

Emergence of scale-free networks from local connectivity and communication trade-offsValmir C. Barbosa,¹ Raul Donangelo,² and Sergio R. Souza²¹*Programa de Engenharia de Sistemas e Computação, COPPE, Universidade Federal do Rio de Janeiro, Caixa Postal 68511, 21941-972 Rio de Janeiro, Rio de Janeiro, Brazil*²*Instituto de Física, Universidade Federal do Rio de Janeiro, Caixa Postal 68528, 21941-972 Rio de Janeiro, Rio de Janeiro, Brazil*
(Received 21 December 2005; revised manuscript received 29 May 2006; published 20 July 2006)

We suggest a mechanism of connectivity evolution in networks to account for the emergence of scale-free behavior. The mechanism works on a fixed set of nodes and promotes growth from a minimally connected initial topology by the addition of edges. A new edge is added between two nodes depending on the trade-off between a gain and a cost function of local connectivity and communication properties. We report on simulation results that indicate the appearance of power-law distributions of node degrees for selected parameter combinations.

DOI: [10.1103/PhysRevE.74.016113](https://doi.org/10.1103/PhysRevE.74.016113)

PACS number(s): 89.75.Hc, 05.65.+b, 89.75.Da, 89.75.Fb

I. INTRODUCTION

The topology of large-scale complex networks such as the Internet and the Worldwide Web (WWW) is, in general, not known. The study of such networks has then relied on modeling them as random graphs [1], and particularly on focusing almost exclusively on the distribution of node degrees. Unlike the classic case pioneered by Erdős and Rényi [2], for the Internet, the WWW, and several other networks, it appears that degrees are distributed according to a power law, not a Poisson distribution. That is, the probability that a randomly chosen node has degree k is proportional to $k^{-\tau}$ in general with $2 < \tau < 3$ [3,4].

These findings are based on probe samplers in the case of both the Internet [5] and the WWW [6], and are generally regarded as reasonably accurate. However, part of the underlying machinery has been recently proven somewhat unreliable in the case of the Internet. For example, it has been demonstrated experimentally that the usual mechanism of inferring breadth-first-search trees from the probe results can underestimate the value of τ significantly when the graph does have a power-law degree distribution [7]. Likewise, it is possible to argue formally that such a mechanism can in some cases lead to the conclusion of a power-law degree distribution when in fact the graph's degrees are distributed in some other way [8].

In recent years, and notwithstanding these limitations, considerable effort has been put into discovering mechanisms of network growth that give rise to a power-law degree distribution. Especially noteworthy is the mechanism of preferential attachment, which underlies the so-called Barabási-Albert model [9,10], as well as variations [11–14] and generalizations [15] thereof. Preferential attachment is the policy whereby a new edge is added to the network between a new node and a preexisting one with probability proportional to how many edges are already incident on the preexisting node, that is, its current degree. The generalization of [15] incorporates both this policy and also the copying mechanism of [16]. We refer the reader to [17] for a review of the essential mathematical results related to these models.

While the study of complex networks from the perspective of node-degree distributions seems sound and has given rise to important discoveries related to global properties, such as the nature and size of a network's connected compo-

nents and diameter [18], explaining the formation of the network from the same perspective (e.g., by evoking preferential attachment) is unreasonable for at least two reasons. The first is that the addition of a particular edge most definitely does not depend on global properties such as the distribution of node degrees at the time of expansion. The second reason is that, at least for computer networks like the Internet, it makes no sense to assume that the degree distribution, rather than some cost- or performance-related entity, is the essential driving force behind the evolution of the network's topology.

Models that depend on node-degree distributions are then adequate descriptive models, in the sense that they give rise to the desired power-law functional form, but constitute poor generative models. This has also been recognized elsewhere from various perspectives (cf., e.g., [19–21]), and has almost always resulted in the appearance of alternative models, such as the ones in [22,23], that are also dependent on global properties (like one-to-all distances) and therefore seem implausible as well. One exception has been the copying model of [16], which attempts to explain the growth of the WWW by mixing random attachments with attachments that essentially copy those of a randomly selected “prototype” node. Another one is the model of [19], where each node is assigned a “fitness” that is local to it and then the decision on whether to interconnect two nodes depends only on their fitnesses. But this model is not quite realistic, since a node's fitness remains unchanged as new connections are attempted, regardless of how many edges are already incident on it.

