Neighborhood properties of complex networks

Roberto F. S. Andrade and José G. V. Miranda

Instituto de Física-Universidade Federal da Bahia, 40.130-240, Salvador, Brazil

Thierry Petit Lobão

Instituto de Matemática-Universidade Federal da Bahia, 40210-340, Salvador, Brazil

(Received 10 July 2005; published 3 April 2006)

A concept of neighborhood in complex networks is addressed based on the criterion of the minimal number of steps to reach other vertices. This amounts to, starting from a given network R_1, generating a family of networks $R_\ell, \ell = 2, 3, \ldots$ such that, the vertices that are ℓ steps apart in the original R_1, are only 1 step apart in R_ℓ. The higher order networks are generated using Boolean operations among the adjacency matrices M_ℓ that represent R_ℓ. The families originated by the well known linear and the Erdös-Renyi networks are found to be invariant, in the sense that the spectra of M_ℓ are the same, up to finite size effects. A further family originated from small world network is identified.

INTRODUCTION

Several properties of complex networks have been addressed recently, much inspired by the identification of their relevance in the description of the relations among the individual constituents of many systems, which have their origin from natural to social sciences, telephone and internet to energy supply, etc. [1–3]. The huge development in this research area came along with the proposition of many parameters that might be relevant to the characterization of properties of networks. The most established quantifiers so far are the distribution of links $n(k)$, clustering coefficient C, mean minimal distance among the nodes $\langle d \rangle$, diameter D, the assortativity degree [4,5]. The evaluation of these and other indices for actual networks help to characterize them, putting into some network classes with well defined properties, such as the small-world (SM) [6], the scale-free [7], the Erdös-Renyi (ER) (or random networks) [8], etc.

As the number of nodes directly connected to node i is k_i, $n(k)$ characterizes the immediate neighborhood of the network nodes. In this work we explore further neighborhood properties of a network R, which are related to the distribution of the number of second, third, …, neighbors. The higher order neighborhoods are characterized with the help of a family of adjacency matrices (AM) [4], which in turn are evaluated within a Boolean operation formalism, as will be detailed below. After that, we discuss some of the neighborhood properties and also explore some concepts related to neighborhood invariance.

For the sake of simplicity, we consider herein networks that are fully connected, i.e., each of its nodes can be reached from any other one. Two nodes are ℓth neighbors, or neighbors of order $O(\ell)$, when the minimal path connecting them has ℓ steps. So, the usual distribution $n(k)$ is based on the $O(1)$ neighborhood. For a given R, the explicit evaluation of $O(\ell)$ along the network, $\ell = 0, 1, 2, 3, \ldots$ promptly indicates the structure of minimal paths connecting the nodes, where we consider that any vertex is $O(0)$ neighbor of itself. If R is finite, the possible values of ℓ are limited by the network

diameter D. The knowledge of $O(\ell)$ classifies uniquely the neighborhood of a vertex, in the sense that if two vertices are $O(\ell)$ neighbors, they are not $O(j)$ neighbors for $\ell \neq j$.

For each ℓ we identify all $O(\ell)$ neighbors of R and construct a family of networks $\mathbf{R}$. Each $R_\ell \in \mathbf{R}$ is defined by the same set of nodes, but edges link only pairs of nodes that are $O(\ell)$ neighbors in $R_1 \equiv R$, so that the family $\mathbf{R}$ characterizes the neighborhood structure of R. On the other hand, each member R_ℓ can be investigated by itself. For instance, we may consider the links described by R_2 and ask which are the properties of this network, regardless of the fact that it describes the second neighbors of R. Thus, the investigation of properties of distinct neighborhoods $O(\ell)$ of R can proceed by the investigation of the family $\mathbf{R}$.

