Topological structural classes of complex networks

Ernesto Estrada*
Complex Systems Research Group, X-Rays Unit, RIAIDT, Edifico CACTUS, University of Santiago de Compostela, Santiago de Compostela 15782, Spain
(Received 27 October 2006; published 19 January 2007)

We use theoretical principles to study how complex networks are topologically organized at large scale. Using spectral graph theory we predict the existence of four different topological structural classes of networks. These classes correspond, respectively, to highly homogenous networks lacking structural bottlenecks, networks organized into highly interconnected modules with low inter-community connectivity, networks with a highly connected central core surrounded by a sparser periphery, and networks displaying a combination of highly connected groups (quasicliques) and groups of nodes partitioned into disjoint subsets (quasibipartites). Here we show by means of the spectral scaling method that these classes really exist in real-world ecological, biological, informational, technological, and social networks. We show that neither of three network growth mechanisms—random with uniform distribution, preferential attachment, and random with the same degree sequence as real network—is able to reproduce the four structural classes of complex networks. These models reproduce two of the network classes as a function of the average degree but completely fail in reproducing the other two classes of networks.

DOI: 10.1103/PhysRevE.75.016103 PACS number(s): 89.75.Hc, 89.75.Fb, 02.10.Ox, 02.70.Hm

I. INTRODUCTION

Classification is usually part of the early stage of understanding science. It allows us to sort our data into unifying categories in order to obtain a better understanding of their meaning in a more efficient way. Complex networks are a good example of systems which pervade different scientific disciplines ranging from natural to technological and social sciences [1–5]. Due in part to this abundance of complex networks, as well as due to their strategic and scientific importance, there are several possible classification schemes for networks. These classification systems can be based on network functionality or on network structure. For instance, a "classical" division of networks into biological, social, informational, technological, and ecological clearly reflects the functionality of these systems [3]. On the other hand, there are several classification schemes based on network structure. For instance, complex networks can be classified according to the existence or not of the "small-world" property [6,7] or according to their degree distribution. The last classification permits classification of networks as "scale-free" [8] if their node degree distribution decays as power-law, "broad-scale" networks, which are characterized by a connectivity distribution that has a power-law regime followed by a sharp cutoff, or "single-scale" networks in which degree distribution displays a fast decaying tail [9]. Even scale-free networks have been classified into two different subclasses according to their exponent in the power-law distribution of the betweenness centrality [10].

Each one of these classification schemes reproduces some different characteristics of complex networks. "Small-worldness" [6] and "scale-freeness" [8] reflect global organizational principles of complex systems. The first characterizes the relatively small separation among pairs of nodes and the high cliquishness of some real-world networks [6]. The second reproduces the presence of a few highly connected hubs that stay glued in the vast majority of poorly connected nodes in certain networks [8]. Both properties are of great relevance in analyzing other important properties of complex networks, such as disease propagation [11–13] or robustness against targeted or random attacks [14–16]. However, there are important organizational principles of complex networks which escape the analysis of these global network characteristics. The most relevant example is the existence of network communities, which has lead to great efforts in finding such structures in complex networks [17–22].

Taking into account the community structure we can classify complex networks into those having a clear community structure and those not having it. However, those networks having a clear community structure can also be classified according to the architectural organization of these communities in the network. For instance, it is possible that some networks display several communities of highly interconnected nodes which display very low intercommunity connectivity. We can think of this network as dominated by a few quasicliques [23], i.e., the highly interconnected communities. Other networks can be characterized by a central community of highly interconnected nodes surrounded by one or several communities forming a sparse periphery. Thus the "communication" between nodes in the periphery is mainly carried out by passing through the central core. We can think of these networks as dominated by a set of nodes in the periphery which are not close to those they are linked with, forming a structure of quasibipartite communities [23]. Finally, it is possible that both types of communities exist in a network forming a mixture of quasicliques and quasibipartite modules, without a predominance of either of them over the other.

In this work we consider the theoretical existence of different structural classes of networks and we introduce a mathematical method to identify them on the basis of the network spectra. Then, we show the existence of the four

*Email address: estrada66@yahoo.com

structural classes of complex networks predicted by the theory in real-world systems by analyzing 61 networks representing ecological, biological, informational, technological, and social systems. Finally, we analyze how some of the existing network growth models reproduce some, but not all, of these classes of networks. We analyze the random generation of networks with uniform degree distribution, with preferential attachment, and with the same degree sequence as real-world networks.

II. PRELIMINARIES

Here we represent a complex network by a graph $G=(V,E)$, where V and E are the set of nodes and links, respectively. Let G be a graph having N nodes. Then the adjacency matrix of G, $\mathbf{A}(G)=\mathbf{A}$, is a square, symmetric matrix of order N, whose elements A_{ij} are ones or zeroes if the corresponding nodes are adjacent or not, respectively. This matrix has N (not necessarily distinct) real-valued eigenvalues, which are denoted here by $\lambda_1,\lambda_2,\ldots,\lambda_N$, and are assumed to be labeled in a nonincreasing manner: $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_N$. The set of eigenvalues of $\mathbf{A}$ together with their multiplicities is denoted here as the spectrum of G. Let $\boldsymbol{\gamma}_j$ be an orthonormal eigenvector corresponding to the eigenvalue λ_j. Then, $\gamma_j(i)$ designates the component of this eigenvector corresponding to the ith node in the network. Through the whole paper we will use base-10 logarithms designated as $\log=\log_{10}$.

In order to characterize the centrality of the nodes in a network we have introduced a measure named the *subgraph centrality* of a node, ${}^{S}C(i)$ [24], which is based on the total number of closed walks (CWs) in a network. A walk of length l is any sequence of (not necessarily) different vertices $v_1,v_2,\cdots,v_l,v_{l+1}$ such that for each $i=1,2,\cdots,l$ there is an edge from v_i to v_{i+1}. A CW of length l is a walk in which $v_{l+1}=v_1$. This measure can be expressed as the sum of contributions coming from even and odd CWs [25]. In this context, the *odd subgraph centrality* of a complex network represents a weighted sum of the CWs of odd length starting and ending in a node in which the longest walks receive lower weights in the sum. We have shown that this measure can be expressed in terms of the network spectrum in the following way [24,25]:

$$ {}^{S}C_{odd}(i) = \sum_{j=1}^{N} [\gamma_j(i)]^2 \sinh(\lambda_j) \tag{1}$$

It is straightforward to realize that we can write expression (1) in the following form:

$$ {}^{S}C_{odd}(i) = [\gamma_1(i)]^2 \sinh(\lambda_1) + \sum_{j=2}^{N} [\gamma_j(i)]^2 \sinh(\lambda_j), \tag{2}$$

where λ_1 and $\boldsymbol{\gamma}_1$ are the principal (Perron-Frobenius) eigenvalue and eigenvector of the network, respectively. This expression can be represented in a logarithmic scale in the following form:

$$\log \gamma_1(i) = 0.5 \log\left[{}^{S}C_{odd}(i) - \sum_{j=2}^{N} [\gamma_j(i)]^2 \sinh(\lambda_j) \right] - 0.5 \log[\sinh(\lambda_1)] \tag{3}$$

