Pseudofractal scale-free web

S. N. Dorogovtsev,[1,2,*] A. V. Goltsev,[2,†] and J. F. F. Mendes[1,‡]
[1]*Departamento de Física and Centro de Física do Porto, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal*
[2]*A.F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia*
(Received 8 December 2001; published 25 June 2002)

We find that scale-free random networks are excellently modeled by simple deterministic graphs. Our graph has a discrete degree distribution (degree is the number of connections of a vertex), which is characterized by a power law with exponent $\gamma = 1 + \ln 3/\ln 2$. Properties of this compact structure are surprisingly close to those of growing random scale-free networks with γ in the most interesting region, between 2 and 3. We succeed to find exactly and numerically with high precision all main characteristics of the graph. In particular, we obtain the exact shortest-path-length distribution. For a large network ($\ln N \gg 1$) the distribution tends to a Gaussian of width $\sim \sqrt{\ln N}$ centered at $\overline{\ell} \sim \ln N$. We show that the eigenvalue spectrum of the adjacency matrix of the graph has a power-law tail with exponent $2 + \gamma$.

DOI: 10.1103/PhysRevE.65.066122 PACS number(s): 87.18.Sn, 05.10.−a, 05.40.−a, 05.50.+q

The essence of the modern situation in network science is the change over from the study of classical random graphs with Poisson degree distributions [1] to the exploration of complex networks with fat-tailed degree distributions [2–7]. The prominent particular case of such nets are networks with power-law degree distributions (scale-free networks) [2,8]. While growing, such nets actually self-organize into scale-free structures. These networks play a great role in Nature [3–5]. The Internet, the World Wide Web, and many basic biological networks belong to this class. Fat tails of the degree distributions produce a number of intriguing effects [9–14].

Such networks are widespread, but very little is still known even about their basic properties [15,16]. Most of real growing scale-free networks have γ exponent of the degree distribution $P(k) \sim k^{-\gamma}$ in the range (2,3), but this case turned to be the most difficult and unexplored one. In particular, no exact results for the average shortest-path length $\overline{\ell}$ are known in this situation. The only known exact shortest-path-length distributions were obtained for the simplest equilibrium networks [17,18]. Notice that if $\gamma \leqslant 3$, standard estimates of $\overline{\ell}$ [19] are inapplicable to equilibrium networks with uncorrelated vertices. Correlations in growing networks are inevitable, and the results are even less encouraging. The generic property of these networks, which makes their analytical study so hard, is a complex structure of their adjacency matrices.

Scale-free random networks naturally have a continuous degree distribution spectrum, but it has recently been shown that discrete degree distributions of some deterministic graphs also have a power-law decay [20]. All known stochastic models of growing scale-free networks are based on three main ''physical'' principles: (i) the growth is stochastic; (ii) the growth produces ''long-distance'' connections, so that networks are ''compact;'' and (iii) the linking is preferential (new vertices become preferentially attached to more connected vertices). One may wonder, whether all these three ingredients are necessary to produce the structure and topology of a growing scale-free network. From the growth principles for a simple network with preferential linking we change only the first one, so that the growth is deterministic, and compare main structural characteristics of this graph with those of random growing networks with $\gamma < 3$.

Here we present results of this program. We succeed to find a number of exact characteristics of the scale-free deterministic graph, some of which are still unknown for random scale-free networks. The structural properties of deterministic and random scale-free growing networks proved to be surprisingly close to each other, so our results can be reasonably applied to random growing nets.

Pseudofractal graph. The most popular and simple models of random growing networks are citation graphs. By definition, new connections in them emerge only between new vertices and old ones. For example, in a scale-free citation graph proposed in Ref. [21], one vertex is created per time step and connects to both the ends of a randomly chosen edge. Here we use the closest deterministic variation of this network growing under the mechanism of ''preferential linking.''