BOOLEAN FORMALISM FOR HIGHER ORDER NEIGHBORHOODS

In order to obtain the R_ℓ's, we set up a corresponding family of AM's $M_\ell \in \mathbf{M}$, $\ell \in [0, D]$: if two sites i and j are ℓ' neighbors in R, then $(M_\ell)_{ij} = \delta_{\ell, \ell'}$. M_ℓ's can be successively evaluated by the systematic use of Boolean (B) operations [9] among matrices. Note that AM's $M_0 = I$ (identity matrix) and M_1, which describes the original network R_1, have only 0's or 1's as entries for all of its elements. If M_1 is applied to a unitary vector v_i, with all but one entry i set to 0, the resulting vector expresses which vertices are linked to the vertex i. If we take the usual matrix product of M_1 by itself, the elements $(M_1^2)_{rs}$ indicate how many possible two-step walks along the network, with end points r and s, exist. Contrary to what happens with M_1, M_1^2 has many elements $(M_1^2)_{rs} > 1$, indicating multiplicity of two-step paths starting at r and ending at s. In particular, all elements $(M_1^2)_{rr}$ of the diagonal can have this property, since they count all two-step walks that start at r, visit any of the vertices s to which r is linked (i.e., $(M_1)_{rs} = 1$), and turn back to r. The same interpretation is valid for all usual powers of M_1.

As the neighborhood concept does not take path multiplicity into account, it is most convenient, for the current

purposes, to use B operations among matrices. Regarding $(M_1)_{rs}=0,1$ as B variables, we use B sum, subtraction and product operations [9], respectively $\oplus,\ominus,\otimes$,

$$0 \oplus 0 = 0, \quad 1 \oplus 0 = 1, \quad 0 \oplus 1 = 1, \quad 1 \oplus 1 = 1;$$

$$0 \ominus 0 = 0, \quad 1 \ominus 0 = 1, \quad 0 \ominus 1 = 0, \quad 1 \ominus 1 = 0;$$

$$0 \otimes 0 = 0, \quad 1 \otimes 0 = 0, \quad 0 \otimes 1 = 0, \quad 1 \otimes 1 = 1 \quad (1)$$

to define further B matrices M_ℓ. Herein, B matrix operations are defined by using the usual matrix element operation rules, replacing the usual sum, subtraction and product among matrix elements by the corresponding B operations. To avoid multiplicity of notation we also use $\oplus,\ominus$ and $\otimes$ to indicate B matrix operations.

As an example, let us consider $\overline{M_2}=M_1 \otimes M_1$ and compare it to M_1^2. The position of all their zero elements coincides, while $\overline{M_2}$ can be obtained from $\underline{M_1^2}$ by collapsing to 1 all of its nonzero elements. In fact, $\overline{M_2}$ just indicates the possibility of two-step walks, but not the multiplicity of walks. Note that $\overline{M_2}$ does not describe the second order neighborhood of R, as it includes 1s in the whole main diagonal ($O(0)$ neighbo rhood), as well as between nodes i and j that, together with a third node k, are connected in a triangle. Such nodes already belong to $O(1)$ and, much as those in $O(0)$, should be present in $O(2)$. Thus, to obtain M_2, it is necessary to B sum M_1 to $\overline{M_2}$, and subsequently subtract, from this sum, the contributions I and M_1. This leads to

$$M_2 = (M_1 \oplus \overline{M_2}) - (I \oplus M_1), \quad (2)$$

that can be further simplified to

$$M_2 = (I \oplus M_1) \otimes M_1 - (I \oplus M_1). \quad (3)$$

Finally, it is possible to proof, e.g., by finite induction, that Eq. (3) can be generalized for arbitrary value of ℓ by

$$M_\ell = \left(\bigoplus_{j=0}^{\ell-1} M_j\right) \otimes M_1 - \left(\bigoplus_{j=0}^{\ell-1} M_j\right). \quad (4)$$

