Scale-Free and Hierarchical Structures in Complex Networks

Albert-László Barabási[*], Zoltán Dezső[*], Erzsébet Ravasz[*], Soon-Hyung Yook[*] and Zoltán Oltvai[†]

[*]*Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA.*
[†]*Department of Pathology, Northwestern University, Illinois 60611, USA.*

Abstract.
Networks with complex topology describe systems as diverse as the cell or the World Wide Web. The emergence of these networks is driven by self-organizing processes that are governed by simple but generic laws. In the last three years it became clear that many complex networks, such as the Internet, the cell, or the world wide web, share the same large-scale topology. Here we review recent advances in the characterization of complex networks, focusing the emergence of the scale-free and the hierarchical architecture. We also present empirical results to demonstrate that the scale-free and the hierarchical property are shared by a wide range of complex networks. Finally, we discuss the impact of the network topology on our ability to stop the spread of viruses in complex networks.

INTRODUCTION

The behavior of many natural and social systems is fundamentally determined by the interwoven web through which the system's constituents interact with each other [1, 2, 3]. For example, the cell's metabolism is maintained by a cellular network, whose nodes are substrates and links are chemical reactions [4, 5, 6, 7, 8]. But equally complex webs describe human societies, whose nodes are individuals and links represent social interactions [9, 10]; the World Wide Web (WWW) [11, 12, 13, 14] whose nodes are Web documents connected by URL links; the scientific literature, whose nodes are publications and links are citations [15, 16], or the language, whose nodes are words and links represent various syntactical or grammatical relationships between them [17, 18, 19, 20]. The networks describing these systems constantly evolve by the addition and removal of new nodes and links. Due to the diversity and the large number of the nodes and interactions, until recently the topology of these evolving networks was largely unknown and unexplored. Yet, the inability of contemporary science to address the properties of complex networks limited advances in many disciplines, including molecular biology, computer science, ecology and social sciences.

Recent results on the topology of real networks indicate that the apparent randomness of complex systems with many degrees of freedom hides generic mechanisms and order that are crucial to the understanding of the interwoven world surrounding us. By reviewing some advances in the area we wish to convey the potential for understanding complex systems through the evolution of the networks behind them.

NETWORK MODELS

Random Networks

While graph theory initially focused on regular graphs, since the 1950's large networks with no apparent design principles were described as random graphs [21], proposed as the simplest and most straightforward realization of a complex network. According to the Erdős-Rényi (ER) model of random graphs [22], we start with N nodes and connect every pair of nodes with probability p, creating a graph with approximately $pN(N-1)/2$ edges distributed randomly (Fig. 1). This model has guided our thinking about complex networks for decades after its introduction in the late 1950s. But the growing interest in complex systems prompted many scientists to ask a simple question: are real networks behind diverse complex systems fundamentally random? Our intuition offers a clear answer: complex systems must display some organizing principles which should be at some level encoded in their topology as well. But if the topology of these networks indeed deviates from a random graph, we need to develop tools and measures to capture in quantitative terms their underlying organizing principles.

Scale-Free Networks

Not all nodes in a network have the same number of edges. The spread in the number of edges of the diverse nodes, or a node's degree, is characterized by the degree distribution $P(k)$ which gives the probability that a randomly selected node has exactly k edges. Since in a random graph the edges are placed at random, the majority of nodes have approximately the same degree, close to the average degree $\langle k \rangle$ of the network. Indeed, the degrees in a random graph follow a Poisson distribution with a peak at $\langle k \rangle$.

A highly nontrivial development in our understanding of complex networks was the discovery that for most large networks, including the World-Wide Web [11], Internet [23], metabolic and protein networks [7, 8], language [17, 18, 19, 20] or sexual [24] networks, the degree distribution follows a power-law

$$P(k) \sim k^{-\gamma}. \tag{1}$$

Networks with a power-law degree distribution are called scale-free [25]. A non-comprehensive list of scale-free networks reported so far is shown in Table 1.

There are major topological differences between random and scale-free networks (Fig. 1). For the former most nodes have approximately the same number of links, $k \approx \langle k \rangle$, the exponential decay of $P(k)$ guaranteeing the absence of nodes with significantly more links than $\langle k \rangle$. In contrast, the power-law distribution implies that nodes with only a few links are numerous, but a few nodes have a very large number of links.
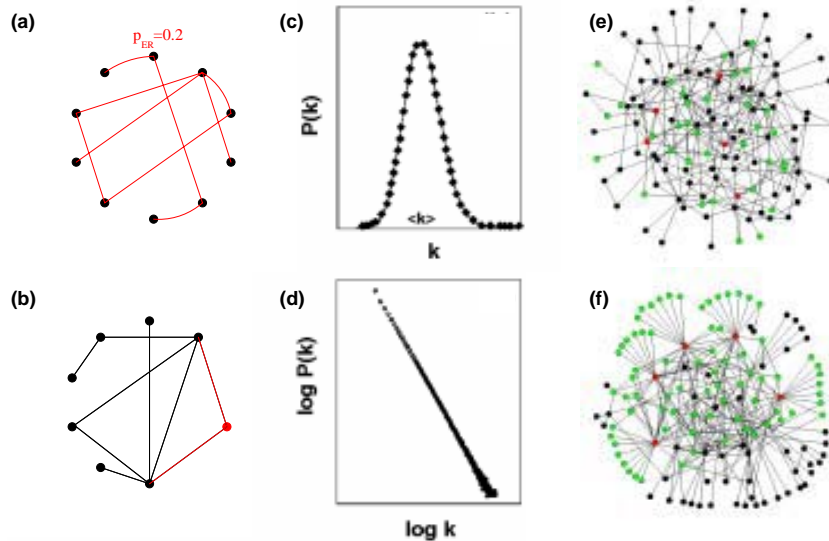
2