This expression can be represented as a straight line in a plot of $\log \gamma_1(i)$ versus $\log\{[\gamma_1(i)]^2 \sinh(\lambda_1)\}$, with a slope of 0.5 and intercept of $-0.5 \log[\sinh(\lambda_1)]$. We have previously found that there is a particular case in which [26]

$$ [\gamma_1(i)]^2 \sinh(\lambda_1) \gg \sum_{j=2}^{N} [\gamma_j(i)]^2 \sinh(\lambda_j) \tag{4}$$

and

$$ {}^{S}C_{odd}(i) \approx [\gamma_1(i)]^2 \sinh(\lambda_1). \tag{5}$$

For instance, this situation can happens when $\lambda_1 \gg \lambda_2 > \cdots > \lambda_N$. This is a characteristic of the good expansion networks (GENs), which are networks in which every subset S of nodes ($S \leqslant 50\%$ of the nodes) has a neighborhood that is larger than some "expansion factor" ϕ multiplied by the number of nodes in S. A neighborhood of S is the set of nodes which are linked to the nodes in S [27]. Formally, for each vertex $v \in V$ (where V is the set of nodes in the network), the neighborhood of v, denoted as $\Gamma(v)$ is defined as $\Gamma(v)=\{u \in V | (u,v) \in E\}$ (where E is the set of links in the network). Then, the neighborhood of a subset $S \subseteq V$ is defined as the union of the neighborhoods of the nodes in S: $\Gamma(S)=\cup_{v \in S}\Gamma(v)$ and the network has GE if $\Gamma(v) \geqslant \phi|S| \forall S \subseteq V$ [27]. GENs show excellent communication properties due to the absence of bottlenecks—a small set S for which $G \backslash S$ has at least two large connected components. We have shown that GENs are more robust to targeted attacks than non-GEN networks irrespective of their degree distribution [28]. It is known that a network having GE properties is characterized by a large spectral gap $(\lambda_1-\lambda_2)$ [27]. In particular, random regular networks are expected to have big spectral gaps with high probability, and thus are GENs. In the case in which the condition (5) is obeyed we will have an "ideal" case of a network without bottlenecks and GE properties, which displays a perfect spectral scaling between $\log \gamma_1(i)$ and $\ln {}^{S}C_{odd}(i)$ [26],

$$\log \gamma_1^{Ideal}(i) = 0.5 \log {}^{S}C_{odd}(i) - 0.5 \log[\sinh(\lambda_1)]. \tag{6}$$

III. THEORETICAL MODELS

Here we will consider the possibility that deviations from the ideal behavior represented by Eq. (6) can occur. This situation is present when the condition (4) is not perfectly fulfilled. Then we can account for these deviations from ideality by measuring the departure of the points from the perfect straight line respect to $\log \gamma_1^{Ideal}(i)$:

$$\Delta \log \gamma_1(i) = \log \frac{\gamma_1(i)}{\gamma_1^{Ideal}(i)} = \log \left\{ \frac{[\gamma_1(i)]^2 \sinh(\lambda_1)}{{}^{S}C_{odd}(i)} \right\}^{0.5} \tag{7}$$

According to the values of $\Delta \log \gamma_1(i)$ there will be four different classes of complex networks:

Class I: Networks displaying perfect spectral scaling

$$\Delta \log \gamma_1(i) \cong 0, \quad \forall i \in V \Rightarrow [\gamma_1(i)]^2 \sinh(\lambda_1) \cong {}^{S}C_{odd}(i); \tag{8}$$

Class II: Networks displaying spectral scaling with negative deviations

$$\Delta \log \gamma_1(i) \leqslant 0 \Rightarrow [\gamma_1(i)]^2 \sinh(\lambda_1) \leqslant {}^{S}C_{odd}(i), \quad i \in V; \tag{9}$$

Class III: Networks displaying spectral scaling with positive deviations

$$\Delta \log \gamma_1(i) \geqslant 0 \Rightarrow [\gamma_1(i)]^2 \sinh(\lambda_1) \geqslant {}^{S}C_{odd}(i), \quad i \in V; \tag{10}$$

Class IV: Networks displaying spectral scaling with mixed deviations

$$\Delta \log \gamma_1(p) \leqslant 0, \quad p \in V \quad \text{and} \quad \Delta \log \gamma_1(q) > 0, \quad q \in V. \tag{11}$$

We have previously analyzed the case for networks showing a perfect spectral scaling, which, in general, correspond to networks without topological bottlenecks, or in other words to networks with GE properties [26]. Now, we will consider the other theoretical possibilities of structural classes of complex networks. First, we consider the case of spectral scaling with negative deviations (class II). Here, we only need to consider the case where $[\gamma_1(i)]^2 \sinh(\lambda_1) < {}^{S}C_{odd}(i)$ because $[\gamma_1(i)]^2 \sinh(\lambda_1) = {}^{S}C_{odd}(i)$ produces null deviations from the perfect scaling, i.e., $\Delta \log \gamma_1(i) = 0$.

Using expression (2) for ${}^{S}C_{odd}(i)$ we obtain

$$[\gamma_1(i)]^2 \sinh(\lambda_1) < [\gamma_1(i)]^2 \sinh(\lambda_1) + \sum_{j=2} [\gamma_j(i)]^2 \sinh(\lambda_j), \tag{12}$$

which obviously implies that

$$\sum_{j=2} [\gamma_j(i)]^2 \sinh(\lambda_j) > 0. \tag{13}$$

Taking into account that we are considering networks without self-loops, the adjacency matrix has zeroes along its principal diagonal, which means that its spectrum must have both positive and negative eigenvalues. Because $\lambda_1 > 0$ we will designate by Σ_+ and Σ_- the sums corresponding to positive and negative eigenvalues for $j \geqslant 2$. Then, we can write the left part of inequality (13) as follows:

$$\sum_{j=2}^{N} [\gamma_j(i)]^2 \sinh(\lambda_j) = \sum_{+} [\gamma_j(i)]^2 \sinh(\lambda_j) + \sum_{-} [\gamma_j(i)]^2 \sinh(\lambda_j). \tag{14}$$

Now, we can rewrite the inequality (13) in terms of the sums of positive and negative eigenvalues,

$$\sum_{+} [\gamma_j(i)]^2 \sinh(\lambda_j) + \sum_{-} [\gamma_j(i)]^2 \sinh(\lambda_j) > 0. \tag{15}$$

Then, because $\Sigma_-[\gamma_j(i)]^2 \sinh(\lambda_j) < 0$ we immediately obtain the following new inequality:

$$\left| \sum_{+} [\gamma_j(i)]^2 \sinh(\lambda_j) \right| > \left| \sum_{-} [\gamma_j(i)]^2 \sinh(\lambda_j) \right|. \tag{16}$$

It is known from spectral clustering techniques that the eigenvectors corresponding to positive eigenvalues give a partition of the network into clusters of tightly connected nodes [29–34]. On the contrary, the eigenvectors corresponding to negative eigenvalues make partitions in which nodes are not close to those which they are linked, but rather with those with which they are not linked. In other words, the nodes will be close to other nodes which have similar patterns of connections with other sets of nodes, i.e., nodes to which they are structurally equivalent [29–34]. In the case of the eigenvectors corresponding to positive eigenvalues the nodes corresponding to larger components tend to form quasicliques. That is, clusters in which every two nodes tend to interact with each other. On the contrary, for eigenvectors corresponding to negative eigenvalues, nodes tend to form quasibipartites, i.e., nodes are partitioned into disjoint subsets with high connectivity between sets but low internal connectivity.

Accordingly, the inequality $|\Sigma_+[\gamma_j(i)]^2 \sinh(\lambda_j)| > |\Sigma_-[\gamma_j(i)]^2 \sinh(\lambda_j)|$ is telling us that the corresponding network is dominated by partitions into quasicliques more than into quasibipartites. In other words, these networks are characterized by two or more clusters of highly interconnected nodes which display a low intercluster connectivity. A simple model illustrating these topological characteristics can be created from a highly connected network having one or more holes. For the sake of simplicity we will consider a small network with only one central hole. In this case the clusters formed at the four corners of the network display large internal connectivity because they were created from a highly connected network. However, the intercluster connectivity is dramatically reduced due to the presence of the central hole in the network. In Fig. 1(a) we illustrate this model, where we can see that the network effectively displays negative deviations in the spectral scaling [Fig. 1(b)]. In the Fig. 1(c) we have plotted the eigenvectors corresponding to the second and third largest eigenvalues (positive eigenvectors), and in Fig. 1(d) the same plot is illustrated for the two largest negative eigenvectors. As can be seen, the topological structure of this network is "dominated" by the presence of quasicliques more than for the presence of quasibipartites. This is observed by a clear presence of well-defined, highly interconnected clusters in Fig. 1(c) and the absence of a clear structure of quasibipartites in Fig. 1(d).

