodes: the node  $j$  and the set  $\mathcal{N}_j(t)$  of its  $k_j$  neighbors at time  $t$ .
- (C) The new node  $n$  links to  $m$  distinct nodes, randomly chosen with uniform probability from the set  $\mathcal{S}_j(t)$ .

Notice that this growing mechanism shares some features with the *vertex copying mechanisms* and *duplication models* [Kim *et al.*, 2002; Chung *et al.*, 2003; Vázquez *et al.*, 2003], and with a particular kind of model based on local rules, the so-called *random walk model*, proposed in [Vázquez, 2003]. Furthermore, in the proposed growing process, each newly added node  $n$  has only a partial information on the existing network structure (limited to the initially randomly selected node  $j$  and its neighborhood  $\mathcal{N}_j$ ), and does not use explicitly a preferential rule for setting up the connectivity (both the initial node  $j$  and the connections are chosen at random with a uniform distribution). Nevertheless, it should be highlighted that high degree nodes automatically receive more new links (in a way linearly proportional to their degree) because they can be reached in a larger number (linearly proportional to the degree) of ways, as first suggested by Cohen *et al.* [2003]. Therefore, our growing procedure can be seen as a way of introducing a linear preferential attachment mechanism without having to *ad hoc* postulating it in a direct way in the network growth.

The resulting degree distribution in the model ( $N_0 = 6$  and  $m = 5$ ) is depicted in Fig. 1(a). The data refer to an ensemble average over ten different realizations of a network with  $N = 10\,000$  nodes. Within the same panel, we draw the line corresponding to the function  $P_k = k^{-\gamma}$ , with  $\gamma = 3$ , in order to make it evident that the proposed mechanism induces a power-law degree distribution that has the very same features of those attained by the original linear preferential attachment mechanism [Barabási & Albert, 1999]. The same value of  $\gamma$  is obtained for different  $N_0$  and  $m$  values.

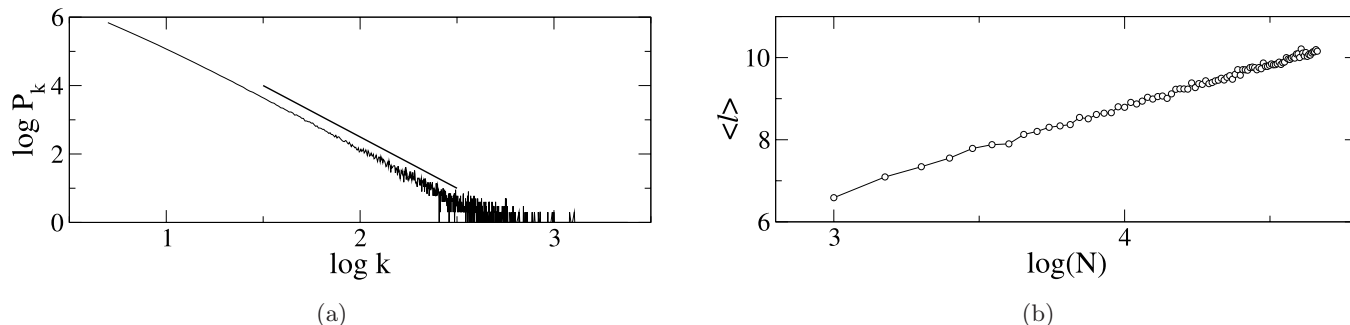


Fig. 1. (a) Degree distribution  $P_k$  (in logarithm scale) versus the logarithm of the node degree, for  $N_0 = 6$  and  $m = 5$ . Both quantities are adimensional. The straight line depicts the function  $P_k = k^{-\gamma}$ , with  $\gamma = 3$ . (b) Characteristic path length  $\langle l \rangle$  versus the logarithm of the network size  $N$ .

A measure of the typical separation between two nodes in a graph is given by the *characteristic path length*, defined as the mean of geodesic lengths over all couples of nodes [Watts & Strogatz, 1998]:

$$\langle l \rangle = \frac{1}{N(N-1)} \sum_{i,j=1,\dots,N,i \neq j} d_{ij}, \quad (1)$$

where  $d_{ij}$  is the length of the geodesic (the number of edges in the shortest path) from node  $i$  to node  $j$ . Figure 1(b) highlights that the proposed mechanism originates a graph possessing the *small-world property*, i.e. such that  $\langle l \rangle$  scales logarithmically with the network's size  $N$ , as it was observed in many real world networks [Watts & Strogatz, 1998].