After obtaining the M_ℓ's, whose Boolean algorithm can be implemented by bit manipulation to strongly reduce the amount of disk and CPU storage, several usual indices for network characterization can be easily worked out. This includes the average shortest distance between the nodes and the diameter. So we observe that: (1) For a finite network with $\mathbf{N}$ vertices, there is a large enough ℓ_{max} such that $M_\ell \equiv 0, \forall \ell > \ell_{max}$. Thus, the diameter $D \equiv \ell_{max}$. (2) As for each r,s pair, $(M_\ell)_{rs}=1$ for only one (M_ℓ), one can collapse in a single matrix

$$\hat{M} = \sum_{j=0}^{\ell_{max}} jM_j \quad (5)$$

all information on the neighborhood of any pairs of vertices. Particularly, all pairs $(r,s) \in O(\ell)$, satisfy $\hat{M}_{rs}=\ell$. To obtain the average shortest distance for each node r, it is sufficient to sum all elements of the rth row (or column) of $\hat{M}$ and divide by $\mathbf{N}-1$. Taking the average over all nodes we immediately obtain the average shortest distance for the entire net-

work. (3) $\hat{M}$ can be also used to visualize the structure of network with the help of color or gray code plots, as shown in Fig. 4.

Further properties of ℓ neighborhoods can be obtained within this framework, just by looking at properties of each R_ℓ as an independent network. This includes two very important indices which characterize the local landscape of a network: (i) the degree distribution $n(k) \equiv n_1(k)$; and (ii) the clustering coefficient $C \equiv n_1$. They are easily obtained from M_1 by performing: (i) usual sums over matrix rows; and (ii) line to line multiplication of matrix elements. If the same procedure is adopted on each M_ℓ, corresponding indices $n_\ell(k)$ and C_ℓ can be evaluated to characterize the local landscape of higher order $O(\ell)$. Note that this introduces no new indices, just extends to the higher order neighborhood the same concepts used for characterizing the immediate neighborhood of the nodes.

Local landscapes described by higher order neighborhoods described by R_ℓ can be rather distinct from those seen by the original R_1. Take, for instance, the simple Cayley tree, which is a loopless network. It is an easy matter to see that, for $\ell \geq 2$, R_ℓ describe structures full of loops, and where the network is split into several disjoint subnetworks.

NEIGHBORHOOD INVARIANT PROPERTIES

We can also ask whether, in opposition to the given example, there are networks for which some specific property G (or a set of properties) remains invariant as ℓ changes. If this is true, it is possible to assign to R a G neighborhood invariant (NI) property. This invariance of the array of edges linking the nodes of R expresses insensitivity of local properties with respect to length scale, provided length is measured by number of steps along the network. It is distinct from both the scale-free distribution of vertex connections, as well as from geometrical scale invariance included in the construction of the network as, for instance, in the class of Apollonian networks [13–15].

To define a NI criterion, it is necessary to identify a relevant network property present in all elements of $\mathbf{R}$ [10]. Though important, $n_\ell(k)$ and C_ℓ are global indices and may not provide a sufficiently fine characterization of a network. We propose to adopt a more precise invariance criterion, based on the eigenvalue spectrum density $\rho_\ell(\lambda)$ of the family M_ℓ [11,12]. Although other criteria can be proposed, this one is much more sensitive to details of the network than, e.g., the global indices just quoted. Also it is worth nothing that the set of eigenvalues is closely connected to the notion of neighborhood (see [12]). Nevertheless, as distinct co-spectral graphs exist, only the complete set of eigenvalues and eigenvectors univocally characterizes a network. Thus, even the spectral criterion may not be an exact one.