The idea of plotting the 2nd and 3rd largest (positive) eigenvectors as well as the two largest negative ones is justified by the following known facts [31]. First, the components of the largest positive eigenvector of the adjacency matrix of a network have the same sign. If the network is regular they will have also the same value. However, the second largest eigenvector, which is orthonormal to the first eigenvector, will have both positive and negative components. In a similar way, the third largest positive eigenvector will have a different pattern of positive and negative signs,

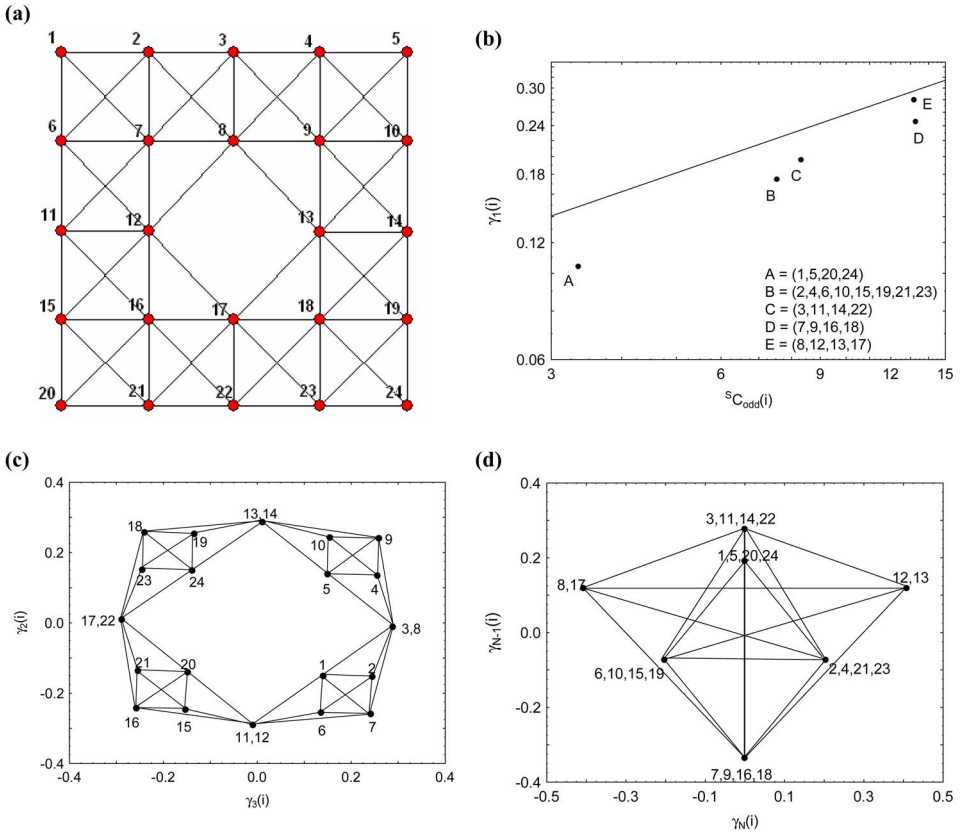

FIG. 1. (Color online) (a) Model of network in class II consisting of highly interconnected clusters, e.g., the four corners of the networks, which display low intercluster connectivity due to the central hole. (b) Spectral scaling of the network in (a) showing negative deviation from perfect scaling (straight line). The letters on the points correspond to series of nodes in the network as indicated in the legend of this plot. (c) Plot of the network (a) using the second and third largest eigenvectors of the adjacency matrix. (d) Plot of the network (a) using the largest negative eigenvectors of the adjacency matrix.

and so forth. Then if we arrange the rows and columns of the adjacency matrix according to the signs of the first and second largest eigenvectors we will obtain a partition of this matrix into biants. These biants correspond to the partition of the network into clusters of highly interconnected nodes, i.e., quasicliques. The use of sign patterns of further eigenvectors will partition the matrix into quadrants, octants, etc. At the end, the sign patterns of the last two eigenvectors, the two largest negative ones, we will obtain a partition of the network in n-tants, in which nodes are closer to those with which they are not linked, i.e., quasibipartites.

Now we can follow a similar procedure for the networks with positive deviations of the spectral scaling (class III). After the same algebraic manipulations we arrive at the following expression analogous to (16):

$$\left|\sum_{+}[\gamma_j(i)]^2 \sinh(\lambda_j)\right| < \left|\sum_{-}[\gamma_j(i)]^2 \sinh(\lambda_j)\right|. \quad (17)$$

According to this inequality the topological organization of the nodes/links in these networks is "dominated" by the negative eigenvalues. This means that networks in the structural class III are characterized by the dominance of quasibipartites more than of quasicliques. Then, we can elaborate a model explaining the general features of these networks by considering a central core of highly interconnected nodes surrounded by a periphery of nodes displaying low connectivity with the central core and among themselves. This model corresponds to the contrary case of the model for networks in class II, as can be seen in Fig. 2(a). In fact, these networks display positive deviations in the spectral scaling, as can be seen in Fig. 2(b), and they are dominated by the presence of quasibipartites [Fig. 2(d)] more than by quasicliques [Fig. 2(c)].

Now it is obvious that when $[\gamma_1(i)]^2 \sinh(\lambda_1) \cong {}^SC_{odd}(i)$ (class I) the network is formed by a tightly connected homogenous cluster, which is characterized by the leading eigenvalue. This situation is very clear from the fact that $[\gamma_1(i)]^2 \sinh(\lambda_1) \gg \Sigma_{j=2}^N [\gamma_j(i)]^2 \sinh(\lambda_j)$, indicating the predominance of one large quasiclique formed by almost all nodes in the network.

On the contrary, the networks in class IV, which have a mixture of positive and negative deviations, are characterized by a combination of both quasicliques and quasibipartites, without the predominance of either structure over the other. This situation can be represented by a network formed by two highly interconnected parts linked directly by an edge. On the one hand, the central nodes connecting both highly interconnected clusters display larger connectivity to all other nodes in the network than the one expected from their local cliquishness, i.e., they display positive deviations from the perfect scaling. On the other hand, the nodes on one side of the graph are not well-connected to the nodes on the other side, despite that they are internally highly connected. Consequently, these nodes display negative deviations from the perfect scaling.

We can quantify the degree of deviation of the nodes from the ideal spectral scaling by accounting for the mean square error of all points with positive and negative deviations in the spectral scaling, respectively,