As for the clustering properties, these can be quantified by means of the graph *clustering*

*coefficient*  $C$ , given by the average of  $c_i$  over all the nodes:

$$C = \langle c \rangle = \frac{1}{N} \sum_{i=1,\dots,N} c_i, \quad (2)$$

where  $c_i$  (the local clustering coefficient of node  $i$ ) is the normalized number of edges (denoted by  $e_i$ ) in the subgraph of neighbors of  $i$ , and is defined as  $c_i = (2e_i)/[k_i(k_i - 1)]$  [Watts & Strogatz, 1998]. The inset of Fig. 2 illustrates the behavior of  $C$  versus  $N$ , and shows that, as the network size increases,  $C$  tends to approach a size-independent value very close to 1, indicating that the grown network displays prominent clustering features, at variance with what happens with the preferential attachment mechanism [Barabási & Albert, 1999]. Our

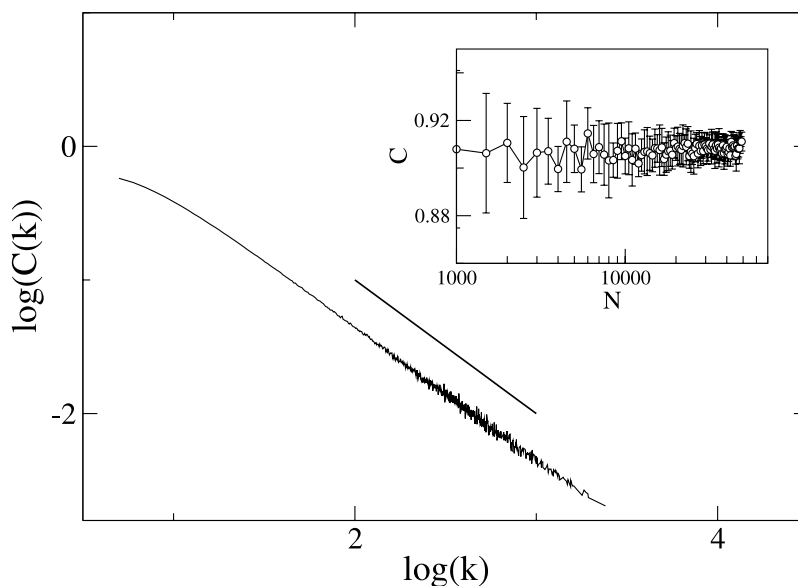


Fig. 2. The clustering coefficient  $C_k$  of a connectivity class  $k$  (in logarithm scale) versus the logarithm of  $k$ . Both quantities are adimensional. The straight line depicts the function  $C_k = k^{-\omega}$ , with  $\omega = 1$ . The inset reports the clustering coefficient  $C$  versus the network size  $N$  (in logarithmic scale).

numerical simulations indicate that this asymptotic behavior of  $C$  persists also for poorly clustered initial graphs.

Usually, another measure is also considered, namely  $C_k$ , the clustering coefficient of a connectivity class  $k$ , defined as in Eq. (2), with the sum restricted over the set of all nodes with a given degree  $k$ . Figure 2 reports  $C_k$  versus the logarithm of the node degree, showing that our growing mechanism, despite being fully nonhierarchical, recovers the power-law behavior  $C_k \sim k^{-1}$ , typical of hierarchically growing networks [Ravasz *et al.*, 2002]. This result can be explained by simple heuristic arguments. Precisely, the leading term in the probability  $\Pi_\Delta$  that the connections from a newly added node  $n$  contribute in forming triangles in the graph is equal to the probability that the added node forms a connection with the initially randomly selected node  $j$  (in this case, indeed, the connections of the new nodes will form  $m - 1$  triangles in the network). If node  $j$  has degree  $k_j$ , the probability that node  $n$  will form a connection with it is  $m/(k_j + 1)$ , and therefore the clustering coefficient  $C_k$  has to scale like  $k^{-1}$ .

As for the power-law degree distribution, one has to explicitly write down the rate equations for the time evolution of the number of nodes  $N_k(t)$  with degree  $k$  at time  $t$  [Krapivsky *et al.*, 2000].  $N_k(t)$  satisfies the normalization condition  $\sum_k N_k(t) = N(t) = N_0 + t \approx t$ . The rate equations, for  $k \geq m$ , read:

$$\begin{aligned} \frac{dN_k(t)}{dt} = & -\frac{N_k(t)}{N(t)} \frac{m}{k+1} - \sum_{k'} \frac{N_{k'}(t)}{N(t)} P(k|k') \frac{m}{k'+1} \\ & + \frac{N_{k-1}(t)}{N(t)} \frac{m}{k} + \sum_{k'} \frac{N_{k'}(t)}{N(t)} P(k-1|k') \\ & \times \frac{m}{k'+1} + \delta_{k,m}. \end{aligned} \tag{3}$$