The numerical evaluation of the eigenvalue spectra and spectral density ρ_ℓ has been carried out for several networks. Exact NI invariance is observed for the linear chain network, where each vertex interacts (as in system of phonons, tight-binding electrons [16,17]) only with its two nearest neighbors (nn). $M_{1,lc}$ has 0's overall, with exception of 1's along the nearest upper and lower diagonals to the matrix main
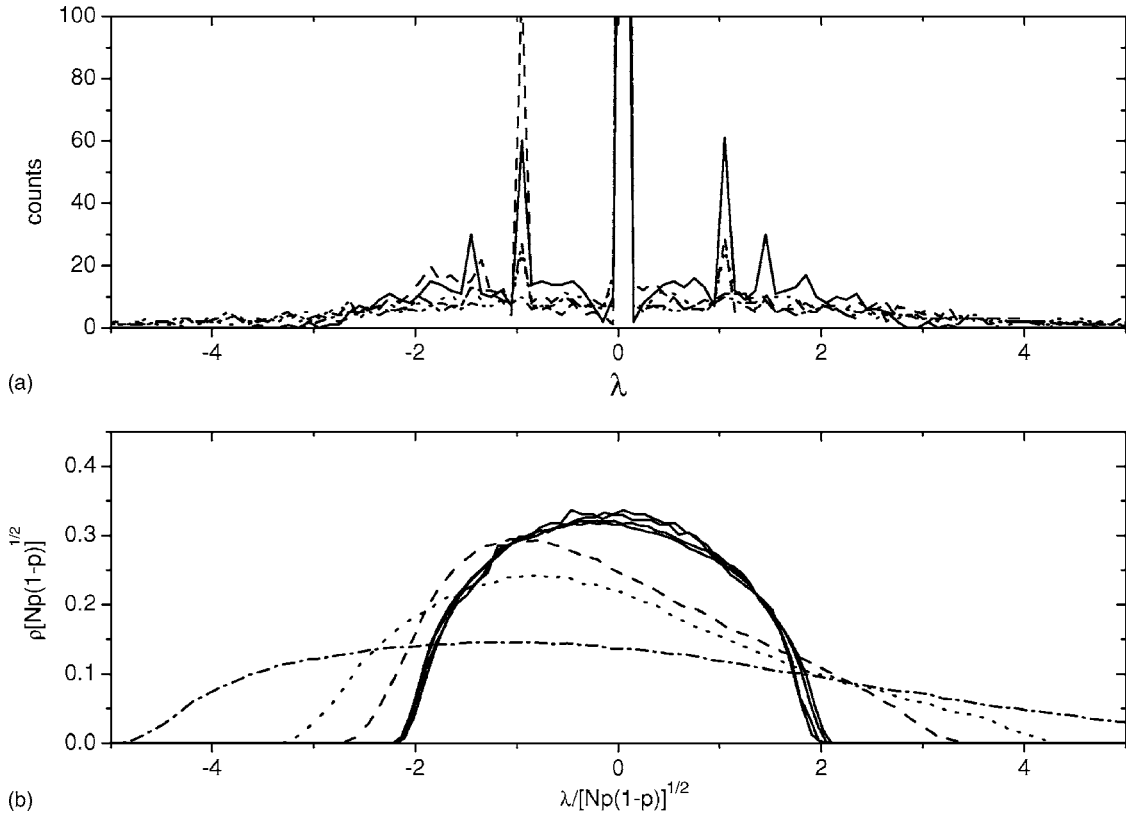
FIG. 1. Spectral density ρ_ℓ for the ER network single realizations. (a) $N=1000$, $p=0.05$, $\ell=1,2,3,4,5$. (b) Four semicircle shaped (solid lines) ρ_2 for $p=c/N^z$, when $c=68$, $z=0.9$, $N=500,1000,3000,3500$, which collapse to corresponding ρ_1. Three left-skewed ρ_2 spectra, when $c=32$, $z=0.92$, $N=900$ (dash), 1000 (dot) and 1700 (dash-dot), for which ρ_1 obeys the semicircle law.