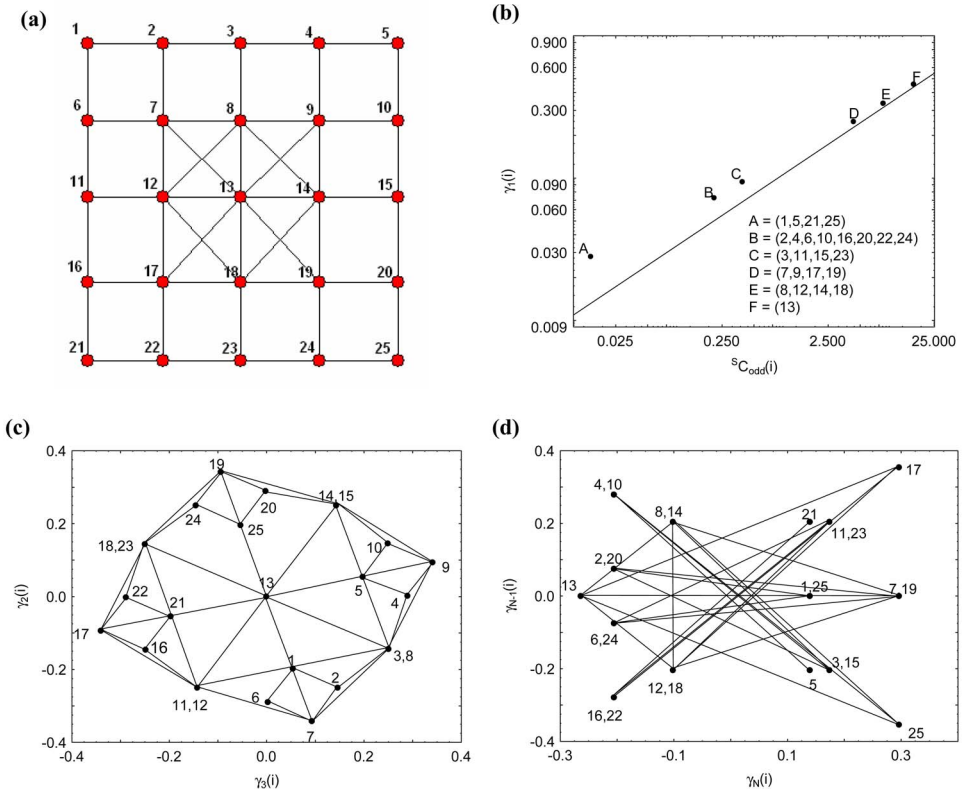

FIG. 2. (Color online) (a) Model of network in class III consisting of a highly interconnected central core surrounded by a sparser periphery. (b) Spectral scaling of the network in (a) showing positive deviation from perfect scaling (straight line). The letters on the points correspond to *a* series of nodes in the network as indicated in the legend of this plot. (c) Plot of the network (a) using the second and third largest eigenvectors of the adjacency matrix. (d) Plot of the network (a) using the largest negative eigenvectors of the adjacency matrix.

$$\xi^{+} = \sqrt{\frac{1}{N_{+}} \sum_{+} \left(\log \frac{\gamma_1(i)}{\gamma_1^{Ideal}(i)} \right)}$$

and

$$\xi^{-} = \sqrt{\frac{1}{N_{-}} \sum_{-} \left(\log \frac{\gamma_1(i)}{\gamma_1^{Ideal}(i)} \right)},$$

where Σ_+ and Σ_- are the sums carried out for the N_+ points having $\Delta \log \gamma_1(i) > 0$ and for N_- having $\Delta \log \gamma_1(i) < 0$, respectively.

IV. STRUCTURAL CLASSES OF REAL-WORLD NETWORKS

We study here 61 real-world complex networks accounting for ecological, biological, protein secondary structures, informational, technological, and social systems. The ecological networks studied correspond to the following food webs [7]: Benguela, Bridge Brook, Canton Creek, Chesapeake Bay, Coachella Valley, El Verde rainforest, Grassland, Little Rock Lake, Reef Small, Scotch Broom, Shelf, Skipwith Pond, St. Marks Seagrass, St. Martin Island, Stony, and Ythan Estuary (1), with and without parasites (2). The biological networks correspond to the protein-protein interaction networks (PINs) for *Saccharomyces cerevisiae* (yeast) [23] and for the bacterium *Helicobacter pylori* [35], and three transcription interaction networks concerning *E. coli*, yeast, and sea urchins and the neural network in *C. elegans* [36]. The protein residue networks correspond to the proteins with Protein Data Bank (PDB) codes 1aaf, 1gds, 1bmtA, 1afrA, 1alkA, 1aozA, 1gof, 1kit, 1qba, and 1bglA. In these networks each residue is represented as a single node, centered on C_β atoms. Then a contact map is represented by taking a 7-Å cutoff radius [37]. The informational networks include two semantic networks, one based on *Roget's Thesaurus of English* (Roget) and the other on the *Online Dictionary of Library and Information Science* (ODLIS) and four citation networks: one consisting of papers published in the *Proceedings of Graph Drawing* in the period 1994–2000 (GD), papers published in the field of "Network Centrality" (Centrality), papers published or citing articles from *Scientometrics* for the period 1978–2000 (SciMet), and papers containing the phrase "Small World" [38]. The technological systems represented by networks correspond to three electronic sequential logic circuits parsed from the ISCAS89 benchmark set, where nodes represent logic gates and flip-flops [36], the airport transportation network in the US in 1997 [38], the Internet at the autonomous systems (AS) level as from April 1998 [39], and five software networks: Abi, Digital, MySQL, VTK, and XMMS [40]. Finally, the social networks studied here include a network of the corporate elite in the U.S. [41], a scientific collaboration network in the field of computational geometry (Geom), inmates in prison, injectable drug users (IDUs), Zachary karate club, college students on a course about leadership [38], a sexual network in Colorado Springs [42], a collaboration between Jazz musicians [43], the friendship ties among 31 physicians (Galesburg), the friendship ties among the employees in a small hi-tech computer firm which sells, installs, and maintains computer systems (High Tech), and a communication network within a small enterprise (Saw Mill) [38].

In Table I we illustrate the values of ξ^- and ξ^+ for the 61 real-world networks studied as well as their number of nodes

TABLE I. Real-world complex networks studied in this work, their size (N), number of links (E), the values of negative (ξ^-) and positive (ξ^+) departures from ideal spectral scaling and classification into the four different classes found in this work.

No.	Network	N	E	ξ^-	ξ^+	Class
	Ecological					
1	Benguela	29	191	6.30×10^{-3}	0.000	I
2	Coachella	30	241	7.18×10^{-5}	0.000	I
3	Skipwith	35	353	6.16×10^{-5}	0.000	I
4	St. Martin	44	218	1.10×10^{-3}	1.60×10^{-3}	I
5	St. Marks	48	218	5.00×10^{-3}	2.20×10^{-3}	I
6	Reef Small	50	503	0.000	4.31×10^{-5}	I
7	Bridge Brook	75	542	0.000	9.00×10^{-3}	I
8	Shelf	81	1451	0.000	6.75×10^{-5}	I
9	Ythan2	92	416	0.000	2.92×10^{-3}	I
10	Ythan1	134	593	1.11×10^{-4}	1.50×10^{-3}	I
11	El Verde	156	1439	4.50×10^{-5}	0.000	I
12	Little Rock	181	2318	3.72×10^{-5}	0.000	I
13	Scotch Broom	154	366	0.047	7.90×10^{-3}	II
14	Canton	108	707	0.000	0.183	III
15	Stony	112	830	0.000	0.219	III
16	Chesapeake	33	71	0.085	0.097	IV
17	Grassland	75	113	0.792	0.233	IV
	Biological					
18	Neurons	280	1973	3.83×10^{-4}	2.32×10^{-5}	I
19	PIN-1	2224	6608	0.142	2.36×10^{-4}	II
20	Trans Urchins	45	80	1.072	0.084	IV
21	Trans Ecoli	328	456	1.231	0.338	IV
22	Trans Yeast	662	1062	0.784	0.822	IV
23	PIN-2	710	1396	0.116	0.099	IV
	Proteins					
24	1aaf	55	119	2.499	4.45×10^{-3}	II
25	1gds	151	443	0.871	0.000	II
26	1bmtA	246	813	0.820	0.000	II
27	1afrA	345	1122	0.872	0.000	II
28	1alkA	449	1573	0.925	0.000	II
29	1aozA	552	1749	0.978	0.000	II
30	1gof	639	2208	1.072	0.000	II
31	1kit	757	2476	2.896	0.000	II
32	1qba	863	2928	1.619	0.000	II
33	1bglA	1021	3266	2.122	0.000	II
	Informational					
34	Centrality	118	613	9.59×10^{-5}	5.30×10^{-5}	I
35	Small World	233	994	3.70×10^{-3}	1.11×10^{-4}	I
36	ODLIS	2898	16376	4.06×10^{-6}	1.57×10^{-5}	I
37	Roget	994	3640	0.230	0.000	II
38	SciMet	2678	10368	0.102	4.50×10^{-4}	II
39	GD	49	635	0.467	0.041	IV
	Technological					
40	USAir97	332	2126	9.04×10^{-5}	0.000	I
41	Internet97	3015	5156	6.21×10^{-4}	1.20×10^{-3}	I
42	Internet98	3522	6324	1.55×10^{-3}	9.44×10^{-4}	I