FIGURE 1. **(a)** The Erdős-Rényi random network model is constructed by laying down *N* nodes and connecting each pair of nodes with probability *p*. The figure shows a particular realization of such a network for $N = 10$ and $p = 0.2$. **(b)** The scale-free model assumes that the network continually grows by the addition of a new nodes. The figure shows the network at time *t* (black nodes and links) and after the addition of a new node at time $t + 1$ (red). The probability that the new node chooses a node with *k* links follows (2), favoring highly connected nodes, a phenomenon called preferential attachment. **(c)** For the random graph generated by the Erdős-Rényi model the degree distribution, $P(k)$, is strongly peaked at $k = <k>$ and decays exponentially for large *k*. **(d)** $P(k)$ for a scale-free network does not have a peak, and decays as a power-law, $P(k) \sim k^{-\gamma}$. **(e)** The random network generated by the Erdős-Rényi model is rather homogeneous, i.e. most nodes have approximately the same number of links. **(f)** In contrast, a scale-free network is extremely inhomogeneous: while the majority of the nodes have one or two links, a few nodes have a large number of links, guaranteeing that the system is fully connected. To show this, we colored with red the five nodes with the highest number of links, and with green their first neighbors. While in the exonential network only 27% of the nodes are reached by the five most connected nodes, in the scale-free network more than 60% are, demonstrating the key role hubs play in the scale-free network. Note that both networks contain the same number of nodes and links, 130 and 430, respectively. After [26].

Scale-Free Model

Two mechanisms, absent from the classical random network models [25, 30], are responsible for the emergence of the power-law degree distribution . First, most networks *grow* through the addition of new nodes, that link to nodes already present in the system. Indeed, for example, the WWW or the scientific literature, two prototype scale-free networks, continuously expand by the addition of new nodes. Second, in most real networks there is a higher probability to link to a node with a large number of connections, a property called *preferential attachment*. Indeed, we link with higher probability to a more connected document on the WWW, or we tend to cite repeatedly

3

TABLE 1. The scaling exponents characterizing the degree distribution of several scale-free networks, for which $P(k)$ follows a power-law (1). We indicate the size of the network and its average degree $\langle k \rangle$. For directed networks we list separately the indegree (γ_{in}) and outdegree (γ_{out}) exponents, while for the undirected networks, marked with a star, these values are identical. Expanded after Ref. [1].

Network	Size	$\langle k \rangle$	γ_{out}	γ_{in}	Reference
WWW	325,729	4.51	2.45	2.1	[11]
WWW	4×10^7	7	2.38	2.1	[27]
WWW	2×10^8	7.5	2.72	2.1	[14]
WWW, site	260,000			1.94	[28]
Internet, domain*	3,015 - 4,389	3.42 - 3.76	2.1 - 2.2	2.1 - 2.2	[23]
Internet, router*	3,888	2.57	2.48	2.48	[23]
Internet, router*	150,000	2.66	2.4	2.4	[29]
Movie actors*	212,250	28.78	2.3	2.3	[30]
Coauthors, SPIRES*	56,627	173	1.2	1.2	[31]
Coauthors, neuro.*	209,293	11.54	2.1	2.1	[32]
Coauthors, math*	70,975	3.9	2.5	2.5	[32]
Sexual contacts*	2810		3.4	3.4	[24]
Metabolic, E. coli	778	7.4	2.2	2.2	[7]
Protein, S. cerev.*	1870	2.39	2.4	2.4	[8]
Ythan estuary*	134	8.7	1.05	1.05	[33]
Silwood park*	154	4.75	1.13	1.13	[33]
Citation	783,339	8.57		3	[15]
Phone-call	53×10^6	3.16	2.1	2.1	[34]
Words, conccurence*	460,902	70.13	2.7	2.7	[20]
Words, synonyms*	22,311	13.48	2.8	2.8	[19]
Protein, S. Cerev*	9,85	1.83	2.5	2.5	[35]
Comic Book Characters	6,486	14.9	0.66	3.12	[36]
E-mail	59,912	2.88	2.03	1.49	[37]
Protein Domains*	876	9.32	1.6	1.6	[38]
Prot. Dom. (PromDom)*	5995	2.33	2.5	2.5	[39]
Prot. Dom. (Pform)*	2478	1.12	1.7	1.7	[39]
Prot. Dom. (Prosite)*	13.60	0.77	1.7	1.7	[39]

much cited papers. These two ingredients, growth and preferential attachment, inspired the scale-free model that leads to a network with a power-law degree distribution. The algorithm of the scale-free model is the following [25, 30]:

(1) *Growth*: Starting with a small number (m_0) of nodes, at every timestep we add a new node with $m(\leq m_0)$ edges that link the new node to m different nodes already

4

present in the system.

(2) *Preferential attachment*: When choosing the nodes to which the new node connects, we assume that the probability Π that a new node will be connected to node i depends on the degree k_i of node i, such that

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}.\qquad(2)$$

Numerical simulations indicate that this network evolves into a scale-invariant state with the probability that a node has k edges follows a power-law with an exponent $\gamma = 3$ (Fig. 2). The scaling exponent is independent of m, the only parameter in the model.
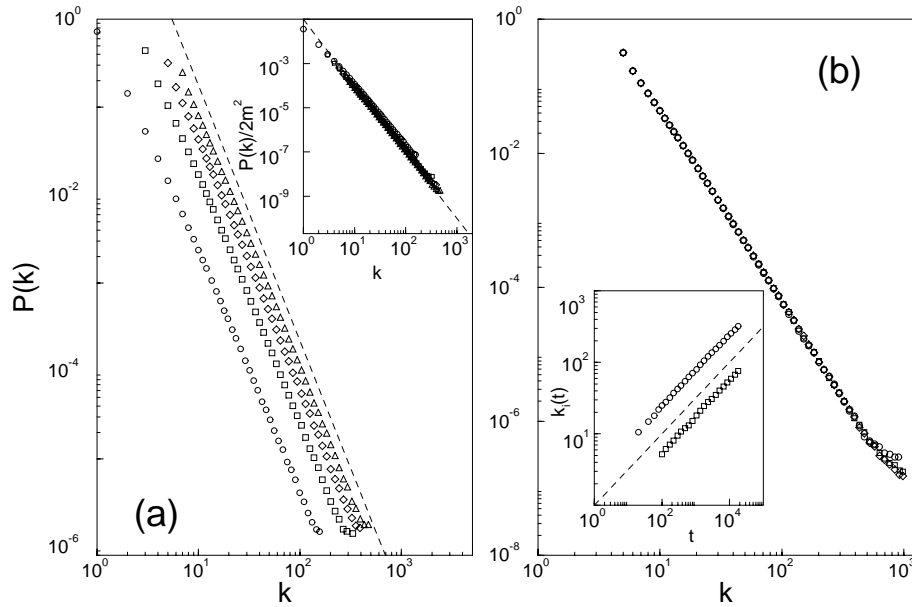