We work on the premise that networks such as the Internet or the WWW, although fast growing, appear not to acquire new nodes fast enough to impact their main topological properties significantly. Thus the model that we study is targeted at the evolution of the connectivity of computer networks, and promotes network growth on a fixed set of nodes by incrementally adding edges between nodes as the result of comparing a gain function and a cost function for each edge addition. If i and j are nodes not currently connected by an edge in the network, the gain incurred when adding an edge between them depends only on the immediate neighborhoods of i and j and on the current distance between i and j . The cost of the addition, in turn, is also dependent solely upon i and j and seeks to reflect both the cost of deploying the communications link itself and the cost of upgrading nodes i and j 's connection capabilities to accommodate the new link.

The edge joining i and j is added to the network if the gain surpasses the cost.

We point out, before proceeding to a detailed description of our model, that it shares with the model of [24] both the property of assuming a fixed set of nodes and the property of having all decisions depend solely on locally available information. On the other hand, that model operates also on a fixed set of edges and targets only the rewiring of those edges for the refinement of each node's estimates of distances to all others. The two models are then substantially distinct from each other: ours not only considers the addition of edges but also relies, for its decisions, on quantities other than mere distance estimates.

II. THE MODEL

We model the evolution of network connectivity as the sequence G^0, G^1, \dots of undirected graphs, all having the same set of n nodes. We assume that G^0 is a tree that spans all the nodes; G^0 is therefore connected and has $n-1$ edges. For $t \geq 0$, G^{t+1} is obtained from G^t by randomly selecting two nodes, say i and j , that are not directly connected by an edge, and then adding an edge between them if the gain incurred with the addition of the edge is greater than its cost. Otherwise, we simply let $G^{t+1} = G^t$. All graphs in the sequence are then guaranteed to be connected and to remain free of multiple edges and self-loops. We let d_{ij}^t denote the distance between i and j in G^t , and n_i^t the degree of node i in G^t . We also let $N_i^t(j)$ be the set comprising every neighbor k of node i in G^t for which $d_{jk}^t > 2$, and similarly N_{ij}^t be the set of unordered node pairs (k, l) such that either k or l is a neighbor of i , the other node in the pair is a neighbor of j , and furthermore $d_{kl}^t > 3$.

Let g_{ij}^t denote the gain incurred with the addition of an edge between i and j to G^t when $d_{ij}^t > 1$. In our model, we let g_{ij}^t be some upper bound on the number of edges by which distances between certain nodes become shorter after the addition of that edge. The distances we consider to establish this upper bound are some of those that involve i or j directly, or yet nodes in their immediate neighborhoods in G^t . Specifically, we consider d_{ij}^t, d_{ik}^t for $k \in N_i^t(j)$ (neighbors of j that are more than two edges away from i in G^t), d_{jk}^t for $k \in N_i^t(j)$ (neighbors of i that are more than two edges away from j in G^t), and finally d_{kl}^t for $(k, l) \in N_{ij}^t$ (node pairs that are more than three edges away from each other in G^t , one being a neighbor of i , the other a neighbor of j).

Upper bounds on each distance in the latter three groups are, clearly, $d_{ij}^t + 1$, $d_{ij}^t + 1$, and $d_{ij}^t + 2$, respectively. An upper bound on the sum of all distances considered is then

$$d_{ij}^t + (d_{ij}^t + 1)|N_j^t(i)| + (d_{ij}^t + 1)|N_i^t(j)| + (d_{ij}^t + 2)|N_{ij}^t|, \quad (1)$$

where we use $|X|$ to denote the cardinality of set X . The addition of an edge joining i and j causes the sum of all these distances to become

$$1 + 2|N_j^t(i)| + 2|N_i^t(j)| + 3|N_{ij}^t|, \quad (2)$$

and consequently the overall number of edges by which the distances become shorter is at most

$$g_{ij}^t = (d_{ij}^t - 1)[1 + |N_j^t(i)| + |N_i^t(j)| + |N_{ij}^t|]. \quad (3)$$

One crucial aspect of the gain expressed in (3) is that, in the context of computer networks, it depends exclusively on information that can be obtained by tracing routes on G^t . This certainly holds for the determination of d_{ij}^t , and holds also for determining the sets $N_j^t(i)$, $N_i^t(j)$, and N_{ij}^t , provided only that the process of tracing routes is controlled for constant depth. However, given the nature of routing algorithms such as those of the Internet [25], tracing a route between i and j on G^t is only guaranteed to provide an upper bound on d_{ij}^t , which is nonetheless consonant with the expression in (3) being itself an upper bound on total distance improvement.