TABLE I. (*Continued.*)

No.	Network	N	E	ξ^-	ξ^+	Class
43	MySQL	1480	4140	1.966	4.14×10^{-4}	II
44	Electronic1	122	189	0.594	0.808	IV
45	Digital	150	198	0.216	0.332	IV
46	Electronic2	252	399	0.753	0.737	IV
47	Electronic3	512	819	0.535	1.191	IV
48	VTK	771	1357	0.204	0.037	IV
49	XMMS	971	1802	1.450	0.105	IV
50	Abi	1035	1719	0.498	0.061	IV
	Social					
51	Jazz	1265	38356	1.03×10^{-3}	0.000	I
52	High Tech	33	91	0.108	6.70×10^{-3}	II
53	Drugs	616	2012	1.390	0.000	II
54	Corporate elite	1586	11540	0.053	0.000	II
55	Geom	3621	9461	0.452	0.000	II
56	Galesburg	31	67	0.279	0.079	IV
57	College	32	96	0.100	0.040	IV
58	Zachary	34	78	0.088	0.031	IV
59	Saw Mill	37	62	0.945	0.933	IV
50	Prison	67	142	0.261	0.119	IV
61	ColoSpring	324	347	0.837	0.764	IV

(N) and links (E). Using the values of ξ^- and ξ^+ we have classified these networks into the four different classes which are predicted to exist from a theoretical point of view. We have carried out a canonical discriminant analysis (CDA) [44] for the 61 networks studied using $\log(\xi^-+10^{-3})$ and $\log(\xi^++10^{-3})$ as classifiers, where the sum of the constant 10^{-3} is necessary to avoid indeterminacies due to zero values. In Fig. 3 we can see the main factors (roots) which perfectly separate the networks studied into the four different structural classes, which were predicted to exist theoretically. The first root mainly separates networks in class I from those in class IV, while root 2 makes the separation of these two classes from classes II and III.

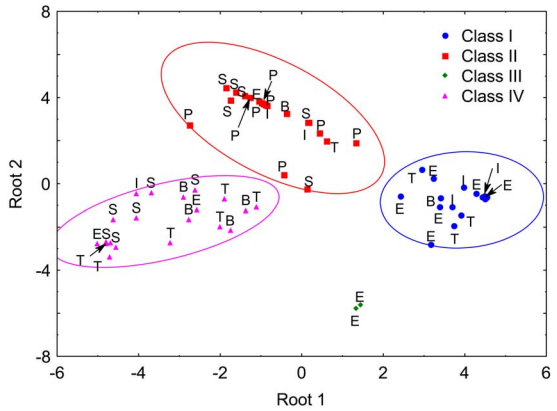

FIG. 3. (Color online) Plot of the two principal roots obtained in the canonical discriminant analysis (CDA) of the 61 networks studied in this work and classified into four different structural classes. Ellipses correspond to 95% of confidence in the CDA.

As can be seen in Table I and Fig. 3, we present here empirical evidence for the existence of the four structural classes of complex networks in real-world systems. Classes I, II, and IV are equally represented among the 61 networks studied, i.e., there are about 32% of networks in each of these classes. On the contrary, there are only two networks in class III, which correspond to the ecological systems of Canton Creek and Stony Stream. In general, most ecological networks correspond to class I (70%) and they represent the only systems in which the four classes of networks are represented. Most biological networks studied correspond to class IV (67%), while all protein secondary structure networks correspond to class II. Informational networks are mainly classified into two classes: class I (50%) and class II (33.3%). On the other hand, technological networks are mainly in class IV (64%), while 27% correspond to class I. Social networks also display great homogeneity in their structural classes as they correspond mainly to classes II and IV (91%).

According to the models we have created in the previous section it is easy to understand the global topological characteristics of the networks in every structural class. As can be seen in Fig. 4(a) for the case of the food web of Coachella valley, class I networks display large structural homogeneity. They are characterized by the lack of structural bottlenecks which separate large regions of the network by disconnecting relatively few nodes/links, which definitively improves their robustness to biodiversity loss [45]. In Fig. 4(b) we also illustrate the perfect spectral scaling obtained for this network.

Networks in class II are characterized by two or more highly interconnected clusters which are connected to each

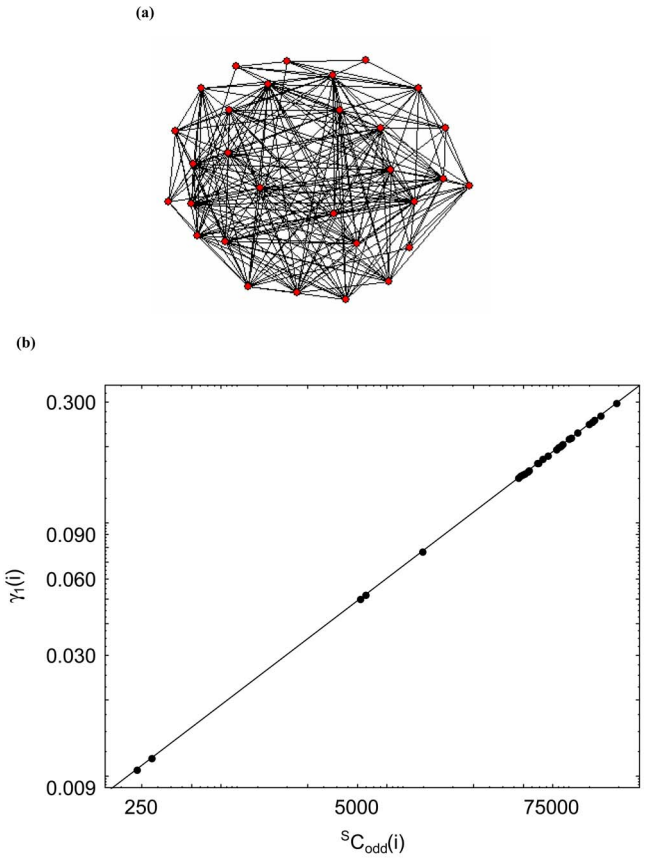

FIG. 4. (Color online) (a) Illustration of a network with good expansion properties corresponding to the Benguela food web. (b) Perfect spectral scaling of the Benguela food web indicating typical characteristics of class I networks.

other by relatively few nodes/links. The prototype of this class of networks is the protein residue network. In Fig. 5 we illustrate the structure of one representative residue network corresponding to the human immunodeficiency virus type 1 (HIV-1) capsid protein (1 gds). It is obvious from this figure and in the general case that proteins are formed by highly interconnected modules or motifs, formed by the secondary structure elements, such as helices and sheets. There are also some interactions between these motifs, which forms the clusters of nodes (residues) with high internal connectivity. However, the existence of cavities is known in proteins [46–49], which corresponds to the holes in our model for class II networks. These holes can be seen in the cartoon structure of the protein [Fig. 5(a)] as well as in the plot of the second and third largest positive eigenvalues [Fig. 5(c)], which is clearly dominating over the structure formed by plotting the largest negative eigenvalues [Fig. 5(d)]. The corresponding spectral scaling displays clear characteristics of networks with negative deviations, which correspond to class II [Fig. 5(b)]. In addition, we have explored the spectral scaling of 595 protein residue networks and we have obtained the same characteristics for all these networks, which will be published elsewhere.