FIGURE 2. **(a)** Degree distribution of the scale-free model, with $N = m_0 + t = 300,000$ and $m_0 = m = 1$ (circles), $m_0 = m = 3$ (squares), $m_0 = m = 5$ (diamonds) and $m_0 = m = 7$ (triangles). The slope of the dashed line is $\gamma = 2.9$, providing the best fit to the data. The inset shows the rescaled distribution $P(k)/2m^2$ for the same values of m, the slope of the dashed line being $\gamma = 3$. **(b)** $P(k)$ for $m_0 = m = 5$ and system sizes $N = 100,000$ (circles), $N = 150,000$ (squares) and $N = 200,000$ (diamonds). The inset shows the time-evolution for the degree of two vertices, added to the system at $t_1 = 5$ and $t_2 = 95$. Here $m_0 = m = 5$, and the dashed line has slope 0.5, as predicted by Eq. (5). After [30].

The dynamical properties of the scale-free model can be addressed using various analytic approaches. The continuum theory proposed in [25, 30] focuses on the dynamics of node degrees. Widely used are the master equation approach of Dorogovtsev, Mendes and Samukhin [40] and the rate equation approach introduced by Krapivsky, Redner and Leyvraz [41]. Here we will focus on the continuum theory; for a discussion of the other methods see [1, 2].

Continuum theory: The continuum approach calculates the time dependence of the degree k_i of a given node i. This degree will increase every time a new node enters the

5

system and links to node i, the probability of this process being $\Pi(k_i)$. Assuming that k_i is a continuous real variable, the rate at which k_i changes is proportional to $\Pi(k_i)$. Consequently, k_i satisfies the dynamical equation

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}. \tag{3}$$

The sum in the denominator goes over all nodes in the system except the newly introduced one, thus its value is $\sum_j k_j = 2mt - m$, leading to

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}. \tag{4}$$

The solution of this equation, with the initial condition that each node i at its introduction has $k_i(t_i) = m$, is

$$k_i(t) = m\left(\frac{t}{t_i}\right)^\beta, \qquad \text{with} \quad \beta = \frac{1}{2}. \tag{5}$$

Equation (5) indicates that the degree of all nodes evolves the same way, following a power-law, the only difference being the intercept of the power-law.

Using (5), the probability that a node has a degree $k_i(t)$ smaller than k, $P(k_i(t) < k)$, can be written as

$$P(k_i(t) < k) = P\left(t_i > \frac{m^{1/\beta}t}{k^{1/\beta}}\right). \tag{6}$$

Assuming that we add the nodes at equal time intervals to the network, the t_i values have a constant probability density

$$P(t_i) = \frac{1}{m_0 + t_i}. \tag{7}$$

Substituting this into Eq. (6) we obtain that

$$P\left(t_i > \frac{m^{1/\beta}t}{k^{1/\beta}}\right) = 1 - \frac{m^{1/\beta}t}{k^{1/\beta}(t + m_0)}. \tag{8}$$

The degree distribution $P(k)$ can be obtained using

$$P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^{1/\beta}t}{m_0 + t}\frac{1}{k^{1/\beta + 1}}, \tag{9}$$

predicting that asymptotically ($t \to \infty$)

$$P(k) \sim 2m^{1/\beta}k^{-\gamma}, \quad \text{with} \quad \gamma = \frac{1}{\beta} + 1 = 3. \tag{10}$$

The scale-free and related models [25, 30, 40, 41, 42, 43, 44, 45] view networks as dynamical systems, assuming that they self-assemble and evolve in time through the

6

addition and removal of nodes and links. Such dynamical modelling attempts to capture what nature did when it assembled these networks, expecting that the structural elements and the topology will follow from these. Local decisions about the addition or removal of a link do not aim at global optimization, but try to gain some local advantage to a node such as, for example, enhancing the visibility of a webpage or the content of a scientific paper. The incompleteness of the information available to the local decision maker about the state of the full network, as well as the different interest driving the individual nodes, are the origin of the stochastic component in network evolution.

The scale-free model is the simplest example of an evolving network. In many systems, due to aging and saturation effects that limit the number of links a node can acquire, the preferential attachment function, $\Pi(k_i)$, can be nonlinear, following $\Pi(k_i) = f(k_i) \sum_j f(k_j)$, where $f(k)$ is an arbitrary function. Nonlinearities in $f(k)$ can result in deviations from the power-law in $P(k)$ [41]. Similarly, the addition and removal of nodes and links can be incorporated by including appropriate terms in $\Pi(k_i)$ [44, 45] changing the exponent γ or the power-law character of $P(k)$. Thus, in contrast with critical phenomena [46], the universal feature of most networks is not reflected by the power-law form of $P(k)$, or the value of the exponent γ. Most complex systems share, however, their dynamical, evolutionary character, captured within the framework provided by evolving networks, indicating that their topology and evolution cannot be divorced from each other.

HIERARCHICAL ORGANIZATION IN COMPLEX NETWORKS

In addition of being scale-free, measurements indicate that most networks display a high degree of clustering. Defining the clustering coefficient for node i with k_i links as $C_i = 2n_i/k_i(k_i - 1)$, where n_i is the number of links between the k_i neighbors of i, empirical results indicate that C_i averaged over all nodes is significantly higher for most real networks than for a random network of similar size [47, 1, 2]. Furthermore, the clustering coefficient of real networks is to a high degree independent of the number of nodes in the network (see Fig. 9 in [1]).

The scale-free property and clustering are not exclusive: for a large number of real networks, including metabolic networks [7, 48], the protein interaction network [8, 35], the world wide web [11] and even some social networks [49, 31, 32] the scale-free topology and high clustering coexist. Yet, most models proposed to describe the topology of complex networks have difficulty capturing simultaneously these two features. For example, the random network model [22, 21] cannot account neither for the scale-free, nor for the clustered nature of real networks, as it predicts an exponential degree distribution, and the average clustering coefficient, $C(N)$, decreases as N^{-1} with the number of nodes in the network. Scale-free networks, capturing the power law degree distribution, predict a much larger clustering coefficient than a random network. Indeed, numerical simulations indicate that for one of the simplest models [25, 30] the average clustering coefficient depends on the system size as $C(N) \sim N^{-0.75}$ [1, 2], significantly larger for large N than the random network prediction $C(N) \sim N^{-1}$. Yet, this prediction still disagrees with the finding that for several real systems C is independent of N [1].

