

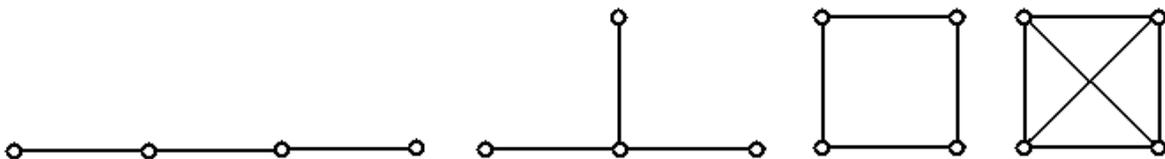
Spectral Theory of Networks: From Biomolecular to Ecological Systems.

Ernesto Estrada

Institute of Complex Systems at Strathclyde, Department of Physics and Department of Mathematics, University of Strathclyde, Glasgow G1 1XH, U.K. E-mail: estrada66@yahoo.com

1. Introduction

The best way for understanding how things work is by understanding their structures [1]. Complex networks are not an exception [2]. In order to understand why some networks are more robust than others, or why the propagation of a disease is faster in one network than in another is necessary to understand how these networks are organized [3-5]. A complex network is a simplified representation of a complex system in which the entities of the system are represented by the nodes in the network and the interrelations between entities are represented by means of the links joining pairs of nodes [3-5]. In analyzing the architecture of a complex network we are concerned only with the topological organization of these nodes and links. That is to say, we are not taking care of any geometric characteristic of the systems we are representing by these networks but only on how the parts are organized or distributed to form the whole system. Some of these topological characteristics of a network can be evident by simple visual inspection. This is particularly easy when the networks (graphs) are small. For instance, the first two graphs displayed below do not contain cycles, i.e., they are *trees*. The first of them is simply a *linear chain* and the second a *star*. The third and four graphs are cyclic. The third graph is the cycle of four nodes, C_4 , and the last is the graph having a connection between every pair of nodes, i.e., the complete graph K_4 [6]. All these graphs are connected, which means that we can travel from any node to another in the graph.



However, this visual analysis is not possible even for medium-sized networks. In addition, most of the real-world complex networks are very large and the questions we have to formulate to understand their structures and functioning are by far more complex [2-5]. Just to have a flavour on how complex this problem is we illustrate in the Fig. 1 the network of protein-protein interactions in human cells [7]. This network is far from being complete but it already contains

more than 3000 nodes [7]. In the Fig. 1 we illustrate some proteins in red that have been identified to be responsible of hereditary diseases in humans [7].

Insert Fig. 1 about here.

It is evident that we need more sophisticated tools than visual inspection for analyzing the structure of complex networks. One of these tools is the *spectral graph theory* [8]. The spectrum of a graph (technically explained in the next section) can be considered as the x-rays test for networks. In a similar way as we obtain information from x-ray spectroscopy about the internal structure of molecules we can obtain information about the internal organization of complex networks with the use of spectral graph theory. This chapter is dedicated to the analysis of graph spectra to extract information about the architectural organization of real-world complex networks.

2. Background on Graph Spectra

A graph $G = (V, E)$ is a set of nodes V , which are connected by means of the elements of the set of links E . Here we are dealing only with simple graphs [6]. That is an undirected graph without multiple links or self-loops. Thus, by graph we mean a simple graph. A node $v \in V$ is a terminal point of a link and represents an abstraction of an entity in a complex network such as a person, a city, a protein, an atom, etc. The links represent the relations between these entities.

A graph $G = (V, E)$ can be represented by different kinds of matrices [6]. The (ordinary) spectrum of a graph always refers to the spectrum of the adjacency matrix of the graph [9]. Thus, we will be concerned here only with this matrix. Excellent reviews about Laplacian spectrum of graphs can be found in the literature [10]. The adjacency matrix $\mathbf{A} = \mathbf{A}(G)$ of a graph $G = (V, E)$ is a symmetric matrix of order $n = |V|$, where $|\cdot|$ means the cardinality of the set, where $\mathbf{A}_{ij} = 1$ if there is a link between the nodes i and j and $\mathbf{A}_{ij} = 0$ otherwise.

The “spectrum” of a network is a listing of the *eigenvalues* of the adjacency matrix of such network. It is well known that every $n \times n$ real symmetric matrix \mathbf{A} has a spectrum of n orthonormal eigenvectors $\phi_1, \phi_2, \dots, \phi_n$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ [11]. The largest

eigenvalue of the graph λ_1 is known as the principal eigenvalue, the spectral radius or the Perron-Frobenius eigenvalue [11]. The eigenvector associated with this eigenvalue is also known as the principal eigenvector of the graph.

A walk of length l is any sequence of (not necessarily) different vertices $v_1, v_2, \dots, v_k, v_{k+1}$ such that for each $i = 1, 2, \dots, k$ there is an edge from v_i to v_{i+1} . A closed walk (CW) of length k is a walk in which $v_{k+1} = v_1$ [7]. The number of CWs of length μ_k is determined by the trace of the k th power of the adjacency matrix, $\mu_k = \text{Tr} \mathbf{A}^k$. This number is also known as spectral moment due to the following relationship with graph eigenvalues,

$$\mu_k = \sum_{j=1}^n (\lambda_j)^k. \quad (1)$$

The number of CWs of length k starting (and ending) at node p in the graph can also be expressed in terms of the graph eigenvalues and eigenvectors [12],

$$\mu_k(p) = \sum_{j=1}^n [\phi_j(p)]^2 (\lambda_j)^k. \quad (2)$$

In a similar way the number of walks of length k starting at node p and ending at node q are given by [12],

$$\mu_k(p, q) = \sum_{j=1}^n \phi_j(p) \phi_j(q) (\lambda_j)^k. \quad (3)$$

The spectrum of certain graphs is completely determined by the structure of the graph [9]. For instance, the complete graph, which is the graph in which every node is connected to every node, has spectrum $(n-1)^1, (-1)^{n-1}$. In the cycle graph, which is a graph on n nodes containing a single cycle through all nodes, the spectrum is given by $2 \cos(2\pi j/n)$ ($j = 0, \dots, n-1$). The path or linear chain is also determined by its spectrum, which is given $2 \cos(2\pi j/(n+1))$ ($j = 0, \dots, n$). The reader is referred to several books, such as [9, 12, 13] for a more thorough discussion and list of references to original papers.

3. Spectral measures of node centrality

A local characterization of networks is made numerically by using one of several measures known as “centrality” [14]. One of the most used centrality measures is the “degree centrality”, DC [15], which can be interpreted as a measure of immediate influence, as opposed to long-term effect in the network [14]. There are several other centrality measures that have been introduced and studied for real world networks, in particular for social networks. They account for the different node characteristics that permit them to be ranked in order of importance in the network. Betweenness centrality (BC) measures the number of times that a shortest path between nodes i and j travels through a node k whose centrality is being measured. The farness of a vertex is the sum of the lengths of the geodesics to every other vertex. The reciprocal of farness is closeness centrality (CC).

The first spectral measure of centrality was introduced by Bonacich in 1987 as the eigenvector centrality (EC) [16]. This centrality measure is not restricted to shortest paths [16], and it is defined as the principal or dominant eigenvector of the adjacency matrix A representing the connected subgraph or component of the network. It simulates a mechanism in which each node affects all of its neighbors simultaneously [17]. EC is better interpreted as a sort of extended degree centrality which is proportional to the sum of the centralities of the node’ neighbors. Consequently, a node has high value of EC either if it is connected to many other nodes or if it is connected to others that themselves have high EC [18].

Here we designate the number of walks of length L starting at node i by $N_L(i)$ and the total number of walks of this length existing in the network by $N_L(G)$. The probability that a walk selected at random in the network has started at node i is simply:

$$P_L(i) = \frac{N_L(i)}{N_L(G)} \quad (4)$$

It is known that for a non-bipartite connected network with nodes $1, 2, \dots, n$, for $L \rightarrow \infty$, the vector $[P_L(1) \ P_L(2) \ \dots \ P_L(n)]$ tends toward the eigenvector corresponding to the largest eigenvalue of the adjacency matrix of the network. Consequently, the elements of EC represent the probabilities of selecting at random a walk of length L starting at node i when $L \rightarrow \infty$: $EC(i) = P_L(i)$ [12].

Another spectral measure of node centrality was introduced recently by Estrada as the *subgraph centrality* of vertex i in the network, which is given by [19]:

$$SC(i) = \sum_{k=0}^{\infty} \frac{\mu_k(i)}{k!}. \quad (5)$$

where $\mu_k(i)$ are the number of closed walks of length k starting and ending at node i . The relation of this measure with the graph spectrum comes from the following results.