The terms in the right-hand side of Eq. (3) with negative (positive) signs correspond to loss (gain) terms. The first of the loss terms accounts for the fact that, at each time step, a node with  $k$  edges can be chosen as the first node in the wiring process.  $N_k(t)/N(t)$  is the probability to get a node with  $k$  links by choosing a node at random, and  $m/(k + 1)$  is the probability that one out of the  $m$  new edges connects the added node with the randomly chosen one. The second contribution to the loss term comes from the fact that we can choose a node with  $k$  edges not as the initial node, but

as the first neighbor of an initial node with degree  $k'$ . The conditional probability  $P(k|k')$  in Eq. (3) is defined as the probability that a node with  $k'$  links is connected to a node with  $k$  links, and satisfies the conditions  $\sum_k P(k|k') = 1$  and  $k'P(k|k')P(k') = kP(k'|k)P(k)$  [Pastor-Satorras *et al.*, 2001; Boguñá *et al.*, 2003]. The gain part of Eq. (3) is made by two contributions similar to those in the loss term, plus a third contribution ( $\delta_{k,m}$ ) accounting for the addition, at each time step, of a new node with  $m$  links. Assuming now the existence of a stationary probability distribution  $P_k = \lim_{t \rightarrow \infty} P_k(t)$ , where  $P_k(t) = N_k(t)/N(t)$  is the degree distribution at time  $t$ , since for large times  $N(t) \approx t$ , we can write  $N_k(t) \approx tP_k$ . Substituting in Eq. (3), since  $dN_k(t)/dt = d[tP_k]/dt = P_k$ , we get:

$$\begin{aligned} P_k = & -P_k \frac{m}{k+1} - \sum_{k'} P_{k'} P(k|k') \frac{m}{k'+1} \\ & + P_{k-1} \frac{m}{k} + \sum_{k'} P_{k'} P(k-1|k') \frac{m}{k'+1} + \delta_{k,m}. \end{aligned} \tag{4}$$

Equation (4) reveals a peculiar feature of our model: the asymptotic degree distribution explicitly depends on the degree-degree correlation [the equation for  $P_k$  depends explicitly on the conditional probability  $P(k|k')$ ]. This is because the set  $\mathcal{S}_j$  in step (B) of the growing procedure, includes both the node  $j$  and its first neighbors. If one considers a simplified version of the model in which the newly attached node  $n$  links to  $m$  (randomly chosen) neighbors of a randomly chosen node  $j$  (i.e.  $\mathcal{S}_j = \mathcal{N}_j$ ), the growth turns out to be well described by the equations (valid for  $k \geq m$ ):

$$\frac{dN_k(t)}{dt} = m \frac{-kN_k(t) + (k-1)N_{k-1}(t)}{\sum_k kN_k(t)} + \delta_{k,m}. \tag{5}$$

These rate equations are simply derived by considering that the degree  $k$  of the first neighbors of a randomly chosen node is distributed as  $kN_k/\sum_k kN_k$  [Newman, 2003; Boccaletti *et al.*, 2006]. The rate equations in (5) are identical to those of the model proposed by Barabasi and Albert [1999] (see also [Barabási *et al.*, 1999; Krapivsky *et al.*, 2000]). Since for large times  $N_k(t) \approx tP_k$  and  $\sum_k kN_k(t) \approx 2mt$ , one gets the following recursive equations for the stationary degree distribution

[Newman, 2003]:

$$P_k = \begin{cases} \frac{k-1}{k+2} P_{k-1} & \text{for } k > m \\ \frac{2}{(m+2)} & \text{for } k = m \end{cases} \quad (6)$$

which give  $P_k = [2m(m+1)]/[k(k+1)(k+2)]$ . In summary, this simplified version can be solved exactly and produces a power-law degree distribution with an exponent  $\gamma = 3$ , equal to that observed in Fig. 1 for the complete model. On the other hand, the simplified model does not produce the hierarchical clustering coefficient  $C \sim 1/k$  observed in Fig. 2.