Besides g_{ij}^t , the decision regarding the addition of an edge between i and j depends also on the cost of this addition. We denote this cost by c_{ij}^t and define it in such a way that both the cost of deploying a communications link and the cost of possibly upgrading the connection capabilities of i or j are taken into account. The former of these we denote by C and assume to be independent of t, i , and j .

As for the latter of the two cost components, we assume that the number of connections a node can sustain at any time is at most $\lceil \alpha^z \rceil$ for some fixed $\alpha > 1$ and some $z \in \{0, 1, \dots\}$ that does not decrease as time elapses (we use $\lceil x \rceil$ to denote the least integer that is no less than x). If the degree of i or j in G^t is precisely such a maximum number of connections, then the cost of connecting i to j directly involves the cost of upgrading the connection capabilities of i or j , as the case may be, to $\lceil \alpha^{z+w} \rceil$, where w is the least integer for which $\lceil \alpha^{z+w} \rceil > \lceil \alpha^z \rceil$. We further assume, for some fixed $\beta > 1$, that the cost of endowing the node with the capability of connecting to $\lceil \alpha^z \rceil$ other nodes is proportional to β^z .

Let n_k denote the number of connections of some node k . We model the scenario in which the cost incurred with the upgrade of n_k from $\lceil \alpha^z \rceil$ to $\lceil \alpha^{z+w} \rceil$ is proportional to $\beta^{z+w} - \beta^z$ (only the cost difference is paid) and is furthermore amortized along the deployment of each new connection [as opposed to being paid in full when the $(\lceil \alpha^z \rceil + 1)$ st connection is deployed]. If we let $f(n_k)$ be the cost portion to be incurred when the number of connections is n_k and for simplicity disregard the fact that connections necessarily occur in discrete numbers, then it follows that

$$\int_{\alpha^z}^{\alpha^{z+w}} f(n_k) dn_k \propto \beta^{z+w} - \beta^z \quad (4)$$

$$= (\alpha^{z+w})^{\log_\alpha \beta} - (\alpha^z)^{\log_\alpha \beta}. \quad (5)$$

Consequently,

$$f(n_k) \propto n_k^{\log_\alpha(\beta/\alpha)}. \quad (6)$$

Setting $\alpha = \beta$ leads $f(n_k)$ to be constant with respect to n_k ; setting $\alpha \neq \beta$ leads $f(n_k)$ to vary either directly ($\alpha < \beta$) or inversely ($\alpha > \beta$) with n_k .

We then have

$$c_{ij}^t = C + D[(n_i^t)^\gamma + (n_j^t)^\gamma] \quad (7)$$

for some constant D and $\gamma = \log_\alpha(\beta/\alpha)$. An edge is added to G^t between nodes i and j to yield G^{t+1} if $g_{ij}^t > c_{ij}^t$. If not, then

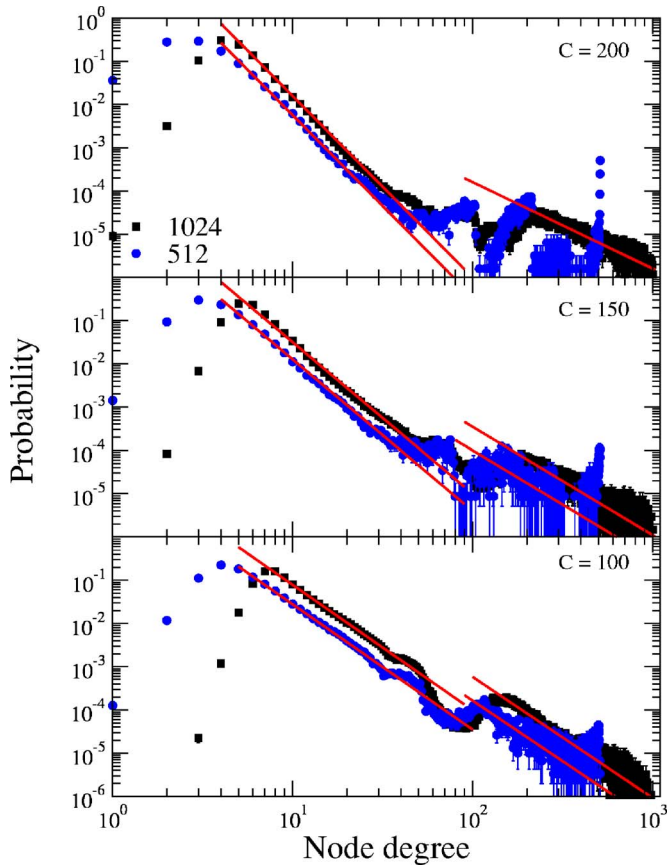


FIG. 1. (Color online) Average node-degree distributions for $n=512, 1024$, $D=0.1$, and $\gamma=0.9$. The degree distribution approaches a power law proportional to $k^{-\tau}$. For values of D and γ as given, τ may be parametrized as either $\tau=2+0.007C+0.00002C^2$, in the region $k < 100$, or $\tau=3.4-0.003C-0.00002C^2$, for $k > 100$.