diagonal. The other $M_{(\ell,lc)}$ are similar to $M_{1,lc}$, with two sequences of 1's along the upper and lower ℓ near diagonals. $M_{1,lc}$ or similar matrices have been used as the starting point to construct Watts-Strogatz small world networks [6], by replacing some of the original links according to a given rewiring probability p_w. The spectral density ρ_1 is expressed analytically by the relation [16–18] $\rho_{1,lc}(\lambda) \sim (4-\lambda^2)^{-1/2}$. If we consider an infinite network, all $\rho_{\ell,lc}$ have exactly the same shape of $\rho_{1,lc}$. Indeed, the search for higher order neighborhoods can be regarded, e.g., in the analysis of tight-binding systems, as a decimation procedure of half of the sites along with a renormalization of the hopping integral [18]. In the finite size cases, $\rho_{\ell,lc}$ depends slightly on ℓ and also on the boundary condition (open or closed) of the network. The spectral invariance means that any ℓth neighborhood of any nodes (or almost all, in the finite size case), is equivalent to that of the first neighborhood, consisting of just two nodes.

Moving far away from the completely ordered scenario, we have looked for NI in Erdös-Renyi (ER) random networks [8] as, if connections are randomly distributed for R_1, so should they also be for all members of **R**. However, the properties of the (ER) networks, hence of ρ_{ER}, depend on the connection probability p between any two nodes. If $p \ll 1$, R_1 is split into several disjoint clusters, in opposition to our assumption. In such cases, $\rho_{1,ER}$ is constituted by some individual peaks superimposed on a shallow background [see Fig. 1(a)]. Though not coincident, $\rho_{\ell,ER}$ share the same quali-

tative structure as an increase of the $\lambda=0$ dominant peak and a decrease of the other peaks are observed. This is related to the clustered structure of the network and to the fact that several clusters are reaching their own diameter.

When $p=c/N^z$, with $z<1$, and for large enough c, almost all nodes are connected in a single cluster, and ρ obeys the well known semicircle law [19,20] $\rho_{sc}(\lambda) \sim (4-\lambda^2)^{1/2}$. In this regime, networks usually have a very small diameter, and invariance can only be noted for a few values of ℓ. For the average node number $\langle k \rangle \simeq 0.5$, we have clearly found NI, as $\rho_2=\rho_1=\rho_{sc}$, as shown in the Fig. 1(b). However, for smaller values of $\langle k \rangle$, a clear skewness in the distribution is observed for ρ_2, despite the semicircle form of ρ_1 (not shown in the figure). This result gives a partial answer to a natural and important question: whether the randomness signature for ER networks with the occupation probability $p \sim 1/N^z$, expressed by the semicircle shape of the spectrum, is also be present for higher order neighborhoods. It hints at the validity of the conjecture, yet restricted to a range of values of c and z.

We have further investigated small-world (SW) networks, starting both from the nearest-next-neighbor (nnn) and the above discussed nn linear chains. We have found NI properties for two p_w intervals. In Fig. 2, we show several ρ_ℓ for very small value of p_w, starting from a nn chain: ρ is split into two parts: the first one, essentially described by $\rho_{1,lc}$, corresponds to the contribution of unperturbed segments of the linear chain. This part remains almost invariant for a
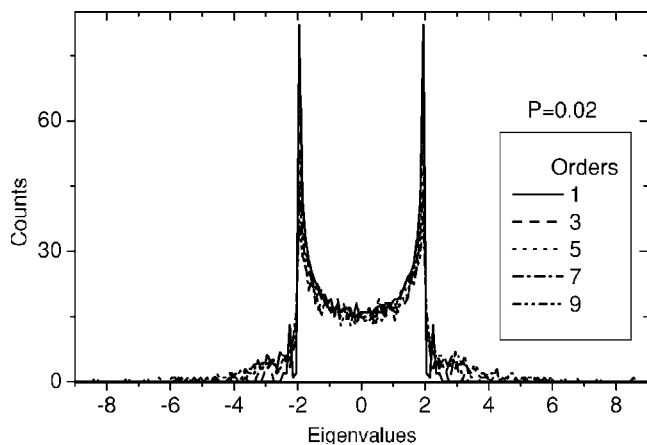
FIG. 2. Plots of the spectra $\rho \times \lambda$ of the matrices $M_\ell, \ell = 1, 3, 5, 7, 9$, for a Watts-Strogatz small-world network starting from a nearest neighbor linear chain and $p_w = 0.02$.