Let λ_1 be the principal eigenvalue of \mathbf{A} . For any nonnegative integer k and any $i \in \{1, \dots, n\}$, $\mu_k(i) \leq \lambda_1^k$, series (2), whose terms are nonnegative, converges.

$$\sum_{k=0}^{\infty} \frac{\mu_k(i)}{k!} \leq \sum_{k=0}^{\infty} \frac{\lambda_1^k}{k!} = e^{\lambda_1} \quad (6)$$

Thus, the subgraph centrality of any vertex i is bounded above by $SC(i) \leq e^{\lambda_1}$. The following result shows that the subgraph centrality can be obtained mathematically from the spectra of the adjacency matrix of the network.

Theorem [19]: Let $G = (V, E)$ be a simple graph of order n . Let $\phi_1, \phi_2, \dots, \phi_n$ be an orthonormal basis of R^n composed by eigenvectors of \mathbf{A} associated to the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$. Let $\phi_j(i)$ denote the i th component of ϕ_j . For all $i \in V$, the subgraph centrality may be expressed as follows:

$$SC(i) = \sum_{j=1}^n [\phi_j(i)]^2 e^{\lambda_j} \quad (7)$$

The sum of the subgraph centralities of all nodes in the network SC depends only on the eigenvalues of the adjacency matrix of the network [19]:

$$SC(G) = \sum_{i=1}^n SC(i) = \sum_{i=1}^n e^{\lambda_i} \quad (8)$$

SC is also known as the Estrada index of a graph and several mathematical results are available in the literature for this index [20-23]. Hereafter we will follow this designation and represent the subgraph centrality as $EE(G)$ or simply EE .

3.1. Subgraph centrality as a partition function

To start with let us now consider a network in which every pair of vertices is weighted by a parameter β . Let \mathbf{B} be the adjacency matrix of this weighted network. It is obvious that $\mathbf{B} = \beta\mathbf{A}$ and $\mu_r(\mathbf{B}) = \text{Tr } \mathbf{B}^r = \beta^r \text{Tr } \mathbf{A}^r = \beta^r \mu_r$. In this case, the subgraph centrality can be generalized as follows [24]:

$$EE(G, \beta) = \sum_{r=0}^{\infty} \frac{\beta^r \mu_r}{r!} = \sum_{j=1}^N e^{\beta \lambda_j} \quad (9)$$

Alternatively, we can write $EE(G, \beta)$ as follows:

$$EE(G, \beta) = \text{Tr} \sum_{r=0}^{\infty} \frac{\beta^r \mathbf{A}^r}{r!} = \text{Tr} e^{\beta \mathbf{A}}. \quad (10)$$

It is straightforward to realize that the subgraph centrality is generalized to the partition function of the complex network in the form [25]:

$$Z(G, \beta) \equiv EE(G, \beta) \equiv \text{Tr} e^{\beta \mathbf{A}}, \quad (11)$$

where the Hamiltonian is $\mathbf{H} = -\mathbf{A}$ and β is the inverse temperature, that is $\beta = 1/(k_B T)$. Note that β can be considered as the ‘‘strength’’ of the interaction between a pair of vertices, assuming

that every pair of vertices has the same interaction strength [25]. For instance, $\beta = 0$, which corresponds to the limit $T \rightarrow \infty$, corresponds to a graph with no links. This case is similar to a gas formed by monoatomic particles. On the other hand, very large values of β in the limit $T \rightarrow +0$ represents very large attractive interactions between pairs of bonded nodes in a similar manner to a solid. The ‘‘classical’’ subgraph centrality is the particular case when $\beta = 1$, i.e., the unweighted network.

Using this approach we can define the probability p_j that the system occupies a microstate j as follows [25]:

$$p_j = \frac{e^{\beta\lambda_j}}{\sum_j e^{\beta\lambda_j}} = \frac{e^{\beta\lambda_j}}{EE(G, \beta)}. \quad (12)$$

Based on Eq. (4) we can also define the information theoretic entropy for the network using the Shannon expression [25]:

$$S(G, \beta) = -k_B \sum_j [p_j (\beta\lambda_j - \ln EE)], \quad (13)$$

where we wrote $EE(G, \beta) = EE$. Then we can obtain the expressions for the total energy $H(G)$ and Helmholtz free energy $F(G)$ of the network [25]:

$$H(G, \beta) = -\sum_{j=1}^n \lambda_j p_j, \quad (14)$$

$$F(G, \beta) = -\beta^{-1} \ln EE. \quad (15)$$

These statistical mechanics functions of networks are bounded as follows [25]:

$$0 \leq S(G, \beta) \leq \beta \ln n, \quad (16)$$

$$-\beta(n-1) \leq H(G, \beta) \leq 0, \quad (17)$$

$$-\beta(n-1) \leq F(G, \beta) \leq -\beta \ln n, \quad (18)$$

where the lower bounds are obtained for the complete graph as $n \rightarrow \infty$ and the upper bounds are reached for the null graph with n nodes.

3.2. Application

As a first illustration of the possibilities of the spectral measures of centrality we selected one example published recently by Choi et al. [26] in which the eigenvector centrality was used in comparing world city networks. The authors ranked the most central cities in the world by considering the Internet backbone and air transport intercity linkages. When the authors considered only the number of direct links in the Internet backbone network, New York emerged as the most connected node, followed by London, Frankfurt, Tokyo and Paris. However, when the eigenvector centrality was considered the most central city was London, followed by New York, Paris, Frankfurt and Amsterdam. In the network of air passengers the ranking according to the degree centrality is dominated by London, followed by Frankfurt, Paris, New York and Amsterdam. The use of the eigenvector centrality ranks London as the most central one, but changes the order of the other cities, Paris becomes the second most central followed by New York, Amsterdam and Frankfurt. The differences arise from the fact that in the eigenvector centrality a city that is connected to central cities has its own centrality boosted. Then, it is not only important to have a large number of connections but to have these connections with highly central nodes in the network.

In order to illustrate the characteristics of the subgraph centrality we selected an example from the collaboration network of *Computational Geometry* authors [19]. We selected at random two authors with the same degree and different subgraph centrality (see Fig. 2): Timothy M. Y. Chan and S. L. Abrams, both having $DC = 10$, but having $SC = 8.09 \cdot 10^9$ and $SC = 974.47$, respectively. Despite both authors' having the same number of coauthors, Chan is connected to five of the hubs of this collaboration network: Agarwal (98), Snoeyink (91), Sharir (87), Tamassia (79) and Yap (76) (DC are given in parenthesis). However, Abrams is connected to authors having lower numbers of coworkers; e.g., Patrikalakis has 31 coauthors and the rest have only five to 16 collaborators. This simple difference means that Chan is separated from 623 other authors by a

distance of only two; i.e., simply connected triplets, while this number is significantly lower for Abrams, i.e., only 116. The risk that Chan is “infected” with an idea circulating among the authors in this field of research is much higher than the risk with Abrams. This difference is accounted for the subgraph centrality [19].

Insert Fig. 2 about here.

4. Global topological organization of complex networks

Our objective here is to give a characterization of the global organization of complex networks. The first step for analyzing the global architecture of a network is to determine whether the network is homogeneous or modular. In a homogeneous network what you see locally is what you get globally from a topological point of view. However, in a modular network the organization of certain modules or clusters would be different from one to another and to the global characteristics of the network [27-29].

Formally, we consider a network is homogeneous if it has good expansion (GE) properties. A network has GE if every subset S of nodes ($S \leq 50\%$ of the nodes) has a neighborhood that is larger than some “expansion factor” Ω 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 [30]. 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 : $\Gamma(S) = \bigcup_{v \in S} \Gamma(v)$ and the network has GE if $|\Gamma(S)| \geq \Omega |S| \quad \forall S \subseteq V$.

Consequently, in a homogeneous network we should expect that some local topological properties scale as a power-law of global topological properties. A power-law relationship between a two variables x and y of the network is known by the term scaling and refers to the relationship [31],

$$y = Ax^\eta, \tag{19}$$

where A and η are constant. The existence of a scaling law reveals that the phenomenon under study reproduces itself on different time and/or space scales. That is, it has self-similarity [31]. Then, if x and y are variables representing some topological features of the network at the local and the global scale, the existence of such scaling implies that the network is topological self-similar and *what we see locally is what we get globally*, which means that the network is homogeneous. In the following section we develop an approach to account for such scaling.