$G^{t+1}=G^t$. By the nature of Eqs. (3) and (7), this decision involves only the distance between i and j in G^t , in addition to other quantities that depend exclusively on the surroundings of i and j within a constant radius in G^t . It is then essentially a local decision.

III. COMPUTATIONAL RESULTS

We have conducted computer simulations for selected combinations of the C , D , and γ parameters. Each simulation starts with a randomly chosen instance of G^0 and proceeds through $t=3000n$. A G^0 instance is generated on the n initially isolated nodes by progressively selecting node pairs at random and directly interconnecting them if no path exists between them; because G^0 is a tree, it is necessary and sufficient that $n-1$ such interconnections be performed.

At each step of a simulation the distance d_{ij}^t must be calculated on G^t . While on a real computer network such a distance (or an upper bound thereof) is readily available from the network's routing structure (as noted earlier), calculating d_{ij}^t seems to be asymptotically no easier than finding the distances between a given node and all others in G^t . For connected graphs, this requires $O(m^t)$ time [26], where we use m^t to denote the number of edges of G^t . It is therefore a

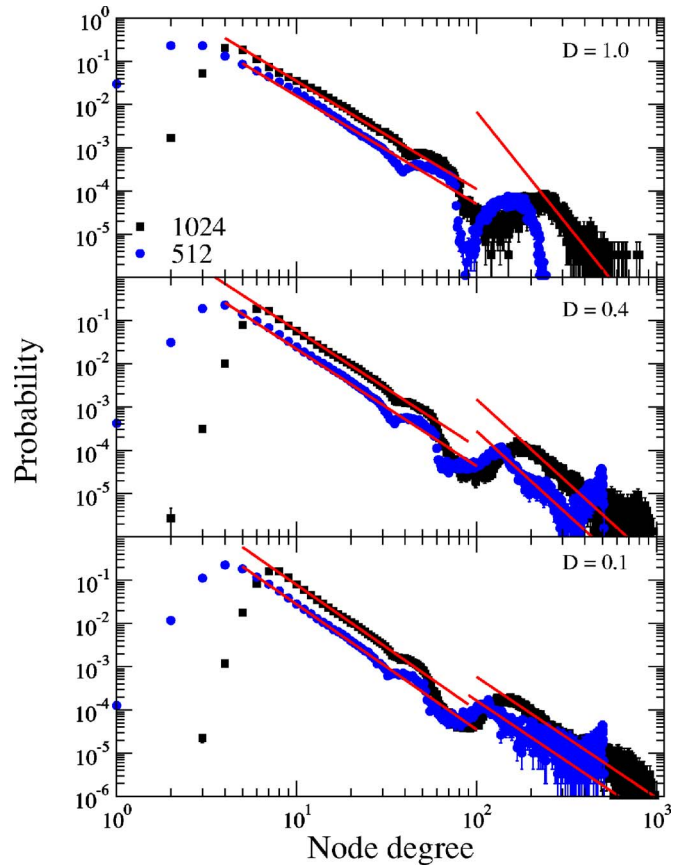


FIG. 2. (Color online) Average node-degree distributions for $n=512, 1024$, $C=100$, and $\gamma=0.9$. The values of τ may be approximated by either $\tau=2.98-0.85D+0.37D^2$, for $k < 100$, or $\tau=2.57+3.37D-0.74D^2$, in the region $k > 100$.

time-consuming procedure, and progressively more so as the simulation is carried on and the graph tends to become denser. The consequence of this for the present study is that the value of n is somewhat limited, and so is the number of independent G^0 instances that can be used for statistical significance.

Our results are shown in Figs. 1–3, where, respectively, the value of each of C , D , and γ is varied while the other two parameters remain fixed at a set of common values ($C=100$, $D=0.1$, and $\gamma=0.9$). For each combination of the three parameters we show results for two values of n , in each case as averages over 500 independent simulations. As the figures indicate, our model for network growth does indeed give rise to a scale-free pattern of behavior in which a vast majority of the nodes has low degrees while a few high-degree nodes are nonetheless present.