The class III of networks is characterized by a highly connected central core surrounded by a sparsely connected periphery. These are the least abundant networks in our dataset as they are represented only by two food webs. However, we think that further explorations of larger pools of real-world complex networks can show the existence of these networks in other complex systems. In Fig. 6(a) we illustrate the network for the Canton food web, represented in a way that we can observe the highly dense core and the sparser

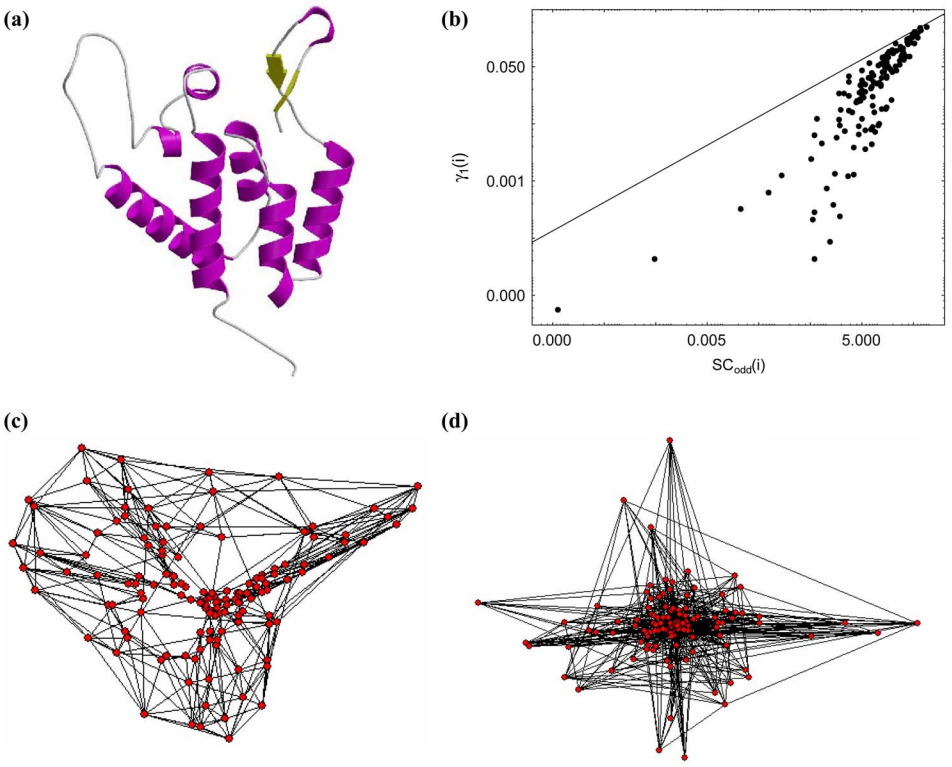

FIG. 5. (Color online) (a) Cartoon representation of the structure of human immunodeficiency virus type 1 (HIV-1) capsid protein (1gds). (b) Spectral scaling of this protein residue network showing negative deviations from perfect scaling, which are characteristic of class II networks. (c) Plot of the protein residue network using the second and third largest eigenvectors of the adjacency matrix. (d) Plot of the protein residue network using the largest negative eigenvectors of the adjacency matrix.

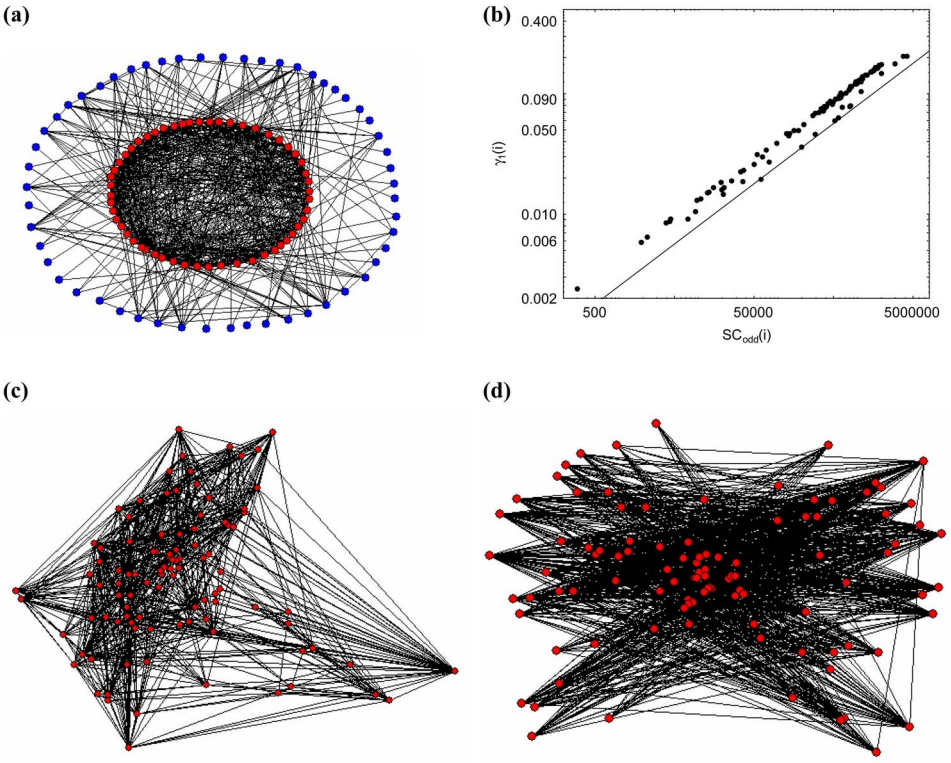

FIG. 6. (Color online) (a) Network representing the Canton Creek food web illustrating its core-periphery structure. (b) Spectral scaling of the Canton food web showing positive deviation from perfect scaling (straight line). (c) Plot of the Canton food web network using the second and third largest eigenvectors of the adjacency matrix. (d) Plot of the Canton food web network using the largest negative eigenvectors of the adjacency matrix.

periphery as well as its spectral scaling [Fig. 6(b)]. The structure of this network is clearly dominated by the formation of quasibipartites over the formation of quasicliques, as can be seen from Figs. 6(d) and 6(c), respectively.

The networks in class IV are characterized by the presence of both quasicliques and quasibipartites without a predominance of any of them over the other. This situation is illustrated in Fig. 7 for the social network of Saw Mill [Fig.

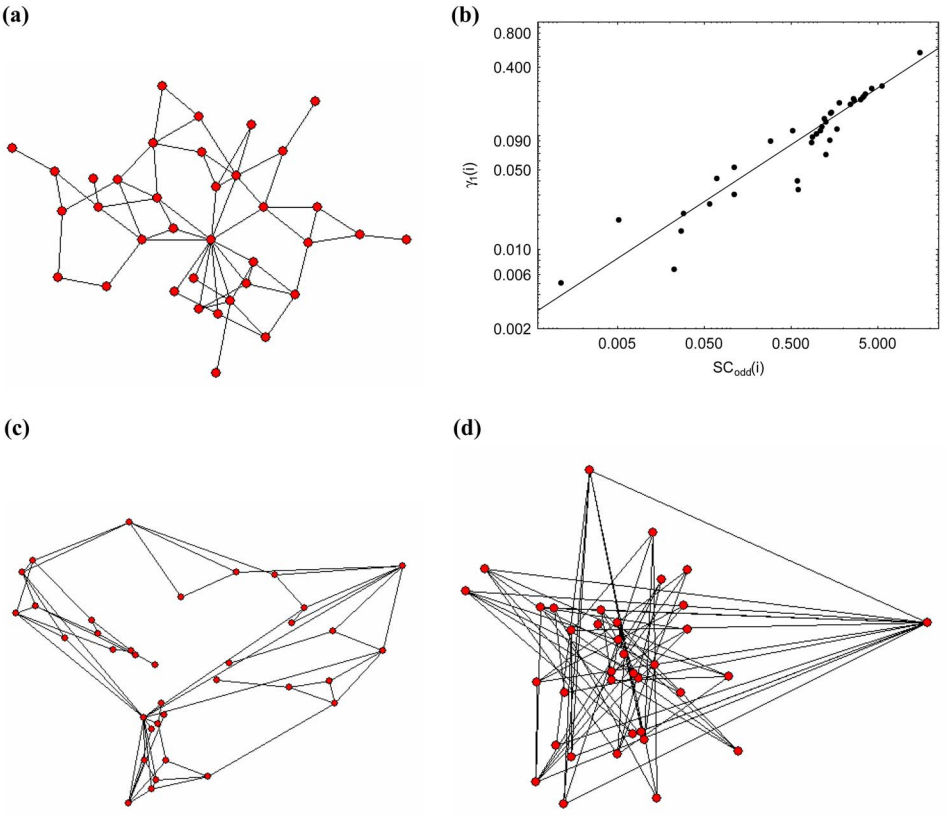

FIG. 7. (Color online) (a) Network representing the Saw Mill social network. (b) Spectral scaling of the Saw Mill social network showing positive and negative deviation from perfect scaling (straight line). (c) Plot of the Saw Mill social network using the second and third largest eigenvectors of the adjacency matrix. (d) Plot of the Saw Mill social network using the largest negative eigenvectors of the adjacency matrix.