The growth starts from a single edge connecting two vertices at $t = -1$ (see Fig. 1). At each time step, to every edge of the graph, a new vertex is added, which is attached to both the end vertices of the edge. Then, at $t = 0$, we have a triangle of edges connecting a triple of vertices, at $t = 1$, the graph consists of six vertices connected by nine edges, and so on. The total number of vertices at ''time'' t is $N_t = 3(3^t + 1)/2$, and the total number of edges is $L_t = 3^{t+1}$, so that the average degree is $\overline{k_t} = 2L_t/N_t = 4/(1 + 3^{-t})$.

This simple rule produces a complex growing network, which is certainly not a fractal [22]. Indeed, at any step, the entire graph can be set inside of a unit triangle. On the other hand, one can depict the graph in another way, namely, as in Fig. 1 where the graph is surrounded by a long chain of edges. This means that the structure has no fixed finite fractal

*Email address: sdorogov@fc.up.pt
†Email address: goltsev@gav.ioffe.rssi.ru
‡Email address: jfmendes@fc.up.pt

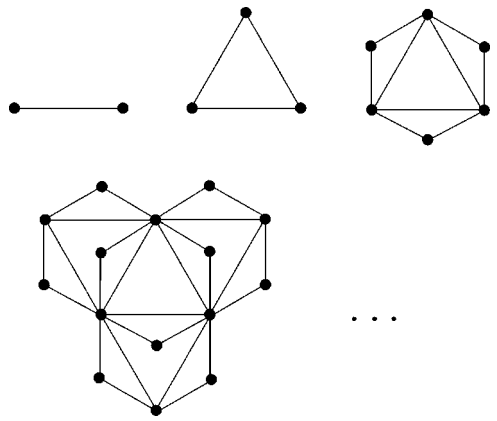

FIG. 1. Scheme of the growth of the scale-free pseudofractal graph. The growth starts from a single edge connecting two vertices at $t=-1$. At each time step, every edge generates an additional vertex, which is attached to both end vertices of the edge. Notice that the graph at time step $t+1$ can be made by connecting together the three t graphs.

dimension. We failed to introduce a well defined spectrum of fractal dimensions (in principle, for this structure, it must extend to infinite dimension), hence the network cannot be called a multifractal. Thus, this graph is not a fractal but only parody of it, and we call it, for brevity, *pseudofractal*. Notice that the graph contains numerous loops and hence is very far from tree like.

Adjacency matrix. By definition, an element a_{ij} of an adjacency matrix is equal to 1 or 0 depending on whether an edge between vertices i and j is present or not. The adjacency matrix $\hat{A}_t$ structure is schematically shown in Fig. 2. At $t=-1$, this is the 2×2 matrix with zeros on the diagonal and two unit elements. At time step t, we add rows and columns $i,j=N_{t-1}+1,\ldots,N_t$ (new vertices) to the matrix. Matrix is symmetric, and each unit element a_{ij} above the diagonal of the matrix $\hat{A}_{t-1}$ generates, in addition, two unit elements a_{is} and a_{js} of $\hat{A}_t$. Here $N_{t-1}+1\leqslant s\leqslant N_t$. Other elements are zeros. This produces the sparse block matrix shown in Fig. 2.

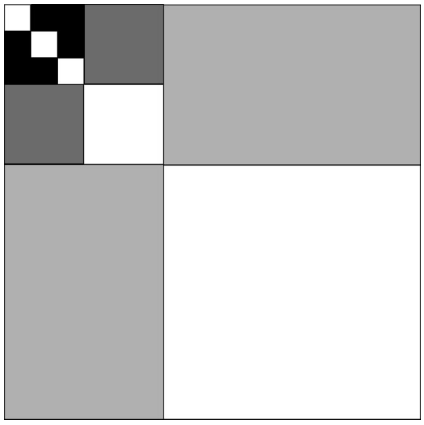

FIG. 2. Structure of the adjacency matrix of the graph ($t=2$, $N_t=15$). Black regions are unit elements of the matrix. In white regions, all matrix elements are zeros. In gray regions, nonzero (unit) elements are present. The matrix is symmetric, and each column in gray blocks above the diagonal contains only two nonzero elements.