4.1. Spectral scaling method

Our first task here is to find a couple of appropriate topological variables for a network which characterize the local and global environment around a node. As for the local property we consider the subgraph centrality. As we already noted this spectral measure characterizes the local cliquishness around a node because it gives larger weights to the participation of a node in smaller subgraphs. It should be noted that $EE(i)$ counts all CWs in the network, which can be of even or odd length. CWs of even length might be trivial on moving back and forth in acyclic subgraphs, i.e., those that do not contain cycles, while odd CWs do not contain contributions from acyclic subgraphs. It is easy to show [32] that:

$$EE(i) = \sum_{j=1}^N [\phi_j(i)]^2 \cosh(\lambda_j) + \sum_{j=1}^N [\phi_j(i)]^2 \sinh(\lambda_j) = EE_{\text{even}}(i) + EE_{\text{odd}}(i) \quad (20)$$

which means that the term $EE_{\text{odd}}(i)$ only accounts for subgraphs containing at least one odd cycle. In this way $EE_{\text{odd}}(i)$ can be considered as a topological property of local organization in networks that characterise the odd-cyclic wiring of a typical neighbourhood. As a global topological characterization of the environment around a node we consider the eigenvector centrality. We have already shown that the eigenvector centrality EC represents the probability of selecting at random a walk of length L starting at node i when $L \rightarrow \infty$ [12]. Due to the infinite length of the walk we are considering, such a walk visit all nodes and links of the network obtaining a global picture of the global topological environment around the corresponding node.

Now, we can establish the relationship between the local and global spectral properties of a network. To start with, we consider a graph with GE properties. Then, it is known that for a network to have good expansion the gap between the first and second eigenvalues of the adjacency matrix ($\Delta\lambda = \lambda_2 - \lambda_1$) need to be sufficiently large. For instance, the following is a well-known result in the field of expander graphs [33-35],

Theorem (Alon-Milman): Let G be a regular graph with spectrum $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Then, the expansion factor is bounded as,

$$\frac{\lambda_1 - \lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_1(\lambda_1 - \lambda_2)}.$$

Then, let us write $EE_{odd}(i)$ as follows

$$EE_{odd}(i) = [EC(i)]^2 \sinh(\lambda_1) + \sum_{j=2} [\phi_j(i)]^2 \sinh(\lambda_j), \quad (21)$$

where $EC(i)$ is the eigenvector centrality (the principal eigenvector $\phi_1(i)$) and λ_1 is the principal eigenvalue of the network. Then, let us assume that $\lambda_1 \gg \lambda_2$ in such a way that we can consider that $[EC(i)]^2 \sinh(\lambda_1) \gg \sum_{j=2} [\phi_j(i)]^2 \sinh(\lambda_j)$. Consequently, we can write the odd-subgraph

centrality as,

$$EE_{odd}(i) \approx [EC(i)]^2 \sinh(\lambda_1), \quad (22)$$

and the principal eigenvector of the network is directly related to the subgraph centrality in GENs according to the following spectral scaling relationship [36, 37]:

$$EC(i) \propto A [EE_{odd}(i)]^\eta. \quad (23)$$

which corresponds to the power-law relationship between the $EC(i)$ and $EE_{odd}(i)$ for GENs, which is similar to the one given by (19) where x and y are variables representing some topological features of the network at the local and the global scale. Here, $A \approx [\sinh(\lambda_1)]^{-0.5}$ and $\eta \approx 0.5$. This expression can be written in a log-log scale as [36, 37]:

$$\log[EC(i)] = \log A + \eta \log[EE_{odd}(i)]. \quad (24)$$

Consequently, a log-log plot of $EC(i)$ vs. $EE_{odd}(i)$ in a homogeneous network has to show a linear fit with slope $\eta \approx 0.5$ and intercept $\log A$ for GENs.

4.2. Universal topological classes of networks

There are several classification schemes grouping networks according to their structures. For instance, complex networks can be classified according to the existence or not of the “small-world” property [38, 39] or according to their degree distribution. The last classification permits to classify networks as “scale-free” [40] if their node degree distribution decays as a 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 [41]. Even scale-free networks have been classified into two different subclasses according to their exponent in the power-law distribution of the betweenness centrality [42].

Each of these classification schemes reproduces different characteristics of complex networks. “Small-worldness” [38] and “scale-freeness” [40] 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 [38]. The second reproduces the presence of a few highly connected hubs that maintain glued the vast majority of poorly connected nodes in certain networks [40]. Both properties are of great relevance in analyzing other important properties of complex networks, such as disease propagation [43-45] or robustness against targeted or random attacks [46-48]. However, there are important organizational principles of complex networks which escape the analysis of these global network characteristics.

The theoretical approach we presented in the previous section permits the classification of complex networks into two groups: homogeneous (GEN) and non-homogeneous networks. Here

we are interested in identifying the topological differences existing among the non-homogeneous networks in such a way that permit us to classify them into some universal classes.

Let us consider the ideal case in which a network displays perfect spectral scaling, such that we can calculate the eigenvector centrality by using the following expression,

$$\log EC^{Ideal}(i) = 0.5 \log EE_{odd}(i) - 0.5 \log[\sinh(\lambda_1)] \quad (25)$$

Now, let us consider the deviations from the ideal behavior represented by Eq. (22) in non-homogeneous networks. We can account for these deviations from *ideality* by measuring the departure of the points from the perfect straight line respect to $\log EC^{Ideal}(i)$:

$$\Delta \log EC(i) = \log \frac{EC(i)}{EC^{Ideal}(i)} = \log \left\{ \frac{[EC(i)]^2 \sinh(\lambda_1)}{EE_{odd}(i)} \right\}^{0.5} \quad (26)$$

Then, according to the values of $\Delta \log EC(i)$ there are four different classes of complex networks. These classes are [49]:

□ **Class I:** networks displaying perfect spectral scaling:

$$\Delta \log EC(i) \cong 0, \forall i \in V \Rightarrow [EC(i)]^2 \sinh(\lambda_1) \cong EE_{odd}(i) \quad (27)$$

□ **Class II:** networks displaying spectral scaling with negative deviations:

$$\Delta \log EC(i) \leq 0 \Rightarrow [EC(i)]^2 \sinh(\lambda_1) \leq EE_{odd}(i), i \in V \quad (28)$$

□ **Class III:** networks displaying spectral scaling with positive deviations:

$$\Delta \log EC(i) \geq 0 \Rightarrow [EC(i)]^2 \sinh(\lambda_1) \geq EE_{odd}(i), i \in V \quad (29)$$

□ **Class IV:** networks displaying spectral scaling with mixed deviations:

$$\Delta \log EC(p) \leq 0, p \in V \text{ and } \Delta \log EC(q) > 0, q \in V \quad (30)$$

We previously showed that 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. Class II corresponds to networks in which there are two or more communities of highly interconnected nodes, which display low inter-module connectivity. This kind of networks looks like networks containing holes in their structures. 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, class IV networks display a combination of highly connected groups (quasi-cliques) and some groups of nodes partitioned into disjoint subsets (quasi-bipartite), without a predominance of any of both structures. In Fig. 3 we illustrate the main structural properties of non-homogeneous networks and their respective spectral scaling plots.

Insert Fig. 3 about here.

In order to quantify the degree of deviation of the nodes from the ideal spectral scaling we account for the mean square error of all points with positive and negative deviations in the spectral scaling, respectively [49]:

$$\xi^+ = \sqrt{\frac{1}{N_+} \sum_+ \left(\log \frac{EC(i)}{EC^{Ideal}(i)} \right)^2} \quad \text{and} \quad \xi^- = \sqrt{\frac{1}{N_-} \sum_- \left(\log \frac{EC(i)}{EC^{Ideal}(i)} \right)^2},$$

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

4.3. Applications

We have studied 61 real-world complex networks accounting for ecological, biological, protein secondary structures, informational, technological and social systems [49]. 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. 4 we can see the main factors (roots) which perfectly separate the networks studied into the four different structural classes.

Insert Figure 4 about here

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. 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%) [49].

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 [49]. Similar results are obtained when generating random networks with the same degree sequence as real-world networks. Our results confirm previous findings about the necessity of investigating new growing mechanisms for generating networks to model real-world systems [50].