IV. DISCUSSION

Except for the highest γ values in our simulations ($\gamma=1.5, 1.9$; see Fig. 3), the node-degree distribution that results from our model seems to settle at two distinct power-law regimes, one for node degrees below roughly 100, the other for those above this threshold. The reason why this happens depends on the process whereby nodes acquire ever higher degrees, as we discuss next.

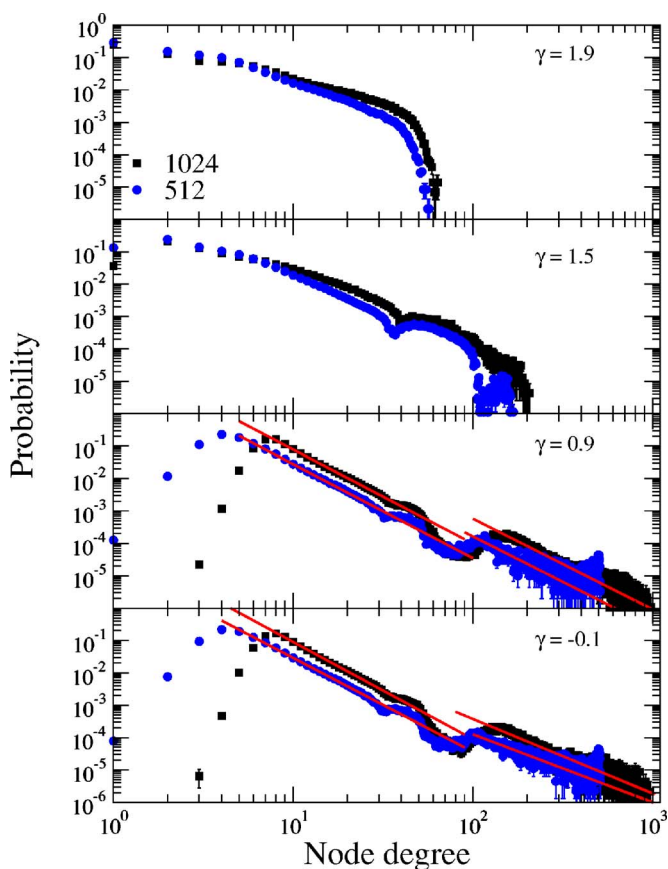


FIG. 3. (Color online) Average node-degree distributions for $n=512, 1024, C=100,$ and $D=0.1$. In this case, the values of τ are given by either $\tau=3+0.1\gamma-0.25\gamma^2$, for $k < 100$, or $\tau=2.3+0.067\gamma-0.67\gamma^2$, for $k > 100$.

In our model, nodes acquire higher degrees one unit at a time when two of them become directly connected to each other as a result of comparing the gain in (3) to the cost in (7). As degrees become larger and the network denser, it also happens that distances between node pairs become shorter. The node sets whose cardinalities appear in (3) tend, therefore, to become smaller. Together, these trends make it progressively harder for gains to surpass costs and for degrees to continue increasing. However, a few high-degree nodes do appear and the dynamics of network growth does occasionally consider joining two of them together. Because they have high degrees, it sometimes happens that the node-set cardinalities in (3) become once again relatively non-negligible and a few high-degree node pairs do indeed become interconnected.

We illustrate this process in Fig. 4, where a sequence of snapshots of the evolving node-degree distribution is shown

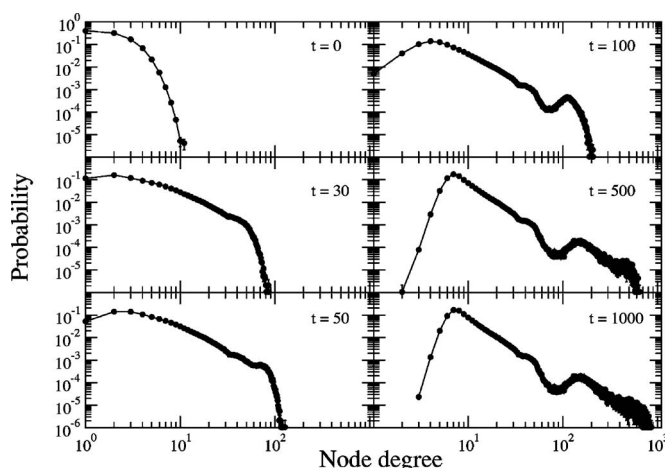


FIG. 4. Average node-degree distributions for $n=1024, C=100,$ and $D=0.1$ (cf. Figure 3) at selected values of t . The position of the peak in the distributions at $k \approx 150$ scales with the number of nodes proportionally to $n^{0.3 \pm 0.06}$.