Degree distribution. The degree spectrum of the graph is discrete: at time t, the number $m(k,t)$ of vertices of degree $k=2,2^2,2^3,\ldots,2^{t-1},2^t,2^{t+1}$ is equal to $3^t,3^{t-1},3^{t-2},\ldots,3^2,3,3$, respectively. Other values of degree are absent in the spectrum. Clearly, for the large network, $m(k,t)$ decreases as a power of k, so the network can be called ''scale-free.'' Spaces between degrees of the spectrum grow with increasing k. Therefore, to relate the exponent of this discrete degree distribution to standard γ exponent of a continuous degree distribution for random scale-free networks, we use a cumulative distribution $P_{cum}(k)\equiv\Sigma_{k'\geqslant k}m(k',t)/N_t\sim k^{1-\gamma}$. Here k and k' are points of the discrete degree spectrum. Thus we obtain

$$\gamma=1+(\ln 3/\ln 2), \tag{1}$$

so that $2<\gamma=2.585\ldots<3$. Compare γ with the characteristic exponent in the relation between the ''mass'' and the ''perimeter'' of the graph. Also, notice that the maximal degree of a vertex is equal to $2^{t+1}\sim N_t^{\ln 2/\ln 3}=N_t^{1/(\gamma-1)}$, which coincides with a standard relation for the cutoff of degree distribution in growing scale-free networks [5].

Distribution of clustering. By definition, the cluster coefficient C of a vertex is the ratio of the total number of existing connections between all k its nearest neighbors and the number $k(k-1)/2$ of all possible connections between them. Usually, only the average value of the clustering coefficient is considered. In our case, it is possible to obtain a more rich characteristic, namely, the distribution of the clustering coefficient in the graph.

One can see that, in this graph, there is a one-to-one correspondence between the clustering coefficient of a vertex and its degree: $C=2/k$. Thus, the number $m_c(C,t)$ of vertices with clustering coefficient $C=1,2^{-1},2^{-2},\ldots,2^{2-t},2^{1-t},2^{-t}$ is equal to $3^t,3^{t-1},3^{t-2},\ldots,3^2,3,3$, respectively. In this case, it is natural to introduce the cumulative distribution of the clustering coefficient $W_{cum}(C)\equiv\Sigma_{C'\leqslant C}m_c(C',t)/N_t\sim C^{\ln 3/\ln 2}=C^{\gamma-1}$, where C and C' are the points of the discrete spectrum. This corresponds to the power-law behavior of the corresponding continuous distribution of clustering $W(C)\sim C^{\gamma-2}$ for random scale-free network at small C. Such distribution of clustering is observed in real scale-free networks of protein-protein interactions [23].

The average clustering coefficient can be easily obtained for arbitrary t,

$$\overline{C}_t=\frac{4}{5}\,\frac{6^t+3/2}{2^t(3^t+1)}. \tag{2}$$

For the infinite graph, $\overline{C}=4/5$, so the clustering is high.

Degree correlations. The number $m(k,k',t)$ of edges, which connect vertices of degree k and k', characterizes short-range degree-degree correlations in the graph. It is convenient to write $k\equiv 2^{p+1}$ and use the notation $m(k,k',t)\equiv c(p,p',t)$. Then one can find directly

$$c(t,t,t)=3,\quad c(t,p'\leqslant t-1,t)=3\times 2^{t-1-p'},$$

$$c(p\leqslant t-1,p'\leqslant p-1,t)=3^{t-p}2^{p-p'-1}. \tag{3}$$

This yields the cumulative distribution $\sim k^{2-\gamma}k'^{-1}$ (we assume that $k\gg k'$), which, in turn, corresponds to the effective continuous distribution

$$P(k,k')\sim k^{1-\gamma}k'^{-2}. \tag{4}$$

This expression coincides with the corresponding asymptotic formula for an arbitrary random scale-free citation graph [5] (by definition, a citation graph is a growing network, in which new edges do not emerge between pairs of old vertices). Originally, Eq. (4) was obtained exactly for a specific model in Ref. [24].