5. Communicability in complex networks

The *communicability* between a pair of nodes in a network is usually considered as taking place through the shortest path connecting both nodes. However, it is known that communication between a pair of nodes in a network does not always take place through the shortest paths but it can follow other non-optimal walks [51-53]. Then, we can consider a communicability measure that accounts for a weighted sum of all walks connecting two nodes in the network. We can design our measure in such a way that the shortest path connecting these two nodes always receives the largest weight. Then, if $P_{pq}^{(s)}$ is the number of shortest paths between the nodes p and q having length s and $W_{pq}^{(k)}$ is the number of walks connecting p and q of length $k > s$, we propose to consider the quantity [54]

$$G_{pq} = \frac{1}{s!} P_{pq} + \sum_{k>s} \frac{1}{k!} W_{pq}^{(k)}. \quad (31)$$

In fact, (31) can be written as the sum of the p, q entry of the different powers of the adjacency matrix,

$$G_{pq} = \sum_{k=0}^{\infty} \frac{(\mathbf{A}^k)_{pq}}{k!}. \quad (32)$$

which converges to [54]

$$G_{pq} = (e^{\mathbf{A}})_{pq} = \sum_{j=1}^n \phi_j(p) \phi_j(q) e^{\lambda_j}. \quad (33)$$

We call G_{pq} the *communicability* between the nodes p and q in the network. The communicability should be minimum between the end nodes of a chain, where it vanishes as the length of the chain is increased. On the other hand, the communicability between an arbitrary pair of nodes in the complete graph diverges as the size of the graph is increased because the oscillation is greatly amplified because of the infinitely many walks between the nodes. Thus, the communicability between a pair of nodes in a network is bounded between zero and infinity,

which are obtained for the two end nodes of an infinite linear chain and for a pair of nodes in an infinite complete graph. For the linear chain P_n the value of G_{pq} is equal to [54]

$$G_{pq} = \frac{1}{n+1} \sum_j \left(\cos \frac{j\pi(p-q)}{n+1} - \cos \frac{j\pi(p+q)}{n+1} \right) e^{2 \cos \left(\frac{j\pi}{n+1} \right)}. \quad (34)$$

Let P_∞ be a chain of infinite length. It is straightforward to realize by simple substitution in (34) that $G_{1,\infty} = 0$ for the end nodes $p=1$ and $q=\infty$. For the complete graph we have that [54]

$$G_{pq} = \frac{e^{n-1}}{n} + e^{-1} \sum_{j=2}^n \varphi_j(p) \varphi_j(q) = \frac{e^{n-1}}{n} - \frac{1}{ne} = \frac{1}{ne} (e^n - 1), \quad (35)$$

and it is easy to see that $G_{pq} \rightarrow \infty$ as $n \rightarrow \infty$ for K_n .

A physical interpretation of the communicability can be done by considering a continuous-time quantum walk on the network. Take a quantum-mechanical wave function $|\psi(t)\rangle$ at time t . It obeys the Schrödinger equation [55]

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = -\mathbf{A} |\psi(t)\rangle, \quad (36)$$

where we use the adjacency matrix as the negative Hamiltonian.

Assuming from now on that $\hbar=1$ we can write down the solution of the time-dependent Schrödinger equation (33) in the form $|\psi(t)\rangle = e^{i\mathbf{A}t} |\psi(0)\rangle$. The final state $e^{i\mathbf{A}t} |q\rangle$ is a state of the graph that results after time t from the initial state $|q\rangle$. The “particle” that resided on the node q at time $t=0$ diffuses for the time t because of the quantum dynamics. Then, we can obtain the amplitude that the “particle” ends up at the node p of the network by computing the product $\langle p | e^{i\mathbf{A}t} | q \rangle$. By continuation from the real time t to the imaginary time, we have the thermal Green’s function defined as $G_{pq} = \langle p | e^{\mathbf{A}} | q \rangle$, which is the communicability between nodes p and q in the network as defined in this work [54]. Consequently, the communicability between nodes p and q in the network represents the probability that a particle starting from the node p ends up

at the node q after wandering on the complex network due to the thermal fluctuation. By regarding the thermal fluctuation as some form of random noise, we can identify the particle as an information carrier in a society or a needle in a drug-user network.

5. 1. Communicability and network communities

Many complex networks in the real-world are not homogeneous as we have already seen previously in this Chapter. Instead, the nodes in most networks appear to group in subgraphs in which the density of internal connections is larger than the connections with the rest of the nodes in the network. This notion was first introduced by Girvan and Newman [56] and it is known as the *community structure* of complex networks [57-61]. In the language of communicability we are using in this section we can say that a community is a group of nodes having larger communicability among them than with the rest of the nodes in the graph. Later on we will give a more formal definition of community.

In order to make further analysis, we now use the spectral decomposition of the Green's function [62]. Imagine that the network has a spring on each link. Each eigenvector indicates a mode of oscillation of the entire network and its eigenvalue represents the weight of the mode. It is known that the eigenvector of the largest eigenvalue λ_1 has elements of the same sign. This means that the most important mode is the oscillation where all nodes move in the same direction at one time.

The second largest eigenvector ϕ_2 has both positive and negative elements. Suppose that a network has two clusters connected through a bottleneck but each cluster is closely connected within. The second eigenvector represents the mode of oscillation where the nodes of one cluster move coherently in one direction and the nodes of the other cluster move coherently in the opposite direction. Then the sign of the product $\phi_2(p)\phi_2(q)$ tells us whether the nodes p and q are in the same cluster or not.

The same analysis can be applied to the rest of the eigenvalues of the network. The third eigenvector ϕ_3 , which is orthonormal to the first two eigenvectors, have a different pattern of signs, dividing the network into three different blocks after appropriate arrangement of the nodes. In general, the second eigenvector divides the graph into biantis, the third divides it into triants, the fourth into quadrants, and so forth, but these clusters are not necessarily independent of each other.

According to this pattern of signs we have the following decomposition of the thermal Green's function [54]:

$$G_{pq} = \left[\phi_1(p)\phi_1(q)e^{\lambda_1} \right] + \left[\sum_{j=2}^{++} \phi_j^+(p)\phi_j^+(q)e^{\lambda_j} + \sum_{j=2}^{--} \phi_j^-(p)\phi_j^-(q)e^{\lambda_j} \right] + \left[\sum_{j=2}^{+-} \phi_j^+(p)\phi_j^-(q)e^{\lambda_j} + \sum_{j=2}^{-+} \phi_j^-(p)\phi_j^+(q)e^{\lambda_j} \right]. \quad (37)$$

where ϕ_j^+ and ϕ_j^- refer to the eigenvector components with positive and negative signs, respectively. According to the partitions made by the pattern of signs of the eigenvectors in a graph, two nodes have the same sign in an eigenvector if they can be considered as being in the same partition of the network, while those pairs having different signs correspond to nodes in different partitions. Thus, the first bracket in (34) represents the background mode of translational movement. The second bracket represents the *intracluster communicability* between nodes in the network and the third bracket represents the *intercluster communicability* between nodes [54].

The above consideration motivates us to define a quantity ΔG_{pq} by subtracting the contribution of the largest eigenvalue λ_1 from Eq. (34) [54]:

$$\Delta G_{pq} = \sum_{j=2}^{\text{intracluster}} \phi_j(p)\phi_j(q)e^{\lambda_j} + \sum_{j=2}^{\text{intercluster}} \phi_j(p)\phi_j(q)e^{\lambda_j}. \quad (38)$$

By focusing on the sign of ΔG_{pq} , we can unambiguously define a community for a group of nodes.

If ΔG_{pq} for a pair of nodes p and q have a positive sign, they are in the same community. If ΔG_{pq}

for the two nodes have a negative sign they are in different clusters [54].

Definition: A community in a network is a groups of nodes $U \subseteq V$ for which the intracluster communicability is larger than the intercluster communicability, i.e., $\Delta G_{pq} > 0, \forall (p, q) \in U$.

5.2. Detection of communities: The communicability graph

To start with we represent the values of ΔG_{pq} in the form of a matrix $\mathbf{\Lambda}$. $\mathbf{\Lambda}$ is a matrix whose nondiagonal entries are given by the values of ΔG_{pq} and zeroes in the main diagonal. Now, let us introduce a Heaviside step function:

$$\Theta(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}. \quad (39)$$

Let $\Theta(\mathbf{\Lambda})$ be the result of applying the Heaviside step function in an elementwise way to the matrix $\mathbf{\Lambda}$. Then, in the resulting matrix $\Theta(\mathbf{\Lambda})$ a pair of nodes p and q is connected if, and only if, they have $\Delta G_{pq} > 0$. Then let us define the following graph [63].

Definition: The *communicability graph* $\Theta(G)$ is the graph having adjacency matrix $\Theta(\mathbf{\Lambda})$.

In such a graph two nodes are connected if they have $\Delta G_{pq} > 0$. That is to say, the nodes forming a community in the original graph are connected in the communicability graph. Now, suppose that there is a link between the nodes p and q and there are also links between them and a third node r . This means that $\Delta G_{pq} > 0$, $\Delta G_{pr} > 0$ and $\Delta G_{qr} > 0$. Consequently, the three nodes form a positive subgraph C . As we want to detect the largest subset of nodes connected to this triple we have to search for the nodes s for which $\Delta G_{is} > 0 \forall i \in C$. Using the *communicability graph*, this search is reduced to finding the cliques in a simple graph, $\Theta(\mathbf{\Lambda})$. These cliques correspond to the communities of the network. A clique is a maximum complete subgraph in the graph. That is a maximum subgraph in which every pair of nodes is connected.