FIG. 4. For the same Watts-Strogatz small network of Fig. 4, graphical illustration of distribution of $O(\ell)$ neighborhoods: $\ell \in [1, 6], [7, 10], [11, 15]$ for (a), (b) and (c) respectively.

large number of M_ℓ, much as exactly observed for $\rho_{\ell, lc}$. The second part of $\rho_{\ell, lc}$, for $|\lambda| > 2$, has no well defined shape. It contains a considerably lower number of eigenvalues, which increases with ℓ: a number of eigenvalues migrates from the unperturbed part into it, as the number of vertices that are affected by the rewiring operation increases with ℓ. The successive migration ends up by affecting the whole form of the spectrum. Of course this behavior depends on p_w. If we increase it, a smooth transition on the shape of ρ_1 occurs, moving towards a pattern with many structures and bands. At the same time, the subsequent ρ_ℓ shows strong dependence on ℓ.

Starting now with nnn chain, we use the Newman process of randomly adding links (instead of rewiring), with probability p, to obtain networks with SW properties. Taking p in the range $p_w \sim 0.15 - 0.22$, we have observed that, as ℓ increases, the spectral density ρ_ℓ evolves towards a very peculiar form, which remains almost invariant for several values of ℓ, as shown in Fig. 3. This spectrum has its own features,

distinct from those discussed before for the fully ordered and disordered networks. For the specific case $p_w = 0.21$ and $N = 1000$, we have a large diameter $D = 15$, and the shown form of the spectra remains invariant for $\ell \in [7, 10]$. For smaller values of ℓ, the shape changes steadily from structured shapes similar to those in Fig. 2 into the invariant form. For larger values of ℓ, finite size effects lead to quite sparse M_ℓ, with a large number of zero eigenvalues: ρ_ℓ evolves to a δ like distribution centered at $\lambda = 0$.

This effect can be graphically illustrated with the help of the matrix $\hat{M}$. In Fig. 4 we draw the position of the $O(\ell)$ neighbors for three distinct ranges of ℓ. In the first range $\ell \leqslant 6$, the typical pattern of linear chain is recognized by diagonal stripes. Isolated islands result from rewiring of some nn or nnn links, and the shape of ρ reminds us of Fig. 2. The particular shape in Fig. 3 is associated with roughly dense matrices when $\ell \in [7, 10]$, and is much more related to robust structures that emerged from rewired links. For larger values of $\ell \in [11, 15]$, the links between all pairs of spins have already been scanned in previous neighborhoods, so that M_ℓ's become very sparse.

For other values of N, we have observed the same evolution. For instance, when $N = 1500$ and 2000, the shape lasts almost invariant for more generations, respectively $\ell \in [8, 12]$ and $[11, 16]$, indicating that this behavior can be more robust as N increases. For p_w larger than the range given above, this persistence in the form is not observed. For smaller values of p_w the spectra changes very slowly as shown in Fig. 2. In such cases, finite size effects set in prior than the form shown in Fig. 3.