7

In order to bring modularity, the high degree of clustering and the scale-free topology under a single roof, we need to assume that modules combine into each other in a hierarchical manner, generating what we call a *hierarchical network*. The presence of a hierarchy and the scale-free property impose strict restrictions on the number and the degree of cohesiveness of the different groups present in a network, which can be captured in a quantitative manner using a scaling law, describing the dependence of the clustering coefficient on the node degree.

Hierarchical Network Model

We start by constructing a hierarchical network model, that combines the scale-free property with a high degree of clustering. Our starting point is a small cluster of five densely linked nodes (Fig. 3a). Next we generate four replicas of this hypothetical module and connect the four external nodes of the replicated clusters to the central node of the old cluster, obtaining a large 25-node module (Fig. 3b). Subsequently, we again generate four replicas of this 25-node module, and connect the 16 peripheral nodes to the central node of the old module (Fig. 3c), obtaining a new module of 125 nodes. These replication and connection steps can be repeated indefinitely, in each step increasing the number of nodes in the system by a factor five.
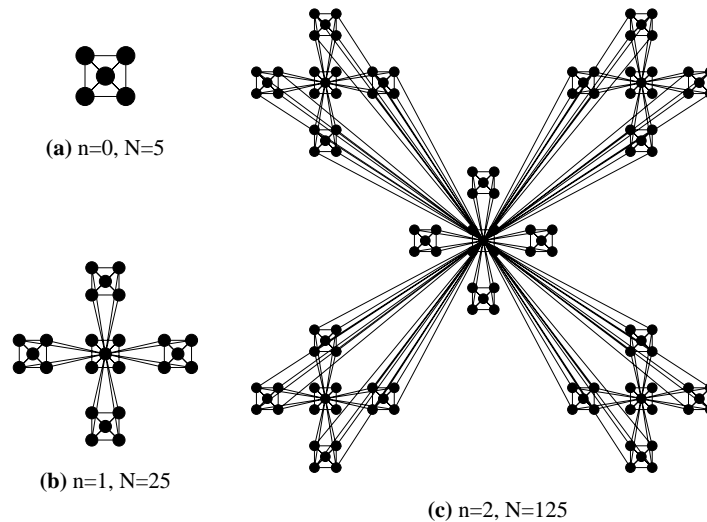

(a) n=0, N=5

(b) n=1, N=25

(c) n=2, N=125

FIGURE 3. The iterative construction leading to a hierarchical network. Starting from a fully connected cluster of five nodes shown in **(a)** (note that the diagonal nodes are also connected – links not visible), we create four identical replicas, connecting the peripheral nodes of each cluster to the central node of the original cluster, obtaining a network of $N = 25$ nodes **(b)**. In the next step we create four replicas of the obtained cluster, and connect the peripheral nodes again, as shown in **(c)**, to the central node of the original module, obtaining a $N = 125$ node network. This process can be continued indefinitely. After [50].

8

Precursors to the model described in Fig. 3 have been proposed in Ref. [51] and discussed in Ref. [52, 53] as a method of generating deterministic scale-free networks. In the following we argue that such hierarchical construction generates an architecture that is significantly different from the networks generated by traditional scale-free models.

First we note that the hierarchical network model seamlessly integrates a scale-free topology with an inherent modular structure. Indeed, the generated network has a power law degree distribution with degree exponent $\gamma = 1 + \ln 5/\ln 4 = 2.161$ (Fig. 4a). Furthermore, numerical simulations indicate that the clustering coefficient, $C \simeq 0.743$, is independent of the size of the network (Fig. 4c). Therefore, the high degree of clustering and the scale-free property are simultaneously present in this network.

The most important feature of the network model of Fig. 3, not shared by either the scale-free [25, 30] or random network models [22, 21], is its hierarchical architecture. The network is made of numerous small, highly integrated five node modules (Fig. 3a), which are assembled into larger 25-node modules (Fig. 3b). These 25-node modules are less integrated but each of them is clearly separated from the other 25-node modules when we combine them into the even larger 125-node modules (Fig. 3c). These 125-node modules are even less cohesive, but again will appear separable from their replicas if the network expands further.

This intrinsic hierarchy can be characterized in a quantitative manner using the recent finding of Dorogovtsev, Goltsev and Mendes [52] that in deterministic scale-free networks the clustering coefficient of a node with k links follows the scaling law

$$C(k) \sim k^{-1}. \tag{11}$$

This scaling law quantifies the coexistence of a hierarchy of nodes with different degrees of clustering, and applies to the model of Fig. 3a-c as well. Indeed, the nodes at the center of the numerous 5-node modules have a clustering coefficient $C = 1$. Those at the center of a 25-node module have $k = 20$ and $C = 3/19$, while those at the center of the 125-node modules have $k = 84$ and $C = 3/83$, indicating that the higher a node's degree the smaller is its clustering coefficient, asymptotically following the $1/k$ law (Fig. 4b). In contrast, for the scale-free model proposed in Ref. [25] the clustering coefficient is independent of k, i.e. the scaling law (11) does not apply (Fig. 4b). The same is true for the random [22, 21] or the various small world models [47, 54], for which the clustering coefficient is independent of the nodes' degree.

Therefore, the discrete model of Fig. 3 combines within a single framework the two key properties of real networks: their scale-free topology and high modularity, which results in a system-size independent clustering coefficient. Yet, the hierarchical modularity of the model results in the scaling law (11), which is not shared by the traditional network models.