Finding the cliques in a graph is a classical problem in combinatorial optimization, which has found applications in diverse areas [64]. Here we use a well-known algorithm due to Bron and Kerbosch [65], which is a depth-first search for generating all cliques in a graph. This algorithm

consumes a time per clique which is almost independent of the graph size for random graphs and for the Moon-Moser graphs of n vertices the total time is proportional to $(3.14)^{n/3}$. The Moon-Moser graphs have the largest number of maximal cliques possible among all n -vertex graphs regardless of the number of edges in the graph [66].

5. 3. Application

As an example of a real-world network, we consider a friendship network known as the Zachary karate club, which has 34 members (nodes) with some friendship relations (links). The members of the club, after some entanglement, were eventually fractioned into two groups, one formed by the followers of the instructor and the other formed by the followers of the administrator [67]. This network has been analyzed in practically every paper considering the problem of community identification in complex networks. In Fig. 5a we illustrate the Zachary network in which the nodes are divided into the two classes observed by Zachary on the basis of the friendship relationships among the members of the club.

In the Fig. 5b we illustrate the communicability graph $\Theta(G)$ of the Zachary network. As can be seen $\Theta(G)$ correctly divides the network into two groups. There is very high internal communicability among the members of the respective groups but there is almost no communicability between the groups. In fact, the node 3 is correctly included in the group of the instructor (node 1).

Insert Fig. 5 about here.

The analysis of the cliques in the communicability graph reveals a more detailed view of the community structure of this network. Accordingly, there are five different cliques representing five overlapping communities in the network. These communities are given below, where the numbers correspond to the labels of the nodes in Fig. 5a:

- 1 : {10,15,16,19,21,23,24,26,27,28,29,30,31,32,33,34}
- 2 : {9,10,15,16,19,21,23,24,27,28,29,30,31,32,33,34}
- 3 : {10,15,16,19,21,23,24,25,26,27,28,29,30,32,33,34}
- 4 : {1,2,3,4,5,6,7,8,11,12,13,14,17,18,20,22}
- 5 : {3,10}

As can be seen, the first three communities, which correspond to the group of the administrator (node 34), are formed by 16 members each, and display an overlap of about 94% (see Fig. 6). The fourth community corresponds to the one of the instructor (node 1) and also has 16 members. The last community is formed by the nodes 3 and 10 only. This community displays overlaps with the communities of the administrator as well as with the one of the instructor. In fact, node 10 appears in communities 1 to 4, and node 3 appears in communities 4 and 5.

Insert Fig. 6 about here.

6. Network Bipartivity

There are numerous natural systems that can be modelled by making a partition of the nodes into two disjoint sets [68, 69]. For instance, in a network representing heterosexual relationships one set of nodes corresponds to female and the other to male partners. In some trade networks one set of nodes can represent buyers and the other sellers, and so forth. These networks are named bipartite networks or graphs and are formally defined below [6].

Definition: A network (graph) $G = (V, E)$ is called *bipartite* if its vertex set V can be partitioned into two subsets V_1 and V_2 such that all edges have one endpoint in V_1 and the other in V_2 .

Now, let us consider the case in which some connections between the nodes in the same set of a formerly bipartite network are allowed. Strictly speaking these networks are not bipartite but we can consider them loosely as *almost-bipartite* networks. For instance, if we consider a sexual relationships network in which not only heterosexual but also some homosexual relations are present the network is not bipartite but it could be almost-bipartite if the number of homosexual relations is low compared to the number of heterosexual ones. It is known that the transmission

rates for homosexual and heterosexual contacts differ [69]. Consequently, the transmission of this disease will depend on how bipartite the corresponding network is. In other words, having an idea of the bipartivity of sexual networks we will have an idea on the rate of spreading of a sexually transmitted disease.

The following is a well-known result due to König that permits us to characterize bipartite graphs [70].

Theorem (König): A graph is bipartite if and only if all its cycles are even.

We will make use of this result in order to characterize the bipartivity of a network. To start with we consider the subgraph centrality of the whole graph defined by (5). We can express this index as the sum of two contributions, one coming from odd and the other from even CWs [32]:

$$EE(G) = \sum_{j=1}^n [\cosh(\lambda_j) + \sinh(\lambda_j)] = EE_{even} + EE_{odd}. \quad (40)$$

If $G(V, E)$ is bipartite then according to the theorem of König [70]: $EE_{odd} = \sum_{j=1}^n \sinh(\lambda_j) = 0$

because there are no odd CWs in the network [32]. Therefore:

$$EE(G) = EE_{even} = \sum_{j=1}^n \cosh(\lambda_j). \quad (41)$$

Consequently, the proportion of even CWs to the total number of CWs is a measure of the network bipartivity [32]:

$$\beta(G) = \frac{EE_{even}}{EE(G)} = \frac{EE_{even}}{EE_{even} + EE_{odd}} = \frac{\sum_{j=1}^n \cosh(\lambda_j)}{\sum_{j=1}^n e^{\lambda_j}}. \quad (42)$$

It is evident that $\beta(G) \leq 1$ and $\beta(G) = 1$ if, and only if, G is bipartite, i.e., $EE_{odd} = 0$. Furthermore, as $0 \leq EE_{odd}$ and $\sinh(\lambda_j) \leq \cosh(\lambda_j)$, $\forall \lambda_j$, then $\frac{1}{2} < \beta(G)$ and $\frac{1}{2} < \beta(G) \leq 1$. The lower bound is reached for the least possible bipartite graph with n nodes, which is the complete

graph K_n . As the eigenvalues of K_n are $n-1$ and -1 (with multiplicity $n-1$), then $\beta(G) \rightarrow \frac{1}{2}$ when $n \rightarrow \infty$ in K_n .

Then, $\beta(G)$ represents a quantitative characterization of the bipartivity of a complex network. Now, we have a quantitative measure which permits us to discern between quasi-bipartite networks as well as to differentiate them from bipartite and not bipartite graphs. However, still an open question remains: Can we identify the bipartite subgraphs existing in a network?

6.1. Detecting bipartite substructures in complex networks

It is known that the eigenvectors corresponding to positive eigenvalues give a partition of the network into clusters of tightly connected nodes [71, 72]. In contrast, 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 [71, 72]. Then we can make use of the communicability function to identify the bipartite structures in complex networks. In general we can say that a positive (negative) value of β in the communicability function (30) increases the contribution of the positive (negative) eigenvalues to the communicability function. Then if we write the communicability function as [73]

$$G_{pq}(\beta) = \sum_{\lambda_j < 0} \phi_j(p)\phi_j(q)e^{\beta\lambda_j} + \sum_{\lambda_j = 0} \phi_j(p)\phi_j(q)e^{\beta\lambda_j} + \sum_{\lambda_j > 0} \phi_j(p)\phi_j(q)e^{\beta\lambda_j}, \quad (43)$$

we have that

$$G_{pq}(\beta > 0) \approx \sum_{\lambda_j > 0} \phi_j(p)\phi_j(q)e^{\beta\lambda_j}, \quad (44)$$

$$G_{pq}(\beta < 0) \approx \sum_{\lambda_j < 0} \phi_j(p)\phi_j(q)e^{-|\beta|\lambda_j}. \quad (45)$$

In other words, $G_{pq}(\beta > 0)$ determines a partition of the network into clusters of tightly connected nodes, which corresponds to the network communities. On the other hand, for $G_{pq}(\beta < 0)$ the network is partitioned in such a way that the nodes are close to other nodes which

have similar patterns of connections with other sets of nodes, i.e., nodes to which they are structurally equivalent. In the first case, we say that the nodes corresponding to larger components tend to form *quasi-cliques*. That is, clusters in which every two nodes tend to interact with each other. In the second case, the nodes tend to form *quasi-bipartites*, i.e., nodes are partitioned into *almost* disjoint subsets with high connectivity between sets but low internal connectivity.

Let us consider a bipartite graph and let p and q be nodes which are in two different disjoint sets of the graph. Then, there are no walks of even length between p and q in the graph and [73]

$$G_{pq}(\beta = -1) = [-\sinh(A)]_{pq} < 0. \quad (46)$$

However, if p and q are nodes in the same disjoint set, then there is no walk of odd length connecting them due to the lack of odd cycles in the bipartite graph, which makes

$$G_{pq}(\beta = -1) = [\cosh(A)]_{pq} > 0. \quad (47)$$

The above argument shows that, in general, the sign of the communicability at a negative temperature, $G_{pq}(\beta = -1) = (e^{-A})_{pq}$, gives an indication as to how the nodes can be separated into disjoint sets.

Our strategy for detecting quasi-bipartite clusters in complex networks is as follows. First we start by calculating $\exp(-A)$, whose (p, q) -entry gives the communicability between the nodes p and q in the network. Then we introduce the following definition [73]

Definition 2. The *node-repulsion graph* is a graph whose adjacency matrix is given by $\Theta[\exp(-A)]$, which results from the elementwise application of the function $\Theta(x)$ to the matrix $\exp(-A)$. A pair of nodes p and q in the node-repulsion graph $\Theta[\exp(-A)]$ is connected if, and only if, they have $G_{pq} > 0$. The Heaviside function $\Theta(x)$ was already introduced in (39).