Going back to the complete model, we notice that Eq. (4) turns into a set of closed equations, similar to Eq. (6), if the growing mechanisms would produce an uncorrelated graph. In such a case the conditional probability  $P(k|k')$  would be independent on  $k'$  and would reduce to  $P(k|k') = kP_k/\langle k \rangle$ . The presence/absence of degree-degree correlations can be checked numerically in the model by computing the average degree of the nearest neighbors of nodes with degree  $k$ , denoted as  $k_{nn}(k)$  [Pastor-Satorras *et al.*, 2001]. Since the latter can be expressed as a function of  $P(k|k')$  as  $k_{nn}(k) = \sum_{k'} k' P(k'|k)$ , in the absence of correlations one would obtain  $k_{nn}(k) = \langle k^2 \rangle / \langle k \rangle$ , independent of  $k$ . Figure 3 reports  $k_{nn}$  versus  $k$  for different network's

sizes, revealing that in all cases the network manifests slightly *disassortative* degree-degree correlations [Newman, 2002] [ $k_{nn}(k)$  is a monotonically decreasing function, i.e. nodes with low degree are more likely connected with highly connected ones]. This is confirmed by the calculation of the Pearson correlation coefficient of the degrees at either ends of an edge  $r(N)$  [Newman, 2002], which turns out to be negative for any finite size  $N$ , as shown in the inset of Fig. 3. Since the correlations are small, in a first approximation we can neglect them. By plugging  $P(k|k') = kP_k/\langle k \rangle$  in Eq. (4), one gets the recursive equations:

$$P_k = \begin{cases} \frac{m \left( \frac{1}{k} + \frac{k-1}{\langle k \rangle} a \right)}{1 + m \left( \frac{1}{k+1} + \frac{k}{\langle k \rangle} a \right)} P_{k-1} & \text{for } k > m \\ \frac{1}{1 + m \left( \frac{1}{m+1} + \frac{m}{\langle k \rangle} a \right)} & \text{for } k = m \end{cases} \quad (7)$$

where  $a = \sum_{k'} (P_{k'}/(k'+1))$ . Equation (7) can be solved numerically. Though representing only the limit of uncorrelated graphs, Eq. (7) produces, for large  $k$ , power-law degree distributions for any value of  $m$  and  $a$  (for instance,  $m = 2$ ,  $\langle k \rangle = 2m$ , and  $a = 0.8$  give  $P(k) \sim k^{-\gamma}$  with  $\gamma = 3.5$ ).

If one wants to solve rigorously the model, the iterative equation for the conditional probability

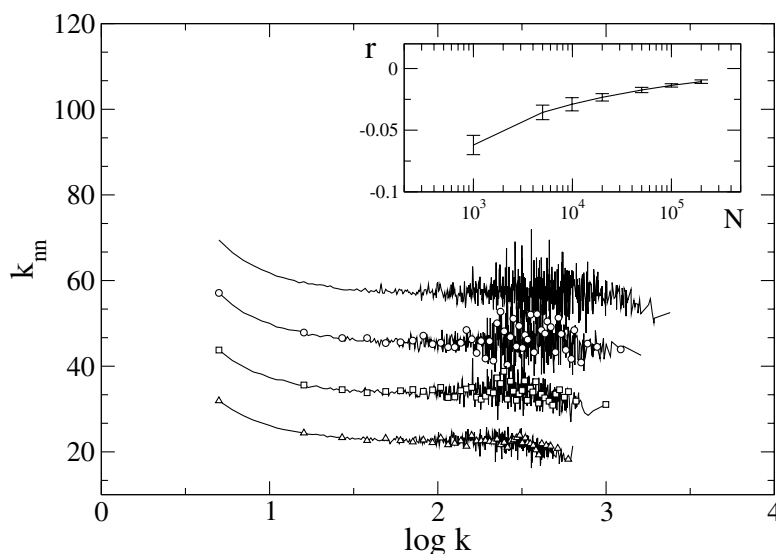


Fig. 3. Average nearest neighbors degree  $k_{nn}$  versus  $k$  (in logarithmic scale). The different curves correspond to different network's sizes. Precisely, the curve with triangles (squares, circles) refers to  $N = 10\,000$  ( $N = 20\,000$ ,  $N = 50\,000$ ), while the solid line refers to  $N = 100\,000$ . Inset: the Pearson coefficient  $r$  (see [Newman, 2002] for definition) versus the system size  $N$ , revealing the disassortative nature of the degree-degree correlations.