Shortest-path-length distribution. Here we briefly outline our exact results for the distribution $\mathcal{P}(\ell,t)\equiv n(\ell,t)/[N_t(N_t-1)/2]$, where $n(\ell,t)$ is the number of pairs of vertices with minimal separation ℓ. Details of the analytical solution and general exact expressions for $n(\ell,t)$ will be published elsewhere.

For the calculation of $n(\ell,t)$ one can use two distinct schemes. First, one may use the following property. The length ℓ_{ij} of the shortest path between vertices i and j is equal to the minimal power of the adjacency matrix with nonzero $\{ij\}$ element: $\{\hat{A}^{\ell-1}\}_{ij}=0,\{\hat{A}^{\ell}\}_{ij}\neq 0$. This property allows us to obtain $n(\ell,t)$ by counting the total numbers of nonzero elements in sequential powers of the adjacency matrix. This can be easily done by computer and yields $n(\ell,t)$,

3				
9	6			
27	57	21		
81	351	369	60	
243	1806	3582	1716	156
. . . ,				

(5)

etc., where t labels lines ($t=0,1,2,3,4,\ldots$) and $\ell=1,2,3,4,5,\ldots$ is the index of columns.

Second, the exact analytical form of the distribution $\mathcal{P}(\ell,t)$ and all its moments were obtained by the solution of recursion relations for $n(\ell,t)$. In particular, an exact expression for the average shortest-path length is of the form

$$\overline{\ell}(t\geqslant 0)=\frac{(4t+11)3^{2t}+10\times 3^t+3}{3(3^t+1)(3\times 3^t+1)}. \tag{6}$$

The distribution quickly approaches an asymptotic regime, where

$$\overline{\ell}(t\gg 1)=\tfrac{4}{9}\,t+\tfrac{11}{9}+O(t3^{-t})\cong\frac{4}{9\ln 3}\ln N_t. \tag{7}$$

Thus, the average shortest-path length logarithmically grows with increasing size of the graph. Expression (7) may be compared with the standard estimate [19]: $\overline{\ell}\sim\ln N/\ln\overline{k}=\ln N/\ln 4$. The relative difference is surprisingly small, $(4/9)(\ln 4/\ln 3)=0.561\ldots$. Notice that according to standard arguments [4,5], the classical formula is not applicable for $\gamma\leqslant 3$ (recall that this formula was obtained for equilibrium uncorrelated graphs). Nevertheless, Eq. (7) demonstrates that the classical estimate is unexpectedly good in our case where degree-degree correlations are strong.

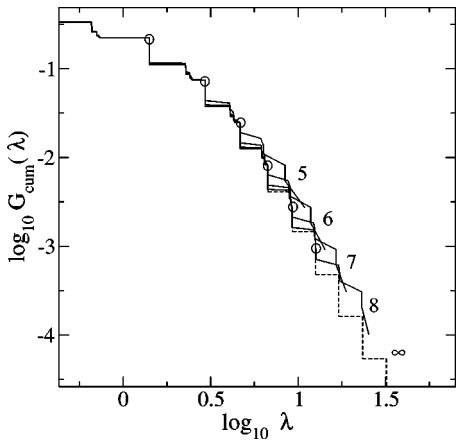

FIG. 3. Log-log plot of the cumulative distribution of eigenvalues of the adjacency matrix, $G_{cum}(\lambda)\equiv\Sigma_{\lambda'\geqslant\lambda}G(\lambda')$. The curves show the spectra for $t=5,6,7,8$. The dashed line depicts the $t\to\infty$ limit. The t-independent points are marked.