Using the *node-repulsion graph*, the search for quasi-bipartite subgraphs in the complex network is reduced to finding the cliques in a simple graph, $\Theta[\exp(-A)]$. These cliques correspond to the quasi-bipartite clusters of the network [73].

6.2. Application

Here we study the food web of Canton Creek, which consists primarily of invertebrates and algae in a tributary, surrounded by pasture, of the Taieri River in the South Island of New Zealand [74]. This network consists of 108 nodes (species) and 707 links (trophic relations). Using our current approach, we find that this network can be divided into two almost-bipartite clusters, one having 66 nodes and the other 42. Only 20 links connect nodes in the same clusters, 13 of them connect nodes in the set containing 66 nodes and the other 7 connect nodes in the set of 42 nodes. Thus 97.2% of links are connections between the two almost-bipartite clusters and only 2.8% links are intracluster connections [73]. In Fig. 7, we illustrate the network and its quasi-bipartite clusters as found in the current work. The value of the bipartivity measure for this network $\beta = 0.775$ indicates that the network in general is not bipartite but that an important presence of bipartite and quasi-bipartite structures are present in the graph, which is corroborated by our algorithm for finding such structures [73].

Insert Fig. 7 about here.

7. Conclusion

The discovery of X-rays more than a century ago has increased our knowledge in many fields, such as the structure of matter, cosmology, security in technology and X-rays diagnostics, among others. The existence of a tool, like X-rays and other spectroscopic techniques, permits us to understand the internal structure of the systems under study from molecules and materials to the human body. In a similar way spectral graph theory is the X-ray machine for studying complex networks. As we have shown here the use of graph spectral techniques permits us to analyze the local and global structure of complex networks.

Using graph spectral theory it is possible to “see” how central a node is based on its weighted participation in all substructures present in the graph. The same techniques permit us to analyze whether a network is homogeneous or modular. In the last case it permits to classify their structures according to certain universal structural classes, no matter if it is representing a cell or a society. In addition, the spectral techniques explained in this Chapter permit us to identify the communities existing in a complex network, as well as the bipartivity structure of certain substructures present in such systems. There many other characteristics of complex networks that can be investigated by using the spectra of graphs. Some of them have been already described by the scientists working in this field, others are still waiting for the development of the appropriate tools. I hope this chapter contributes to inspire the development of new spectral measures for characterizing the structure and functioning of complex networks.

Acknowledgements

This chapter was written between the University of Santiago de Compostela, Spain, the University of Strathclyde, UK and the University of Tokyo, Japan. Then, the author thanks the program “Ramón y Cajal”, Spain, the support by the Royal Society of Edinburgh and the Edinburgh Mathematical Society during a visit to the Department of Mathematics, Strathclyde University (March 2008), to the IIS, University of Tokyo for a fellowship as Research Visitor during April-June, 2008 and both institutions for warm hospitality.

References

- 1 Bar-Yam, Y., *Making things work: Solving complex problems in a complex world*. NECSI Knowledge Press, Cambridge, MA, USA, **2004**
- 2 da F. Costa, L., Rodríguez, F. A., Travieso, G., Villa Boas, P. R., *Adv. Phys.* 56 (2007) p. 167
- 3 Albert, R., Barabási, A.-L., *Rev. Mod. Phys.* 74 (2002) p. 47
- 4 Newman, M. E. J., *SIAM Rev.* 45 (2003) p. 167
- 5 Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U., *Phys. Rep.* 424 (2006) p. 175
- 6 Harary, F. *Graph Theory*. Addison-Wesley, Reading, MA, **1994**
- 7 Rual, J.-F. et al., *Nature* 437 (2005) p. 1173
- 8 Biggs, N. L., *Algebraic Graph Theory*. Cambridge University Press, Cambridge, **1993**
- 9 Cvetković, D., Doob, M., Sach, H., *Spectra of Graphs*. Academic Press, NY, **1980**
- 10 Chung, F. R. K., *Spectral Graph Theory*. CBMS 92, AMS, Providence, **1997**
- 11 Horn, R., Johnson, C. R., *Matrix Analysis*. Cambridge University Press, Cambridge, **1985**
- 12 Cvetković, D., Rowlinson, P., Simić, S., *Eigenspaces of Graphs*. Cambridge University Press, Cambridge, **1997**
- 13 Cvetković, D., Doob, M., Gutman, I., Torgašev, A. *Recent Results in the Theory of Graph Spectra*. North-Holland, Amsterdam, **1988**
- 14 Freeman, L. C., *Social Networks* 1 (1979) p. 215
- 15 Albert, R., Jeong, H., Barabási, A.-L., *Nature* 401 (1999) p. 130
- 16 Bonacich, P., *J. Math. Sociol.* 2 (1972) p. 113.
- 17 Bonacich, P., *Am. J. Sociol.* 92 (1987) p. 1170
- 18 Newman, M. E. J., *Phys. Rev. E* 70 (2004) p. 056131
- 19 Estrada, E., Rodríguez-Velázquez, J. A., *Phys. Rev. E* 71 (2005) p. 056103

- 20 de la Peña, J. A., Gutman, I. Rada, J., *Lin. Algebra Appl.* 427 (2007) p. 70
- 21 Gutman, I. Graovac, A., *Chem. Phys. Lett.* 436 (2007) p. 294
- 22 Ginosar, Y., Gutman, I., Mansour, T. Schork, M., *Chem. Phys. Lett.* 454 (2008) p. 145
- 23 Carbó-Dorca, R. *J. Math. Chem.* (2008) (in press) DOI 10.1007/s10910-007-9314-y
- 24 Rodríguez-Velázquez, J. A., Estrada, E., Gutiérrez, A., *Lin. Multil. Algebra* 55 (2007) 293
- 25 Estrada E., Hatano, N., *Chem. Phys. Lett.* 439 (2007) 247
- 26 Choi, J. H., Barnett, G. A., Chon, B.-S., *Global Networks* 6 (2006) p. 81
- 27 Variano, E. A., McCoy, J. H., *Phys. Rev. Lett.* 92 (2004) p. 188701
- 28 Newman, M. J. E., *Proc. Natl. Ac. Sci USA* 103 (2006) p. 8577
- 29 Wagner, G. P., Pavlicev, M., Cheverud, J. M., *Nature Rev. Genet.* 8 (2007) p. 921
- 30 Hoory, S., Linial, N., Wigderson, A., *Bull. Am. Math. Soc.* 43 (2006) 439
- 31 Barenblatt, G. I., *Scaling.* Cambridge University Press, Cambridge, 2003
- 32 Estrada, E., Rodríguez-Velázquez, J. A., *Phys. Rev. E* 72 (2005) 046105
- 33 Dodziuk, J., *Trans. Amer. Math. Soc.*, 284 (1984) p. 787
- 34 Alon, N., Milman, V. D., *J. Combin. Theory Ser. B*, 38 (1985) p. 73
- 35 Alon, N., *Combinatorica*, 6 (1986) p. 83
- 36 Estrada, E., *Europhys. Lett.* 73 (2006) p. 649
- 37 Estrada, E., *Eur. Phys. J. B* 52 (2006) p. 563
- 38 Watts, D. J., Strogatz, S. H., *Nature* 393 (1998) p. 440
- 39 Dunne, J. A., Williams, R. J., Martinez, N. D., *Proc. Natl. Acad. Sci. USA* 99 (2002) p. 12917
- 40 Barabási, A.-L., Albert, R., *Science* 286 (1999) p. 509
- 41 Amaral, L. A. N., Scala, A., Barthélémy, M., Stanley, H. E., *Proc. Natl. Acad. Sci. USA* 97 (2000) p. 11149
- 42 Goh, K.-I., Oh, E., Jeong, H., Kahng, B., Kim, D., *Proc. Natl. Acad. Sci. USA* 99 (2002) p. 12583