for the $\gamma=0.9$ case of Fig. 3 with $n=1024$. Clearly, all nodes start out with relatively low degrees at $t=0$, but many migrate toward higher degrees with the passage of time. At $t=50$ a power law is already delineated, but the migrations continue and lead some of the nodes, particularly those of relatively high degrees, to have their degrees increased even further. In fact, nodes whose degrees surpass the parameter-dependent threshold mentioned above migrate quickly, leaving the degrees in the region from which they departed much less represented. This may be repeated over and over again, thus causing the oscillatory pattern we observe near the threshold.

Our results indicate that, as the node migrations toward higher degrees occur, a new power-law regime for the highest degrees is inaugurated. In this case, what we witness may be interpreted as the emergence of some sort of hierarchical organization within the network, not unlike what happens with the Internet [25], which is inherently organized in just such a way. That a single power-law regime should be reported in topology measurements like those of [5] may be due exclusively to the fact that they are constrained to within one single level of the hierarchy.

ACKNOWLEDGMENTS

We acknowledge partial support from CNPq, CAPES, and FAPERJ BBP grants, and the joint PRONEX initiative of CNPq and FAPERJ under Contract No. 26.171.176.2003. We also thank the T2-HEPGRID Brazil team at UERJ, where we ran our most time-consuming simulations.

[1] B. Bollobás, *Random Graphs*, 2nd ed. (Cambridge University Press, Cambridge, U.K., 2001).
 [2] P. Erdős and A. Rényi, *Publ. Math. (Debrecen)* **6**, 290 (1959).
 [3] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
 [4] S. N. Dorogovtsev and J. F. F. Mendes, *Adv. Phys.* **51**, 1079

(2002).
 [5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, in *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication* (ACM Press, New York, 1999), pp. 251–262.

- [6] A.-L. Barabási, R. Albert, and H. Jeong, *Physica A* **281**, 69 (2000).
- [7] A. Clauset and C. Moore, *Phys. Rev. Lett.* **94**, 018701 (2005).
- [8] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore, in *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing* (ACM Press, New York, 2005), pp. 694–703.
- [9] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [10] A.-L. Barabási, R. Albert, and H. Jeong, *Physica A* **272**, 173 (1999).
- [11] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Phys. Rev. Lett.* **85**, 4633 (2000).
- [12] P. L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).
- [13] P. Holme and B. J. Kim, *Phys. Rev. E* **65**, 026107 (2002).
- [14] B. Bollobás and O. M. Riordan, *Combinatorica* **24**, 5 (2004).
- [15] C. Cooper and A. Frieze, *Random Struct. Algorithms* **22**, 311 (2003).
- [16] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, in *Proceedings of the Forty-First IEEE Symposium on Foundations of Computer Science* (IEEE Computer Society, Los Alamitos, CA, 2000), pp. 57–65.
- [17] B. Bollobás and O. M. Riordan, in *Handbook of Graphs and Networks*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, Weinheim, Germany, 2003), pp. 1–34.
- [18] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
- [19] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, *Phys. Rev. Lett.* **89**, 258702 (2002).
- [20] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, in *Proceedings of the 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication* (ACM Press, New York, 2002), pp. 147–159.
- [21] L. Li, D. Alderson, W. Willinger, and J. Doyle, in *Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (ACM Press, New York, 2004), pp. 3–14.
- [22] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou, in *Proceedings of the Twenty-Ninth International Colloquium on Automata, Languages, and Programming* (Springer-Verlag, Berlin, 2002), pp. 110–122.
- [23] A. Fabrikant, A. Luthra, E. Maneva, C. H. Papadimitriou, and S. Shenker, in *Proceedings of the Twenty-Second Annual ACM Symposium on Principles of Distributed Computing* (ACM Press, New York, 2003), pp. 347–351.
- [24] M. Rosvall and K. Sneppen, *Phys. Rev. Lett.* **91**, 178701 (2003).
- [25] J. F. Kurose and K. W. Ross, *Computer Networking*, 3rd ed. (Addison-Wesley, Boston, MA, 2005).
- [26] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, 2nd ed. *Introduction to Algorithms* (MIT Press, Cambridge, MA, 2001).