$P(k|k')$  has to be given. The rate equation for  $P(k|k')$  can be written down by further assuming that  $k$  is chosen simply proportional to  $N_{k-1}(t)$  (i.e. as if each link is attached at a node chosen at random). The details of the calculations are rather complicated and somehow cumbersome, and will be presented elsewhere. The final rate equations for  $P(k|k')$  are:

$$\begin{aligned}
& P(k|k')k'P(k') \\
&= P(k-1)P(k'|k-1)\frac{m}{k} \\
&+ P(k'-1)P(k|k'-1)\frac{m}{k'} \\
&- P_kP(k'|k)\frac{m}{k+1} - P(k')P(k|k')\frac{m}{k'+1} \\
&+ \sum_{k''} P(k'')\frac{m}{k''+1}[P(k-1|k'')P(k'|k-1) \\
&+ P(k'-1|k'')P(k|k'-1) - P(k|k'')P(k'|k) \\
&- P(k'|k'')P(k|k')] + m \delta_{k',1}P(k-1), \quad (8)
\end{aligned}$$

that, once coupled with Eq. (4), allow for the analytical solution of both the degree distribution  $P_k$  and the conditional probability  $P(k|k')$ .

In conclusion, we have shown that a simple nonhierarchical growing mechanism is able to produce complex network structures, featuring *all* the main statistical properties observed in most of the real networks. Though not explicitly postulating any preferential attachment rule, the procedure generates: power-law degree and clustering distributions, and disassortative degree-degree correlations. The model, therefore, overcomes the two main drawbacks typical of preferentially growing procedure, especially it does not require full information on the degree sequence when each time a new node is added. Our results therefore contribute to a better understanding of the growth mechanisms at the basis of the formation of complex topologies in technological, social and biological systems.

## Acknowledgments

The authors are indebted to A. Amann, F. T. Arecchi, M. Chavez, R. López-Ruiz and Y. Moreno for the many helpful discussions on the subject. Work was partly supported by projects MIUR-FIRB n.RBNE01CW3M-001 and the INFN project TO61. S. Boccaletti acknowledges support from the Horowitz Center for Complexity.

## References

- Albert, R. & Barabási, A.-L. [2000] “Topology of evolving networks: Local events and universality,” *Phys. Rev. Lett.* **85**, 5234–5237.
- Albert, R. & Barabási, A.-L. [2002] “Statistical mechanics of complex networks,” *Rev. Mod. Phys.* **74**, 47–97.
- Barabási, A.-L. & Albert, R. [1999] “Emergence of scaling in random networks,” *Science* **286**, 509–512.
- Barabási, A.-L., Albert, R. & Jeong, H. [1999] “Mean-field theory for scale-free random networks,” *Physica A* **272**, 173–187.
- Bender, E. A. & Canfield, E. R. [1978] “The asymptotic number of labeled graphs with given degree sequences,” *J. Combinat. Th. A* **24**, 296–307.
- Bianconi, G. & Barabási, A.-L. [2001] “Bose–Einstein condensation in complex networks,” *Phys. Rev. Lett.* **86**, 5632–5635.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. [2006] “Complex networks: Structure and dynamics,” *Phys. Rep.* **424**, 175–308.
- Boguñá, M., Pastor-Satorras, R. & Vespignani, A. [2003] “Epidemic spreading in complex networks with degree correlations,” *Lecture Notes in Physics*, Vol. 625, p. 127.
- Chung, F., Lu, L., Dewey, T. G. & Galas, D. J. [2003] “Duplication models for biological networks,” *J. Computat. Biol.* **10**, 677–687.
- Cohen, R., Havlin, S. & ben-Avraham, D. [2003] “Efficient immunization strategies for computer networks and populations,” *Phys. Rev. Lett.* **91**, 247901.
- Kim, J., Krapinsky, P. L., Kahng, B. & Redner, S. [2002] “Infinite-order percolation and giant fluctuations in a protein interaction network,” *Phys. Rev. E* **66**, 055101.
- Krapivsky, P. L., Redner, S. & Leyvraz, F. [2000] “Connectivity of growing random networks,” *Phys. Rev. Lett.* **85**, 4629–4632.
- Newman, M. E. J. [2002] “Assortative mixing in networks,” *Phys. Rev. Lett.* **89**, 208701.
- Newman, M. E. J. [2003] “The structure and function of complex networks,” *SIAM Rev.* **45**, 167–256
- Pastor-Satorras, R., Vázquez, A. & Vespignani, A. [2001] “Dynamical and correlation properties of the Internet,” *Phys. Rev. Lett.* **87**, 258701.
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A.-L. [2002] “Hierarchical organization of modularity in metabolic networks,” *Science* **297**, 1551–1556.
- Vázquez, A. [2003] “Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations,” *Phys. Rev. E* **67**, 056104.
- Vázquez, A., Flammini, A., Maritan, A. & Vespignani, A. [2003] “Modeling of protein interaction networks,” *Complexus* **1**, 38–44.
- Watts, D. J. & Strogatz, S. H. [1998] “Collective dynamics of small-world networks,” *Nature* **393**, 440–443.