The exact distribution is rather complex and we do not present it here, but we found analytically that at large t, it takes the Gaussian form

$$\mathcal{P}(\ell,t)\cong\frac{1}{\sqrt{2\pi(2^2/3^3)t}}\exp\left[-\frac{(\ell-\overline{\ell}(t))^2}{2(2^2/3^3)t}\right], \tag{8}$$

which is violated only in narrow regions of width $\sim t^{1/3}$ near the points $\ell=1$ and $\ell=t+1$. One sees that the width of the distribution is of the order of $\sqrt{t}\sim\sqrt{\ln N_t}\ll\overline{\ell}(t)$. Notice that the simulations of the Barabási-Albert growing random network also yield a Gaussian-like $\mathcal{P}(\ell)$ [7].

Eigenvalue spectrum of the adjacency matrix. The observation of the power-law eigenvalue spectrum of the adjacency matrix of the Internet [25] makes this problem very challenging. The eigenvalue spectrum of the adjacency matrix of the graph $G(\lambda)$ contains N_t eigenvalues, which are denoted by λ. For $t\geqslant 2$, $N_{t-1}-3$ of them are equal to zero and, for $t\geqslant 3$, there are $N_{t-2}-3$ eigenvalues equal to $\sqrt{2}$ and the same number of those equal to $-\sqrt{2}$. Here we do not derive analytical results for the entire eigenvalue spectrum but only study its tail using a simple numerical analysis.

It is convenient to consider a cumulative distribution of eigenvalues $G_{cum}(\lambda)\equiv\Sigma_{\lambda'\geqslant\lambda}G(\lambda')$. The results of numerical diagonalization of the adjacency matrix for several time steps are shown in Fig. 3. One sees that, in the large graph limit, the resulting cumulative distribution approaches a staircaselike form for $\lambda\gg 1$.

We found network-size-independent points of the spectrum. Using the coordinates of these points, we calculated the series of slopes of lines connecting these points in Fig. 3: 1.498 47, 2.381 92, 3.030 23, 3.406 83, 3.531 35, 3.557 (the

last value is actually a very good estimate obtained for $t=8$). Interpolation of these values yields the exponent $\delta-1=3.575\pm0.015$ of the cumulative distribution $G_{cum}(\lambda)\sim\lambda^{-(\delta-1)}$. This value is $2+\ln 3/\ln 2=1+\gamma=3.585\ldots$, to within a precision of our numerics.

Thus, one can suggest that exponent δ of the corresponding continuous eigenvalue spectrum $G(\lambda\gg1)\sim\lambda^{-\delta}$ is $\delta=2+\gamma$. We should mention that the direct study of the eigenvalue spectrum for a growing random network with $\gamma=3$ (the Barabási-Albert model) showed power-law dependence in a too narrow range of λ to make precise conclusions [26] (see also Ref. [27]). An estimate for the exponent in this situation was $\delta\approx5$ [26], which supports our conjecture.

Percolation properties. Let us delete, at random edges of the pseudofractal. We denote the probability that an edge is present as p. Then the standard real-space renormalization group transformation is exactly $1-p'\leftarrow(1-p)(1-p^2)$, or $p'\leftarrow p+p^2-p^3$. This transformation actually means the ''decimation'' of the graph (see Fig. 1 in reverse order). So, we see that in the infinite network limit, for any nonzero value of p, the system finally approaches the fixed point $p=1$. This means that one has to remove at random ''almost all'' edges to eliminate the giant connected component of the graph. The same is valid for the removal of vertices. This is a standard property of scale-free networks with $\gamma\leqslant3$ [10].

Discussion. The network that we study in this paper was chosen to be as close as possible to random growing networks. Moreover, one can say that this graph grows under the mechanism of ''preferential linking:'' vertices with higher numbers of connections attach higher numbers of new edges (see Fig. 1). Therefore, it is not so strange that the properties of the pseudofractal network resemble those of scale-free random citation graphs. However, it is really surprising how close they appear to each other. Hardly one can propose a more simple deterministic scale-free growing network. In a similar way, we can easily construct various pseudofractal variations with various values of γ but their general properties are similar. Therefore, we have found a convenient tool for exploring complex scale-free networks. In the real-space renormalization group technique for critical phenomena and percolation on lattices, the lattices are actually changed to fractal structures. We have demonstrated that the replacement of a random growing network by a pseudofractal is also reasonable.