Hierarchical Organization in Real Networks

To investigate if such hierarchical organization is present in real networks we measured the $C(k)$ function for several networks for which large topological maps are available. Next we discuss each of these systems separately.
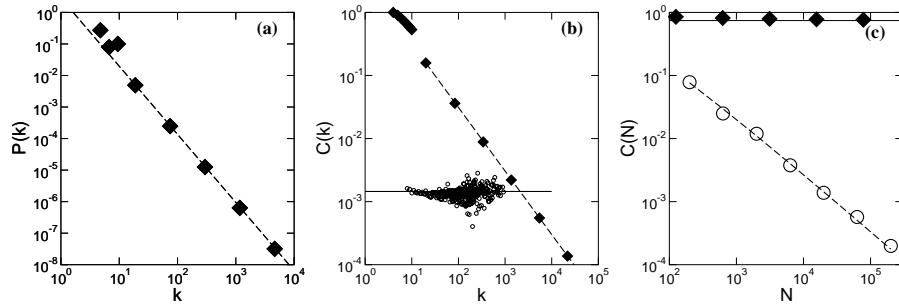
9

FIGURE 4. Scaling properties of the hierarchical model shown in Fig. 3 ($N = 5^7$). **(a)** The numerically determined degree distribution. The asymptotic scaling, with slope $\gamma = 1 + \ln 5 / \ln 4$, is shown as a dashed line. **(b)** The $C(k)$ curve for the model, demonstrating that it follows Eq. (11). The open circles show $C(k)$ for a scale-free model [25] of the same size, illustrating that it does not have a hierarchical architecture. **(c)** The dependence of the clustering coefficient, C, on the size of the network N. While for the hierarchical model C is independent of N (diamond), for the scale-free model $C(N)$ decreases rapidly (circle). After [50].

Actor Network: Starting from the `www.IMDB.com` database, we connect any two actors in Hollywood if they acted in the same movie, obtaining a network of 392,340 nodes and 15,345,957 links. Earlier studies indicate that this network is scale-free with an exponential cutoff in $P(k)$ for high k [25, 44, 43]. As Fig. 5a indicates, we find that $C(k)$ scales as k^{-1}, indicating that the network has a hierarchical topology.

Language network: Recently a series of empirical results have shown that the language, viewed as a network of words, has a scale-free topology [20, 19, 17, 18]. Here we study the network generated connecting two words to each other if they appear as synonyms in the Merriam Webster dictionary [19]. The obtained semantic web has 182,853 nodes and 317,658 links and it is scale-free with degree exponent $\gamma = 3.25$. The $C(k)$ curve for this language network is shown in Fig. 5b, indicating that it follows (11), suggesting that the language has a hierarchical organization.

World Wide Web: On the WWW two documents are connected to each other if there is an URL pointing from one document to the other one. The sample we study, obtained by mapping out the `www.nd.edu` domain [11], has 325,729 nodes and 1,497,135 links, and it is scale-free with degree exponents $\gamma_{out} = 2.45$ and $\gamma_{in} = 2.1$, characterizing the out and in-degree distribution, respectively. While the obtained $C(k)$, shown in Fig. 5c, does not follow as closely the scaling law (11) as observed in the previous two examples, there is clear evidence that $C(k)$ decreases rapidly with k, supporting the coexistence of many highly interconnected small nodes with a few larger nodes, which have a much lower clustering coefficient. Note that $C(k) \sim k^{-1}$ for the WWW was observed and briefly noted in Ref. [55].

Internet at the AS level: The Internet is often studied at two different levels of resolution. At the router level we have a network of routers connected by various physical communication links. At the interdomain or autonomous system (AS) level each administrative domain, composed of potentially hundreds of routers, is represented by a single node. As Fig. 5d shows, we find that at the domain level the Internet,
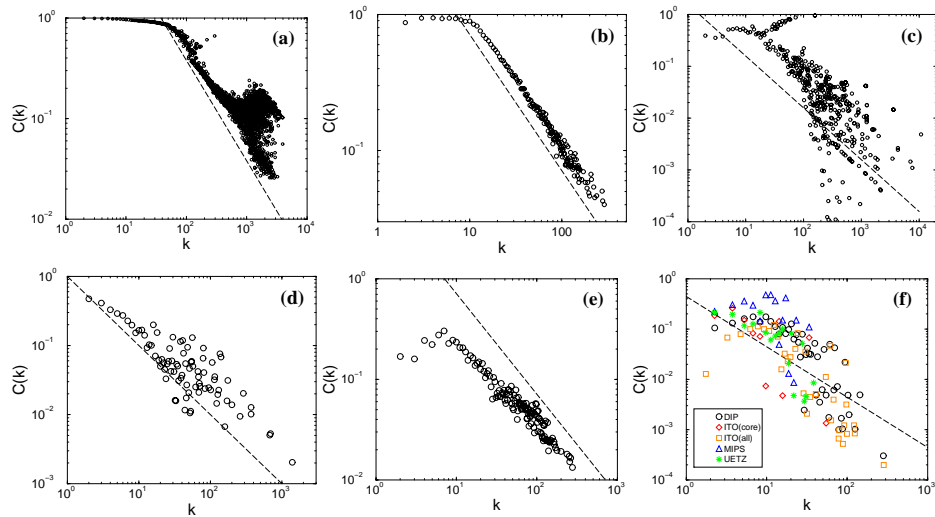
10

FIGURE 5. The scaling of $C(k)$ with k for four large networks: **(a)** Actor network, two actors being connected if they acted in the same movie according to the `www.IMDB.com` database. **(b)** The semantic web, connecting two English words if they are listed as synonyms in the Merriam Webster dictionary [19]. **(c)** The World Wide Web, based on the data collected in Ref. [11]. **(d)** Internet at the Autonomous System level, each node representing a domain, connected if there is a communication link between them. **(e)** The metabolic networks of 43 organisms with their averaged $C(k)$ curves. **(f)** The protein-protein physical interaction networks using four different databases [56, 57, 58, 59]. The dashed line in each figure has slope -1, following Eq. (11). After [50, 60, 61].

consisting of 65,520 nodes and 24,412 links [62], has a hierarchical topology as $C(k)$ is well approximated with (11). The scaling of the clustering coefficient with k for the Internet was earlier noted by Vazquez, Pastor-Satorras and Vespignani (VPSV) [63, 64], who observed $C(k) \sim k^{-0.75}$.

Metabolic networks: We measured the $C(k)$ function for the metabolic networks of 43 organisms [60]. As shown in Fig. 5e, for each organism $C(k)$ is well approximated by $C(k) \sim k^{-1}$.