7(a)], where we can see the spectral scaling having both positive and negative deviations [Fig. 7(b)] and that the topological structure of the network is dominated by neither positive [Fig. 7(c)] nor negative eigenvalues [Fig. 7(d)].

Another interesting characteristic of the community organization in complex networks can be extracted from the present study. As can be seen in Fig. 3, with the exception of the networks of residue interactions in proteins, there is not a clear definition of community structure as a function of the broad functional categories in which these networks are grouped. For instance, the communities in biological, social, or technological networks are not always structurally organized in the same way and these networks appear classified in more than one structural class.

V. FORMATION OF NETWORKS OF DIFFERENT CLASSES

A natural question that arises from the observation of the different classes of networks existing in natural and man-made complex systems is "what are the evolving mechanisms giving rise to the different classes of networks?" We do not pretend to answer this question exhaustively in this work. Instead we will investigate whether some of the existing theoretical models for network formation are able to reproduce the characteristics of networks in the structural classes found here. In particular, we investigate the random network growing model giving rise to networks with uniform degree distributions and the Barabási–Albert (BA) model [8,50]. In both models each random network starts with m nodes and new nodes are added consecutively in such a way that a new node is connected to exactly m of the already existing nodes, which are chosen randomly. The new edges are attached according to the probability distribution used, e.g., uniform distribution for the uniform model and the preferential attachment mechanism in the BA model.

We have studied random networks generated by these two growing mechanisms having $n=1000$ nodes by changing systematically the value of m from 2 to 8. For every value of m we have generated 100 random networks. Then, we have averaged the values of ξ^- and ξ^+ for every value of m. In Fig. 8 we illustrate the plot of ξ^- and ξ^+ in logarithmic scale for both growing mechanisms. We can see that the networks generated by both growing mechanisms which have low average degree $m \leqslant 3$ correspond to class IV. As the value of m increases a transition occurs in the network structure, giving rise to class I networks for values of $m \geqslant 4$ (BA model) or $m \geqslant 5$ (uniform model). In the case of the uniform growing model for $m=4$ we obtained some networks showing characteristics of class II networks. This sounds reasonable by considering that the average degree of the real-world networks in class II are around a value of 3.3. However, the networks obtained in this case by the growing models are more in the borderline of the transition than clearly in one or another class. In addition, we have to say that the average degree cannot be used as a rule of thumb for classifying real-world networks into the different structural classes found here. For instance, despite that most networks in class I display very large average degree, there are other networks

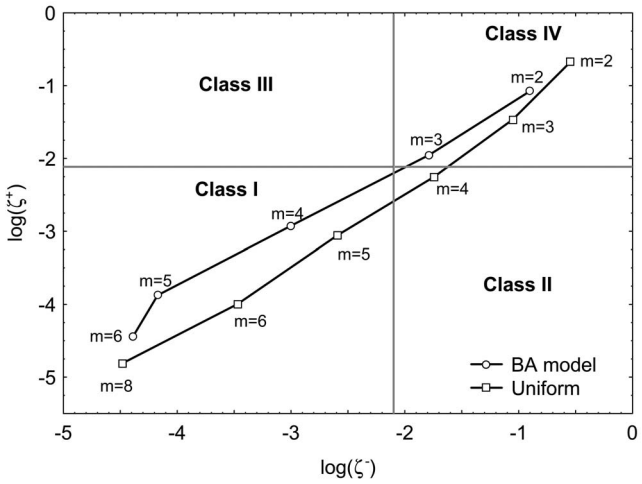

FIG. 8. Change in the deviations from perfect scaling (ξ^- and ξ^+) of random networks generated by two growing mechanisms with the change in the number of starting nodes (m). Uniform corresponds to random networks generated to follow a uniform degree distribution. BA corresponds to random networks generated by a Barabási–Albert preferential attachment mechanism. The plot is represented on the space of topological network classes found in this work and the logarithms are base 10.

in this class, such as Internet 1997 and 1998, which have very low average degrees, i.e., $\langle k \rangle = 1.71$ and $\langle k \rangle = 1.79$, respectively. Thus, we will further explore the evolution of networks in class II using real-world networks.

In closing, we have seen here that these two mechanisms for growing networks mainly reproduce the characteristics of one single class of networks, i.e., class I for all $m \geqslant 5$. Only for the limited cases of networks with very low average degree ($m \leqslant 3$) it is possible to obtain networks in class IV. However, neither of both growing models is able to clearly generate networks in classes II and III. The results obtained in this work showing that the BA growing model mainly generates networks with GE properties is expected from previous theoretical results. For instance, Gkantsidis *et al.* [51] have shown using arguments from max-flow min-cut theory that networks obeying power-law degree distribution have good expansion properties, in that they allow routing with $O(N \log^2 N)$ congestion, which is close to the optimal value of $O(N \log N)$ achieved by regular expanders.

The next step in our analysis is to study how the degree distribution can determine the structural class of a complex network. In this case we study four real-world networks pertaining to the different structural classes found in this work. They are the network of Centrality literature (class I), the protein residue network 1aozA (class II), the food web of Stony stream (class III), and the electronic circuit 1 (class IV). Then, we generate 100 random networks having the same number of nodes, the same average degree and the same degree sequence as the corresponding real-world network by using the Mfinder computer software [52]. Using the spectral scaling information for these randomly generated networks we calculate the values of ξ^- and ξ^+. In Table II we give the results of these calculations where we can see that the random models with the same degree sequence of the

TABLE II. Departure from ideal scaling of real-world networks in the four different classes found in this work and its comparison with random networks generated with the same degree sequence.

	ξ^-	ξ^+	Class	ξ^-	ξ^+	Class
Network	Real			Random[a]		
Centrality	9.59×10^{-5}	5.30×10^{-5}	I	1.50×10^{-5} (2.38×10^{-5})	5.13×10^{-5} (4.51×10^{-5})	I
1aozA	0.978	0.000	II	0.126 (0.008)	0.096 (0.006)	IV
Stony	0.000	0.219	III	2.52×10^{-5} (3.16×10^{-5})	2.83×10^{-5} (2.99×10^{-5})	I
Electronic1	0.594	0.808	IV	0.390 (0.223)	0.724 (0.112)	IV

[a]Standard deviations are given in parenthesis.

real-world networks reproduce very well the structural characteristics of the networks in classes I and IV. The Centrality network (class I) displays an average degree of $\langle k\rangle=5.19$ and an exponential degree distribution. On the other hand, Electronic1 (class IV) has $\langle k\rangle=1.55$ and uniform degree distribution. As we previously observed in both cases—large $\langle k\rangle$ (class I) and low $\langle k\rangle$ (class IV)—results are well reproduced by growing models based on uniform random networks or on a preferential attachment model. However, neither of the networks in classes II and III are well reproduced by the growing method based on the same degree sequences. In the case of the protein residue, the random network with the same degree sequence displays a structure characteristic of class IV networks instead of class II. In the other case, the random network having the same degree sequence than Stony food web is predicted to be on class I instead of class III. These results confirm our previous findings that the existing growing mechanisms for generating random networks do not reproduce the characteristics features of real-world networks in classes II and III. These random network models, however, are useful in reproducing the structure of complex networks in classes I and IV, which in some extension depends on the degree sequence and the average degree of these networks. Consequently, there is need for new theoretical models which reproduce the topological structures of real-world complex networks in all structural classes found in this work. Recently, Jungsbluth *et al.* [53] have also questioned whether random-network models, such as Erdös–Rényi, Small-World, and BA, really reflect all important properties of the real world, indicating that network models have to be more specific for each application.