- 43 Pastor-Satorras R., Vespignani, A., *Phys. Rev. Lett.* 86 (2001) p. 3200
- 44 Newman, M. E. J., *Phys. Rev. E* 66 (2002) p. 016128.
- 45 May, R. M., Lloyd, A. L., *Phys. Rev. E* 64 (2001) p. 066112
- 46 Albert, R., Jeong, H., Barabási, A.-L., *Nature* 406 (2000) p. 378
- 47 Paul, G., Tanizawa, T., Havlin, S., Stanley, H. E., *Eur. Phys. J. B* 38 (2004) p. 187
- 48 Balthrop, J., Forrest, S., Newman, M. J. E., Williamson, M. M., *Science* 304 (2004) p. 527
- 49 Estrada, E., *Phys. Rev. E* 75 (2007) p. 016103
- 50 Jungsbluth, M., Burghardt, B., Hartmann, A. K., *Physica A* 381 (2007) p. 444
- 51 Borgatti, S. P., *Social Networks* 27 (2005) p. 55
- 52 Hromkovic, J., Klasing, R., Pelc, A., Ruzicka, P., Unger, W., *Dissemination of Information in Communication Networks: Broadcasting, Gossiping, Leader Election, and Fault Tolerance*. Springer, Berlin, 2005
- 53 Shi, T. J., Mohan, G., *Comput. Comm.* 29 (2006) p. 1284
- 54 Estrada, E., Hatano, N., *Phys. Rev. E* 77 (2008) p. 036111
- 55 Morse, P. M., Feshbach, H., *Methods of Theoretical Physics*. McGraw-Hill, New York, 1953
- 56 Girvan, M., Newman, M. E. J., *Proc. Natl. Acad. Sci. USA* 99 (2002) p. 7821
- 57 Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D., *Proc. Natl. Acad. Sci. USA* 101 (2004) p. 2658
- 58 Newman, M. E. J., *Eur. Phys. J. B* 38 (2004) p. 321
- 59 Newman, M. E. J., *Phys. Rev. E* 69 (2004) p. 066133
- 60 Palla, G., Derényi, I., Farkas, I., Vicsek, T., *Nature* 435 (2005) p. 814
- 61 Newman, M. E. J., *Phys. Rev. E* 74 (2006) p. 036104
- 62 Amaral, L. A. N., Ottino, J., *Eur. Phys. J. B* 38 (2004) p. 147
- 63 Estrada, E., Hatano, N. Submitted 2008

- 64 Bomze, I. M., Budinich, M., Pardalos, P. M., Pelillo, M., in: Du, D.-Z., Pardalos, P. M. (Eds.), *Handbook of Combinatorial Optimization, Supplement Vol. A*, pp. 1-74. Kluwer Academic Publishers, Dordrecht **1999**
- 65 Bron, C., Kerbosch, J., *Comm. ACM* 16 (1973) 575
- 66 Moon, J. W., Moser, L., *Israel J. Math.* 3 (1965) 23
- 67 Zachary, W. W., *J. Anthropol. Res.* 33 (1977) 452
- 68 Guillaume, J.-L., Lapaty, M., *Inform. Proces. Lett.* 90 (2004) 215
- 69 Holme, P., Liljeros, F., Edling, C. R., Kim, B. J., *Phys. Rev. E.* 68 (2003) p. 056107
- 70 König, D., *Theorie der endlichen un enendlichen Graphen.* Leipzig, **1936**
- 71 Seary, A. J., Richards, Jr., W. D., in: *Proceedings of the International Conference on Social Networks*, London, edited by M. G. Everett and K. Rennolds. Greenwich University Press, London, **1995**, Vol. 1, p. 47.
- 72 Xiao, T. Gong, S., *Pattern Recog.* 41 (2008) 1012
- 73 Estrada, E., Higham, D. J., Hatano, N., **2008**, submitted
- 74 Townsend, C., Thompson, R. M., McIntosh, A. R., Kilroy, C., Edwards, E. Scarsbrook, M. R., *Ecol. Lett.* 1 (1998) 200

Figure captions

Fig. 1. Representation of the human protein-protein interaction network. The proteins marked in red are those which have been identified to be involved in human diseases in the Online Mendelian Inheritance in Man (OMIM).

Fig. 2. Part of the collaboration network in Computational Geometry for two author with the same degree centrality but different subgraph centrality: Timothy M. Y. Chan and S. L. Abrams and all their coworkers.

Fig. 3. Models of networks in class II, III and IV according to the classification carried out by using the spectral scaling approach. In the left we show the corresponding spectral scaling for these networks.

Fig. 4. Plot of the two principal roots obtained in the canonical discriminant analysis (CDA) of the 61 networks classified into four different structural classes. Ellipses correspond to 95% of confidence in the CDA.

Fig. 5. (a) The friendship network from the karate club and the two communities identified by Zachary. (b) The communicability graph associated to the karate club network. The numbering is the same in both figures.

Fig. 6. Illustration of the overlapping between two groups or neighborhoods formed among the followers of the administrator (node 34) in the Zachary karate club network.

Fig. 7. (a) Network representation of the food web of Canton Creek. (b) Bipartite structure of this network as found by the method explained here. Nodes in each quasi-bipartite cluster are represented by squares and circles of two different colours. The red thick lines represent the intracluster connections and the gray lines the intercluster links.