Protein Interaction Network: To address the organization of protein interactions we studied the protein interaction network of *S. cerevisae*, based on four independent databases [56, 57, 58, 59]. The result, shown in Fig. 5, indicate that for each database the obtained network is heirarchical, $C(k)$ scaling as k^{-1} [61].

Our measurements indicate that some real networks lack a hierarchical architecture, and do not obey the scaling law (11). In particular, we find that the power grid and the router level Internet topology have a k independent $C(k)$ [65].

In summary, we offered evidence that for several large networks $C(k)$ is well approximated by $C(k) \sim k^{-1}$, in contrast to the k-independent $C(k)$ predicted by both the scale-free and random networks. This indicates that these networks have an inherently hierarchical organization. In contrast, hierarchy is absent in networks with strong geographical constraints, possibly because the limitation on the link length strongly con-

11

straints the network topology.

HALTING VIRUSES IN SCALE-FREE NETWORKS

After discussing the topological features of complex networks, we now turn to an application of these ideas to network dynamics, looking at the possibility of stopping virus spreading on scale-free networks. Classical epidemiological models predict that infectious diseases with transmission probability under an epidemic threshold will inevitably die out [66, 67]. Thus, short of a universally available cure, lowering the transmission probability represents an effective action against pandemics. Two recent results fundamentally revise this paradigm, however. First, Pastor-Satorras and Vespigniani [68] have shown that in scale-free networks the epidemic threshold is reduced to zero, that is, even extremely weakly infectious viruses spread and prevail. Second, Liljeros *et al.* [24] have found that the network of human sexual contacts has a scale-free topology. Taken together, these results indicate that reducing the transmission probability cannot eradicate sexually transmitted infectious diseases. That is, given the inhomogeneous, scale-free topology of the sexual web, short of a cure or vaccine available to all, the HIV virus will eventually reach the so far uninfected segments of the population exposed to the disease. The continued spreading of the HIV virus is remarkable because relatively effective therapies are available, that not only expand the lifetime of the infected individual, but also lower the transmission probability. The problem is that these expensive therapies are beyond reach in developing countries [69]. Given the limited number of cures available, can we use our improved understanding of the interplay between network topology and disease spreading to design allocation policies that restore the epidemic threshold, offering hope that the pandemic can be constrained?

Epidemics spread without a threshold on a scale-free network thanks to hubs. Once infected, hubs offer an efficient conduit for disease spreading by reaching an unusually high percentage of other nodes [68]. Random immunization cannot restore the epidemic threshold, as it leaves the scale-free nature of the network unaltered. However, one can show that immunizing all hubs with degree larger than a given connectivity k_0 restores the finite epidemic threshold. Indeed, the origin of the zero threshold is the infinite variance [70] of the power law distribution 1. If nodes with $k > k_0$ links do not transmit the disease, the variance is finite, which results in a finite epidemic threshold [71] This indicates that the most effective response to an epidemic must focus all resources on reaching as many hubs as economically feasible. This would also be the most cost-effective, as hubs are relatively rare in scale-free networks. The problem with this optimal policy is that we cannot effectively identify the hubs, because the number of sexual partners for individuals is unknown. Here we study a model that incorporates our limited ability to identify the hubs, assuming that the likelihood of providing cure to a node with k sexual partners in a given time frame increases as k^α. In this model $\alpha = 0$ corresponds to random immunization, which is expected to have zero epidemic threshold, while $\alpha = \infty$ corresponds to the optimal policy that treats all hubs with degree larger than k_0. Thus any finite α describes a policy which focuses its resources to treat at a higher rate the more connected members of the sexual web. To study the effect of such

12

a selective policy we use the susceptible-infected-susceptible (SIS) model, which offers a simple phenomenological description of epidemic spreading in a complex network [68] We assume that the disease spreads on a scale-free network with degree exponent [25] $\gamma = 3$ within the range of the experimentally identified values for the sexual web [24] . Each node is infected at a rate ν if it has a link to an infected node and infected nodes are cured and become susceptible again to the disease with a probability $\delta = \delta_0 k^\alpha$, defining the effective spreading rate as $\lambda \equiv \frac{\nu}{\delta_0}$. The analytical calculations, supported by numerical simulations, indicate that the epidemic threshold varies as $\lambda_c = \alpha m^{(\alpha-1)}$ (Fig. 6). Thus for $\alpha = 0$ we recover $\lambda_c = 0$ confirming that random immunization cannot eradicate sexually transmitted infectious diseases. The good news is that for any nonzero α we obtain that $\lambda_c \neq 0$, indicating that policies that attempt to identify and preferentially treat the hubs restore the epidemic threshold. The fact that the threshold is restored for *any* $\alpha > 0$ value indicates that even modestly effective attempts to uncover and treat hubs, characterized by small α, if carried out systematically, are more successful than policies based on large-scale but random distribution of the available treatments.

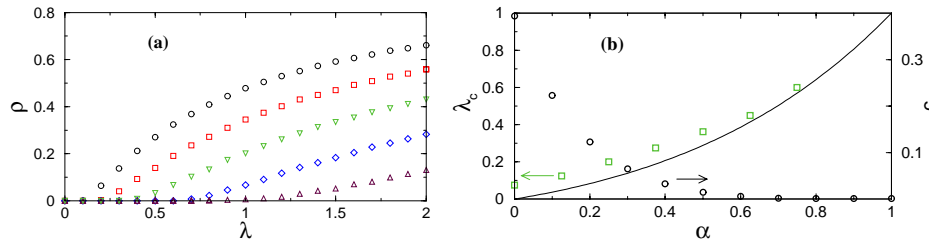**FIGURE 6.** **Curing the hubs. (a)** Prevalence, ρ, measured as the fraction of infected nodes in function of the effective spreading rate λ for $\alpha = 0$ (circle), 0.25 (square), 0.50 (triangle down), 0.75 (diamond) and 1 (triangle up), as predicted by Monte-Carlo simulations using the SIS model on a scale-free [25] network with N=10,000 nodes. While for $\alpha = 0$ the epidemic threshold is zero, a nonzero α leads to the emergence of a finite epidemic threshold. **(b)** The dependence of the epidemic threshold λ_c on α as predicted by our calculations (continuous line) based on the continuum approach described in Ref. [68], and by the numerical simulations based on the SIS model (green boxes). The small deviation between the numerical results and the analytical prediction is due to the uncertainty in determining the precise value of the threshold in Monte-Carlo simulations. The vertical axis on the r.h.s. labels the number of cures, c, administered in an unit time per node for different values of α, shown as black circles on the figure. The rapidly decaying c indicates that more successful is a policy in selecting and curing hubs (larger is α), fewer cures are required for a fixed spreading rate ($\lambda = 0.75$). The data points in (a) and (b) are averaged over 10 independent runs. After [72].