The extreme simplicity of the pseudofractal graph has allowed us to obtain a number of results for growing networks. In particular, for this network with strong correlations, we have obtained the exact shortest-path-length distribution and the eigenvalue spectrum of the complex adjacency matrix. From the latter, we have made a conjecture that the exponent of eigenvalue spectra of scale-free citation graphs is $2+\gamma$. Of course, real growing networks are not deterministic graphs. Nevertheless, we have failed to find any principal difference between the structural properties of pseudofractals and those of random growing nets.

Note added. After submission of this manuscript we came to know of the paper by Jung and co-workers [28] where deterministically growing scale-free trees, which also belong to the class of pseudofractals, were studied. We thank B. Kahng, who informed us about this paper before its publication.

S.N.D. thanks PRAXIS XXI (Portugal) for research Grant No. PRAXIS XXI/BCC/16418/98. S.N.D and J.F.F.M. were partially supported by the project POCTI/99/FIS/33141. A.G. acknowledges the support of the NATO program OUTREACH. We also thank A.N. Samukhin and A. Krzywicki for useful discussions.

[1] P. Erdös and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959); Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).
[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[3] S.H. Strogatz, Nature (London) **401**, 268 (2001).
[4] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[5] S.N. Dorogovtsev and J.F.F. Mendes, e-print cond-mat/0106144; Adv. Phys. **51**, 1079 (2002).
[6] Z. Burda *et al.*, Phys. Rev. E **64**, 046118 (2001).
[7] A. Krzywicki, e-print cond-mat/0110574.
[8] L.A.N Amaral *et al.*, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[9] R. Albert *et al.*, Nature (London) **406**, 378 (2000).
[10] R. Cohen *et al.*, Phys. Rev. Lett. **85**, 4626 (2000).
[11] D. S. Callaway *et al.*, Phys. Rev. Lett. **85**, 5468 (2000).
[12] R. Cohen *et al.*, Phys. Rev. Lett. **86**, 3682 (2001).
[13] R. Pastor-Satorras *et al.*, Phys. Rev. Lett. **86**, 3200 (2001).
[14] S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. Lett. **87**, 219801 (2001).
[15] P.L. Krapivsky *et al.*, Phys. Rev. Lett. **85**, 4629 (2000).
[16] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[17] S.N. Dorogovtsev *et al.*, Europhys. Lett. **52**, 33 (2000).
[18] P. Bialas, Phys. Lett. B **384**, 75 (1996).
[19] M.E.J. Newman *et al.*, Phys. Rev. E **64**, 026118 (2001).
[20] A.-L. Barabási *et al.*, Physica A **299**, 559 (2001).
[21] S. N. Dorogovtsev *et al.*, Phys. Rev. E **63**, 062101 (2001).
[22] Compare this graph with perfect fractals generated in the Migdal-Kadanoff renormalization procedure: A.A. Migdal, Sov. Phys. JETP **42**, 743 (1976); L.P. Kadanoff, Ann. Phys. (N.Y.) **100**, 359 (1976). In the corresponding version of the Migdal-Kadanoff approach, every edge is transformed to a pair of parallel chains, each one consisting of two edges. Note that Migdal-Kadanoff fractals also have ''scale-free'' discrete spectrum of degrees but, of course, their mean shortest-path lengths are very large.
[23] S.N. Dorogovtsev and J.F.F. Mendes (unpublished).
[24] P.L. Krapivsky *et al.*, Phys. Rev. E **63**, 066123 (2001).
[25] M. Faloutsos *et al.*, Soc. Sci. Q. **29**, 251 (1999).
[26] I.J. Farkas *et al.*, Phys. Rev. E **64**, 026704 (2001).
[27] K.-I. Goh *et al.*, Phys. Rev. E **64**, 051903 (2001).
[28] S. Jung, S. Kim, and B. Kahng, e-print cond-mat/0112361.