To summarize, we have investigated higher order neighborhood and NI of networks. The first issue has been done by a systematic use of B operations and the definition of an AM family. The formalism we use allows to uniquely assign links between all pairs of nodes according to the minimal number ℓ of steps required to connect them walking on the given network. This defines a set of ℓ neighborhood of the network, each one of them being characterized as a network in its own. Further, we considered a concept of NI based on the properties of the the spectral density ρ. Finally, we explored well-known networks, showing that NI can be exactly found for the linear chain, and approximately, for the Erdös-Renyi network and Watts-Strogatz small-world networks. In the later cases, NI is observed for particular intervals of, respectively, connection probability p and rewiring probability p_w.
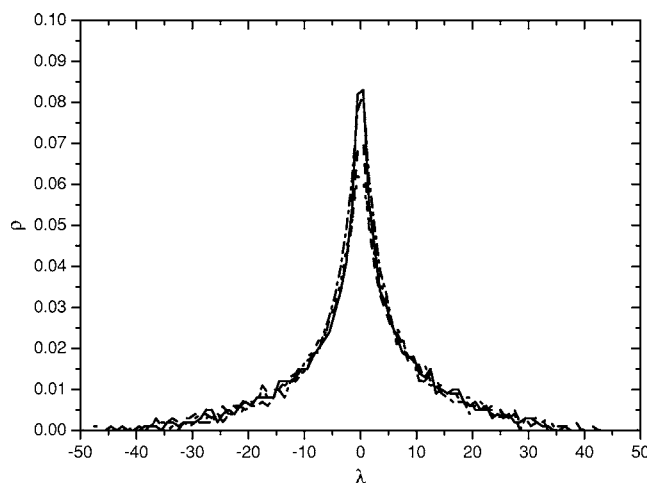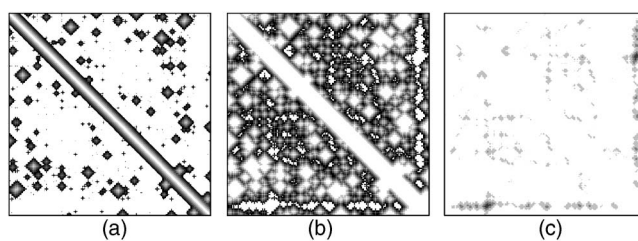

FIG. 3. Almost invariant spectra $\rho \times \lambda$ of the matrices $M_\ell, \ell = 7 - 10$, for a Watts-Strogatz small network starting from a next-nearest neighbor linear chain and $p_w = 0.2$.

Though quite distinct, all of them share a certain homogeneity character. We failed to find NI for Barabasi's scale-free as well as Apollonian networks, where nodes with quite distinct connections coexist. This suggests that NI may not be observed for highly inhomogeneous networks. Moreover, it shows that NI is not equivalent to other concepts of scale and geometrical invariance that are used to characterized other networks. This proposed framework may be useful to characterize actual networks.

[1] D. J. Watts *Small Worlds: The Dynamics of Networks Between Order and Randomness* (Princeton University Press, Princeton, NJ, 1999).

[2] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, UK, 2003).

[3] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, UK, 2004).

[4] R. Albert and A. L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).

[5] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).

[6] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[7] A. L. Barabasi and R. Albert, Science **286**, 509 (1999).

[8] P. Erdös and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).

[9] J. E. Whitesitt, *Boolean Algebra and its Applications* (Dover, New York, 1995).

[10] A. Fronczak, J. A. Holyst, M. Jedynak, and J. Sienkiewicz, Physica A **316**, 688 (2002).

[11] B. Bolobás, *Random Graphs* (Academic, London, 1985).

[12] D. M. Cvetkovic, M. Dods, and H. Sachs, *Spectra of Graphs* (Academic, New York, 1979).

[13] J. S. Andrade, Jr., H. J. Herrmann, R. F. S. Andrade, and L. R. da Silva, Phys. Rev. Lett. **94**, 018702 (2005).

[14] J. P. K. Doye and C. P. Massen, Phys. Rev. E **71**, 016128 (2005).

[15] R. F. S. Andrade and J. G. V. Miranda, Physica A **356**, 1 (2005).

[16] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt-Saunders, New York, 1976).

[17] E. N. Economou, *Green's Function in Quantum Physics* (Springer, Berlin, 1979).

[18] C. E. T. Gonçalves da Silva, and B. Koiller, Solid State Commun. **40**, 215 (1981).

[19] E. P. Wigner, Ann. Math. **62**, 548 (1955).

[20] A. Crisanti, G. Paladin, and A. Vulpiani, *Products of Random Matrices in Statistical Physics*, Springer, Berlin, 1993.