Such selective cures are particularly important when successful therapies are limited by the absence of financial resources. Our results indicate that under such conditions it is most desirable to make the best efforts to uncover and cure individuals with a high number of sexual partners, whose reduced ability to spread the disease will drastically lower the disease rate within the whole population. Indeed, we find that within the model the number of administered cures decreases rapidly as α increases (Fig.1b), indicating that treating the hubs is also the financially most responsible policy. Note that similar conclusions have been reached by several independent studies [73].

13

OUTLOOK

The fact that many large networks are scale-free is now well established. It is also clear that most networks have a modular topology, quantified by the high clustering coefficient they display. Such modules have been proposed to be a fundamental feature of biological systems [4, 60], but have been discussed in the context of the WWW [12, 74], and social networks as well [75, 76]. The hierarchical topology offers a new avenue for bringing under a single roof these two concepts, giving a precise and quantitative meaning for the network's modularity. It indicates that we should not think of modularity as the coexistence of relatively independent groups of nodes. Instead, we have many small clusters, that are densely interconnected. These combine to form larger, but less cohesive groups, which combine again to form even larger and even less interconnected clusters. This self-similar nesting of different groups or modules into each other forces a strict fine structure on real networks.

The presence of such a hierarchical architecture reinterprets the role of the hubs in complex networks. Hubs, the highly connected nodes at the tail of the power law degree distribution, are known to play a key role in keeping complex networks together, playing a crucial role from the robustness of the network [77, 78] to the spread of viruses in scale-free networks [68]. Our measurements indicate that the clustering coefficient characterizing the hubs decreases linearly with the degree. This implies that while the small nodes are part of highly cohesive, densely interlinked clusters, the hubs are not, as their neighbors have a small chance of linking to each other. Therefore, the hubs play the important role of bridging the many small communities of clusters into a single, integrated network.

While it is difficult to identify such universal characteristics from single examples, once they are uncovered, they offer strong support for an emerging theme: networks in nature are far from being random, but they evolve following robust self-organizing principles and evolutionary laws that cross disciplinary boundaries. Progress is possible only if the numerical and analytical work is combined with empirical studies on real networks, potentially opening an unexpectedly revealing window on the structure of complex systems. The results reviewed here likely represent only the tip of the iceberg, and systematic data driven studies focusing on the topology and evolution of real networks could fundamentally change how we approach the complex world around us.

ACKNOWLEDGMENTS

REFERENCES

1. Albert, R., and Barabási, A.-L., *Rev. Mod. Phys.*, **74**, 67–97 (2002).
2. Dorogovtsev, S. N., and Mendes, J. F. F., *Adv. Phys.*, **51**, 1079 (2002).

3. Schuster, H. G., *Complex Adaptive Systems*, Scator Verlag, Saarbruskey, Germany, 2002.
4. Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W., *Nature*, **402**, 47 (1999).
5. Overbeek, R., and et al., *Nucleic Acids Res.*, **28**, 123 (2000).
6. Karp, P. D., and et al., *Nucleic Acids Res.*, **28**, 56 (2000).
7. Jeong, H., Tombor, B., Albert, R., Oltvai, Z., and Barabási, A.-L., *Nature*, **407**, 651 (2000).
8. Jeong, H., Mason, S., Barabási, A.-L., and Oltvai, Z. N., *Nature*, **411**, 41 (2001).
9. Kochen, M., editor, *The Small World*, Ablex, Norwood, NJ, 1989.
10. Wasserman, S., and Faust, K., *Social Network Analysis: Methods and Applications*, Cambridge University, Cambridge, 1994.
11. Albert, R., Jeong, H., and Barabási, A.-L., *Nature*, **401**, 130 (1999).
12. Lawrence, S., and Giles, C. L., *Nature*, **400**, 107 (1999).
13. Kleinberg, J., Kumar, S. R., Raghavan, P., Rajagopalan, S., and Tomkins, A., *Proc. of the Int. Conf. on Combinatorics and Computing* (1999).
14. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajalopagan, S., Stata, R., Tomkins, A., and Wiener, J., *Comput. Netw.*, **33**, 309 (2000).
15. Redner, S., *European Phys. J.*, **B4**, 131135 (1998).
16. de Solla Price, D. J., *Science*, **169**, 510 (1965).
17. Sigman, M., and Cecchi, G., *Proc. Nat. Acad. Sci.*, **99**, 1792 (2002).
18. Dorogovtsev, S. N., and Mendes, J. F. F., *Proc. Reg. Soc. London B*, **268**, 2603 (2001).
19. Yook, S., Jeong, H., and Barabási, A.-L., Preprint (to be published) (2002).
20. Ferrer i Cancho, R., and Solé, R. V., *Proc. Reg. Soc. London B*, **268**, 2261 (2001).
21. Bollobás, B., *Random Graphs*, Academic Press, London, 1985.
22. Erdős, P., and Rényi, A., *Publ. Math. Debrecen*, **6**, 290 (1959).
23. Faloutsos, M., Faloutsos, P., and Faloutsos, C., *Comput. Commun. Rev.*, **29**, 251 (1999).
24. Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., and Åberg, Y., *Nature*, **411**, 907 (2001).
25. Barabási, A.-L., and Albert, R., *Science*, **286**, 509 (1999).
26. Barabási, A.-L., *Physics World*, **July**, 33 (2001).
27. Kumar, R., Raghavan, P., Rajalopagan, S., and Tomkins, A., *Proceedings of the 9th ACM Symposium on Principles of Database Systems*, p. 1 (1999).
28. Huberman, B. A., and Adamic, L. A., *Nature*, **401**, 131 (1999).
29. Govindan, R., and Tangmunarunkit, H., *Proceedings of IEEE INFOCOM 2000, Tel Aviv, Israel,* (IEEE, Piscataway, N. J.), **3**, 1371 (2000).
30. Barabási, A.-L., Albert, R., and Jeong, H., *Physica A*, **272**, 173 (1999).
31. Newman, M. E. J., *Phys. Rev. E*, **64**, 016131 (2001).
32. Barabási, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., and Vicsek, T., *Physica A*, **311**, 590 (2002).
33. Montoya, J. M., and Solé, R. V., *T. Theor. Biol.*, **216**, 605 (2002).
34. W., F. C., and Lu, L., *Proceedings of the 32nd ACM Symposium on the Theory of Computing* (ACM, New York), p. 271 (2000).
35. Wagner, A., *Mol. Biol. Evol.*, **18**, 1283 (2001).
36. Alberich, R., Miro-Julia, J., and Rosselló, F., *Marvel Universe looks almost like a real social network*, Los Alamos Archive, cond-mat/0202174, (2002).
37. Ebel, H., Mielsch, L. I., and Bormholdt, S., *Phys. Rev. E*, **66**, 035103 (2002).
38. Park, J., Lappe, M., and Teichmann, A., *J. Mol. Biol.*, **307**, 929 (2001).
39. Wuchty, S., *Mol. Biol. Evol.*, **18**, 1699 (2001).
40. Dorogovtsev, S. N., Mendes, J. F. F., and Samukhin, A. N., *Phys. Rev. Lett.*, **85**, 6633 (2000).
41. Krapivsky, P. L., Redner, S., and Leyvraz, F., *Phys. Rev. Lett.*, **86**, 5901 (2000).
42. Dorogovtsev, S. N., and Mendes, J. F. F., *Phys. Rev. E*, **62**, 1862 (2000).
43. Amaral, L. A. N., Scala, A., Barthélémy, M., and Stanley, H. E., *Proc. Nact. Acad. Sci. U.S.A*, **97**, 11149 (2000).
44. Albert, R., and Barabási, A.-L., *Phys. Rev. Lett.*, **85**, 5234 (2000).
45. Dorogovtsev, S. N., and Mendes, J., *Europhys. Lett.*, **52**, 33 (2000).
46. Stanley, H., *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, New York, 1971.
47. Watts, D. J., and Strogatz, S. H., *Nature*, **393**, 440 (1998).
48. Wagner, A., and Fell, D. A., *Proc. R. Soc. Lond. B. Biol. Sci.*, **268**, 1803 (2001).