VI. CONCLUSIONS

In this work we have found that there are four theoretically possible topological structures of complex networks. We have used a first-principles approach based on spectral graph theory to predict the existence of these topological structural classes. The first of such classes corresponds to networks displaying good expansion properties. That is, networks in which nodes and links are homogeneously distributed through the network in such a way that there are not structural bottlenecks. The other three classes correspond to different organizations of the community structure in the networks. For instance, class II corresponds to networks in which there are two or more communities of highly interconnected nodes which display low intermodule connectivity. In class III the networks display a typical "core-periphery" structure characterized by a highly interconnected central core surrounded by a sparser periphery of nodes. Finally, there are networks (class IV) displaying a combination of highly connected groups (quasicliques) and some groups of nodes partitioned into disjoint subsets (quasibipartites), without a predominance of any of both structures.

The method developed in this work, which is based on the spectral scaling approach, permits not only clear and effective differentiation among these types of networks but also understanding of the structural characteristics giving rise to these different topological organizations. Consequently, we have identified the existence of the four classes of networks in real-world systems by studying a large pool of networks representing ecological, biological, informational, technological, and social systems. While classes I, II, and IV are equally populated, each having about 32% of the total networks, class III is less frequent and only appeared in two ecological networks. We finally have explored the possible growing mechanisms determining the structural classes observed in this work. We found that a random growing mechanism giving rise to uniform distributions of node degrees and the preferential attachment mechanism of Barabási-Albert reproduces very well the characteristics of networks in group I when the average degree is larger than 5. For sparser networks, such as those having average degree lower than 3, both mechanisms reproduce the characteristics of networks in class IV. However, neither of both growing mechanisms are able to reproduce the topological organization of networks in classes II and III. Similar results are obtained when generating random networks with the same degree sequence than real-world networks. Our results confirm previous findings about the necessity of investigating new growth mechanisms for generating networks to model real-world systems.

ACKNOWLEDGMENTS

The author would like to thank J. A. Dunne, R. Milo, U.

Alon, J. Moody, V. Batagelj, J. Davis, J. Potterat, P. M. Gleiser, D. J. Watts, C. R. Myers, and C. Baysal for generously providing datasets. This work was partially supported by the "Ramón y Cajal" program, Spain.

[1] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] A.-L. Barabási and Z. N. Oltvai, Nat. Rev. Genet. **5**, 101 (2004).
[5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[6] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[7] J. A. Dunne, R. J. Williams, and N. D. Martinez, Proc. Natl. Acad. Sci. U.S.A. **99**, 12917 (2002).
[8] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[9] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[10] K.-I. Goh, E. Oh, H. Jeong, B. Kahng, and D. Kim, Proc. Natl. Acad. Sci. U.S.A. **99**, 12583 (2002).
[11] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[12] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
[13] R. M. May and A. L. Lloyd, Phys. Rev. E **64**, 066112 (2001).
[14] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[15] G. Paul, T. Tanizawa, S. Havlin, and H. E. Stanley, Eur. Phys. J. B **38**, 187 (2004).
[16] J. Balthrop, S. Forrest, M. J. E. Newman, and M. M. Williamson, Science **304**, 527 (2004).
[17] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).
[18] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Proc. Natl. Acad. Sci. U.S.A. **101**, 2658 (2004).
[19] M. E. J. Newman, Eur. Phys. J. B **38**, 321 (2004).
[20] M. E. J. Newman, Phys. Rev. E **69**, 066133 (2004).
[21] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature (London) **435**, 814 (2005).
[22] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).
[23] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, G. Li, and R. Chen, Nucleic Acids Res. **31**, 2443 (2003).
[24] E. Estrada and J. A. Rodríguez-Velázquez, Phys. Rev. E **71**, 056103 (2005).
[25] E. Estrada, and J. A. Rodríguez-Velázquez, Phys. Rev. E **72**, 046105 (2005).
[26] E. Estrada, Europhys. Lett. **73**, 649 (2006).
[27] S. Hoory, N. Linial, and A. Wigderson, Bull., New Ser., Am. Math. Soc. **43**, 439 (2006).
[28] E. Estrada, Eur. Phys. J. B **52**, 563 (2006).
[29] B. Aspvall and J. Gilbert, SIAM J. Algebraic Discrete Methods **5**, 526 (1984).
[30] A. J. Seary and W. D. Richards, Jr., in *Proceedings of the International Conference on Social Nets., London*, edited by M. G. Everett and K.Rennolls(Greenwich University Press, London, 1995), Vol. 1, p. 47.
[31] W. D. Richards, Jr., and A. J. Seary, Convergence Analysis of Communication Networks, http://www.sfu.ca/~richards/Pages/converge.pdf
[32] C. L. Alpert, A. B. Kahng, and S.-Z. Yao, Discrete Appl. Math. **90**, 3 (1999).
[33] R. Kannan, S. Vempala, and A. Vetta, J. ACM **51**, 497 (2004).
[34] A. Capocci, V. D. P. Servedio, G. Caldarelli, and P. Colaiori, Physica A **352**, 669 (2005).
[35] J. C. Rain, L. Selig, H. De Reuse, V. Battaglia, C. Reverdy, S. Simon, G. Lenzen, F. Petel, J. Wojcik, V. Schachter, Y. Chemana, A. Labigne, and P. Legrain, Nature (London) **409**, 211 (2001).
[36] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Science **303**, 1538 (2004).
[37] A. Rana Atilgan, P. Akan, and C. Baysal, Biophys. J. **86**, 85 (2004).
[38] V. Batagelj and A. Mrvar, http://vlado.fmf.uni-lj.si/pub/networks/data/
[39] COSIN database: http://www.cosin.org
[40] C. R. Myers, Phys. Rev. E **68**, 046116 (2003).
[41] G. F. Davis, M. Yoo, and W. E. Baker, Strat. Org. **1**, 301 (2003).
[42] J. J. Potterat, L. Philips-Plummer, S. Q. Muth, R. B. Rothenberg, D. E. Woodhouse, T. S. Maldonado-Long, H. P. Zimmerman, and J. B. Muth, Sex Transm. Infect. **78**, i159 (2002).
[43] A. Arenas, L. Danon, A. Díaz-Guilera, P. M. Gleiser, and R. Guimerá, Eur. Phys. J. B **38**, 373 (2004).
[44] W. R. Dillon and M. Goldstein, *Multivariate Analysis, Methods and Applications* (John Wiley & Sons, New York, 1984).
[45] E. Estrada, J. Theor. Biol. **244**, 296 (2007).
[46] A. A. Rashin, M. Iofin, and B. Honig, Biochemistry (Mosc.) **25**, 3619 (1986).
[47] M. A. Williams, J. M. Goodfellow, and J. M. Thornton, Protein Sci. **3**, 1224 (1994).
[48] S. J. Hubbard and P. Argos, Protein Sci. **3**, 2194 (1994).
[49] S. J. Hubbard, K. H. Gross, and P. Argos, Protein Eng. **7**, 613 (1994).
[50] D. Dreier, http://www.cs.ucr.edu/~ddreier
[51] G. Gkantsidis, A. Saberi, and M. Mihail, in *Proceedings of the ACM Sigmetrics* (ACM Press, New York, 2003).
[52] U. Alon, http://www.weizmann.ac.il/mcb/UriAlon/
[53] M. Jungsbluth, B. Burghardt, and A. K. Hartmann, e-print physics/0607150.