49. Newman, M. E. J., *Proc. Nat. Acad. Sci. U.S.A.*, **98**, 404 (2001).
50. Ravasz, E., and Barabási, A.-L., *Phys. Rew. E*, **(in press)** (2002).
51. Barabási, A.-L., Ravasz, E., and Vicsek, T., *Physica A*, **299**, 559 (2001).
52. Dorogovtsev, S. N., Goltsev, A. V., and Mendes, J. F. F., *Phys. Rev. E*, **65**, 066122 (2002).
53. Jung, S., Kim, S., and Kahng, B., *Phys. Rev. E*, **65**, 056101 (2002).
54. Newman, M. E. J., *J. Stat. Phys.*, **101**, 819 (2000).
55. Eckmann, J.-P., and Moses, E., *Proc. Nact. Acad. Sci. U.S.A*, **99**, 5825 (2002).
56. Xenarios, I., Rice, D. W., Salwinski, L., Baron, M. K., Marcotte, E. M., and Eisenberg, D. M. S., *Nucl. Acids. Res.*, **28**, 289 (2000).
57. Mewes, H. W., Frishman, D., Güldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Münsterkötter, M., Rudd, S., and Weil, B., *Nucl. Acids. Res.*, **30**, 31 (2002).
58. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., and et al., P. P., *Nature*, **403**, 623 (2000).
59. Tashiro, T., Ito, K. Muta, R. O., K., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S., and Sakaki, Y., *PNAS*, **99**, 1143 (2000).
60. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabási, A.-L., *Science*, **297**, 1551 (2002).
61. Yook, S. H., and et al. (2002), in preparation.
62. On-line source, `http://moat.nlanr.net/infrastructure.html`.
63. Vázquez, A., Pastor-Satorras, R., and Vespignani, A., *Phys. Rev. E*, **65**, 066130 (2002).
64. Vázquez, A., Pastor-Satorras, R., and Vespignani, A., *Internet topology at the router and autonomous system level*, Los Alamos Archive, cond-mat/0206084, (2002).
65. Ravasz, E., and Barabási, A.-L., *Hierarchical organization in complex networks*, Los Alamos Archive, cond-mat/0206130, (2002).
66. Nowak, M. A., and May, R. M., *Virus dynamics*, Oxford University Press, Oxford, 2000.
67. May, R. M., and Anderson, R. M., *Phil. Trans. R. Soc. London B*, **321**, 565 (1988).
68. Pastor-Satorras, R., and A. Vespignani, A., *Phys. Rev. Lett.*, **86**, 3200 (2001).
69. Piot, P., Bartos, M., Ghys, P. D., Walker, N., and Schwartlander, B., *Nature*, **410**, 986 (2001).
70. Avraham, D. B., and Havlin, S., *Diffusion and Reactions in Fractals and Disordered Systems*, Cambridge University Press, Cambridge, 2000.
71. Lloyd, A. L., and May, R. M., *Science*, **292**, 1316 (2001).
72. Dezső, Z., and Barabási, A.-L., *Phys. Rev. E.*, **65**, 055103 (2002).
73. Pastor-Satorras, R., and Vespignani, A., *Phys. Rev. E.*, **65**, 036104 (2002).
74. Flake, G. W., Lawrence, S., and Giles, C. L., *Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining,* (ACM, Boston, MA, USA), p. 150 (2000).
75. Granovetter, M. S., *Am. J. Sociol.*, **78**, 1360 (1973).
76. Watts, D. J., Dodds, P. S., and Newman, M. E. J., *Science*, **296**, 130 (2002).
77. Albert, R., Jeong, H., and Barabási, A.-L., *Nature*, **406**, 378 (2000).
78. Cohen, R., Erez, K., ben Avraham, D., and Havlin, S., *Phys. Rev. Lett.*, **86**, 3682 (2001).