Fig. 1

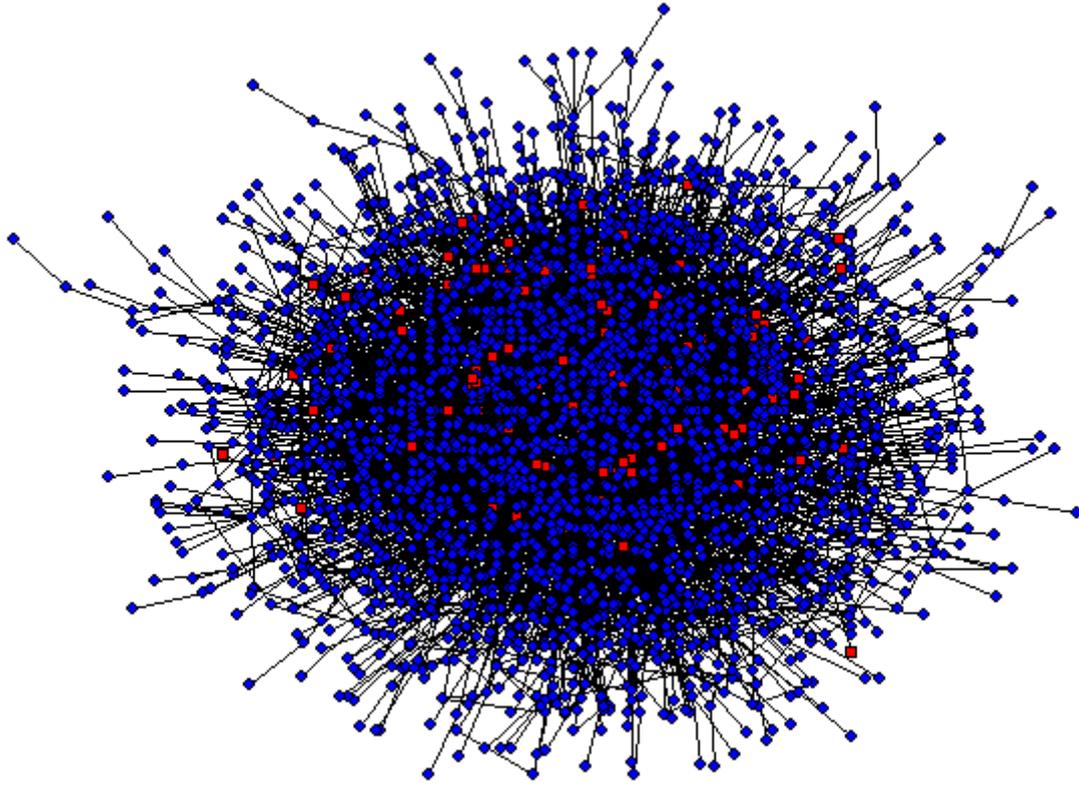


Fig. 2

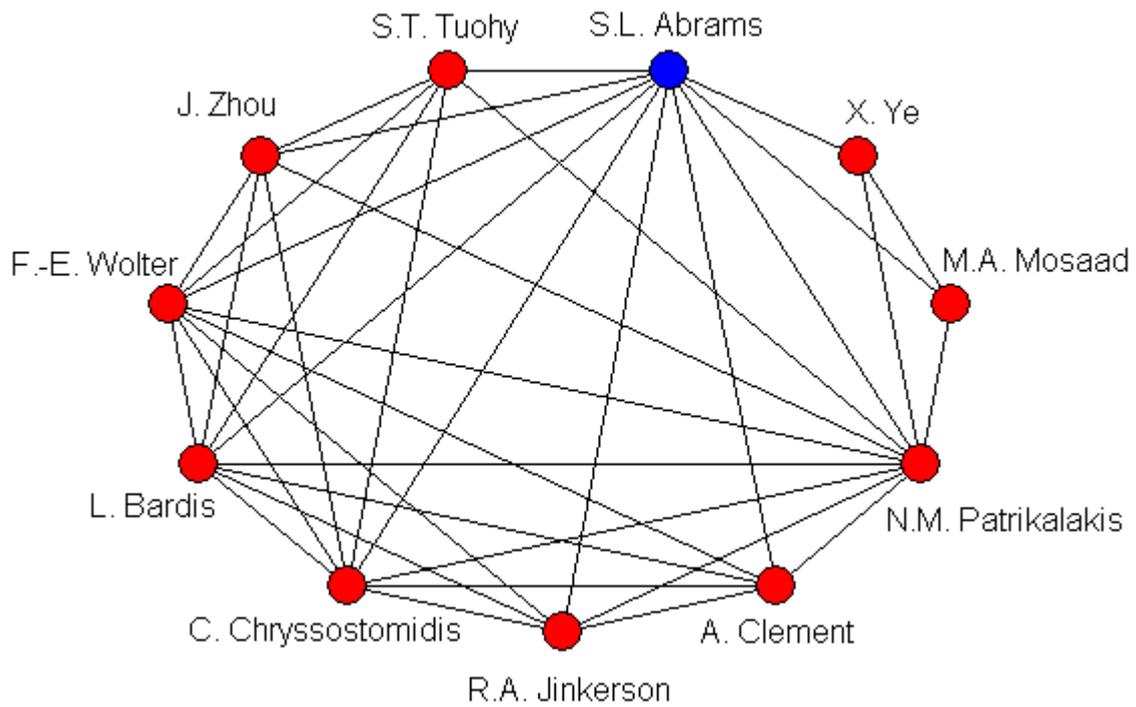
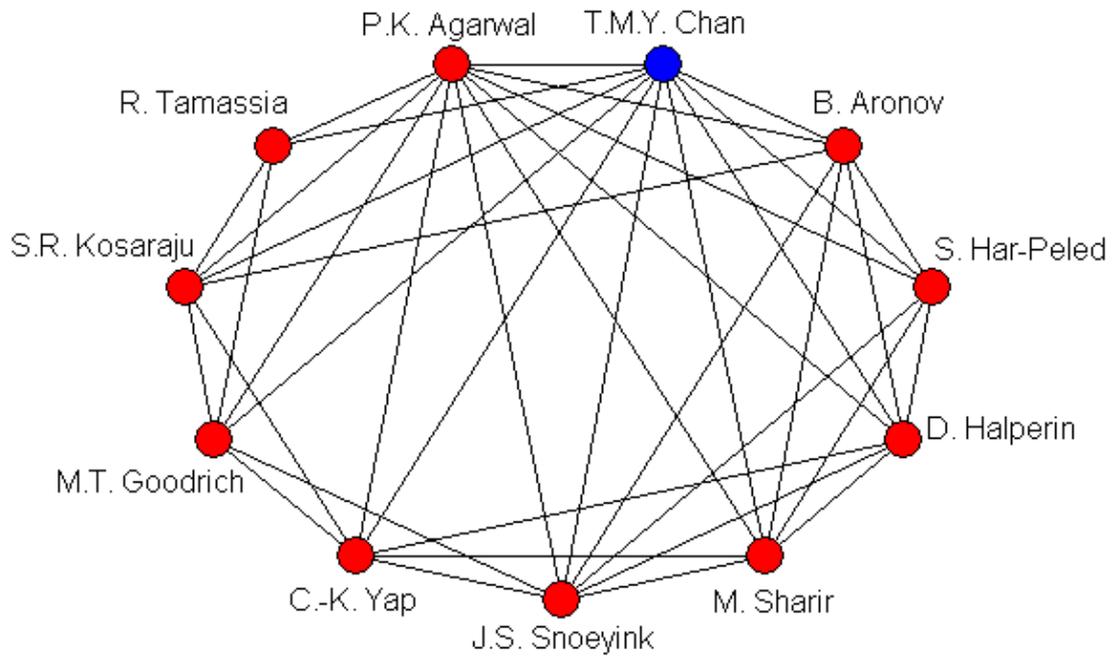


Fig. 3.

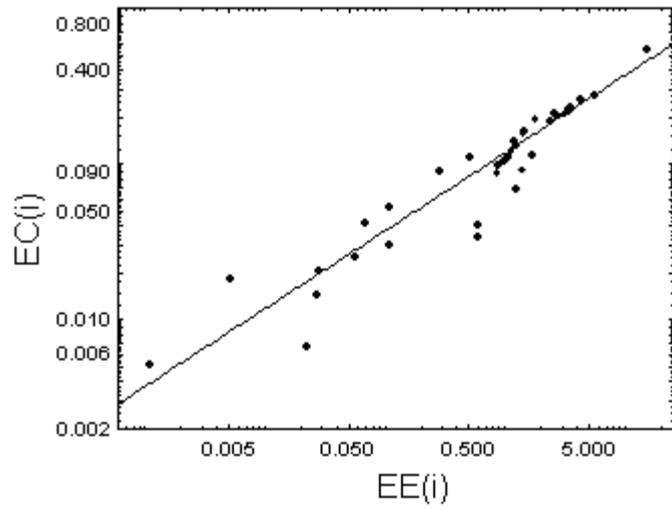
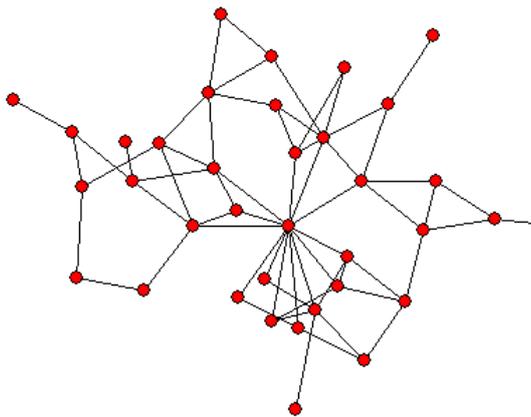
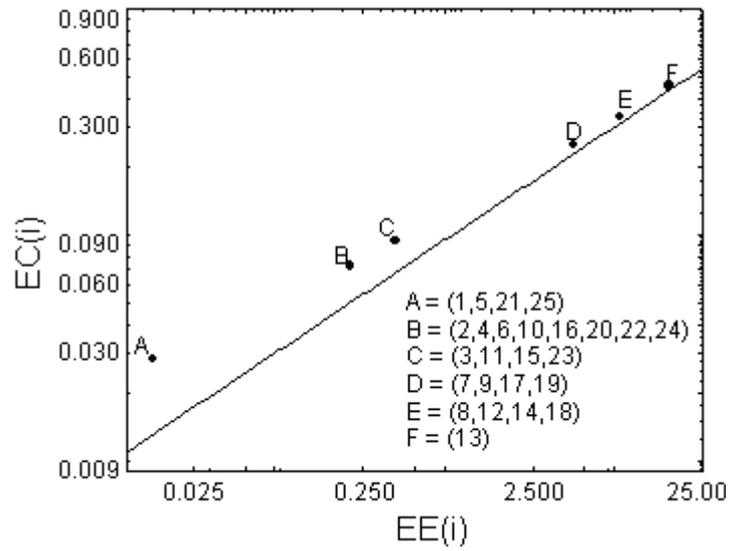
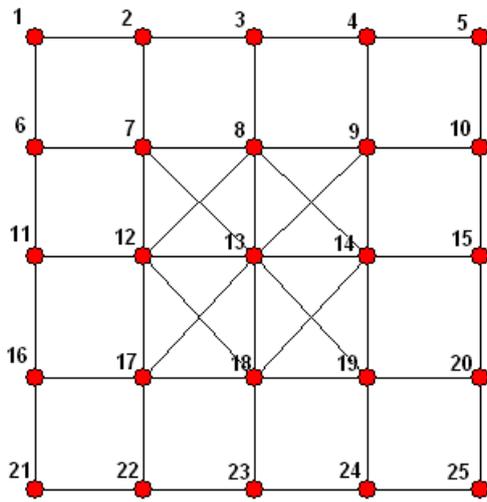
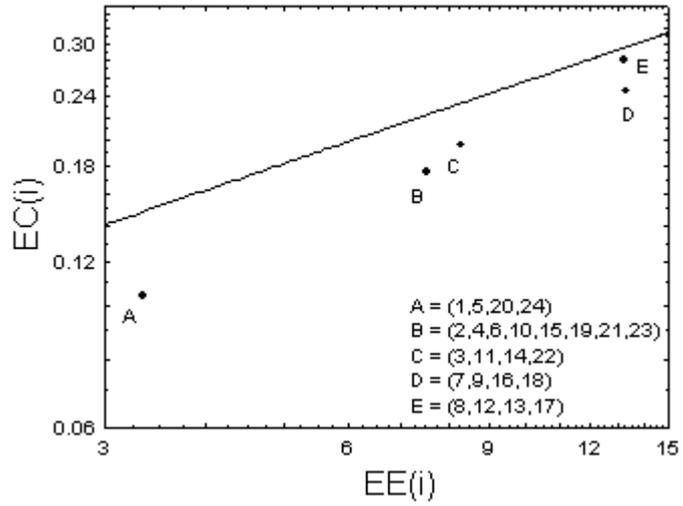
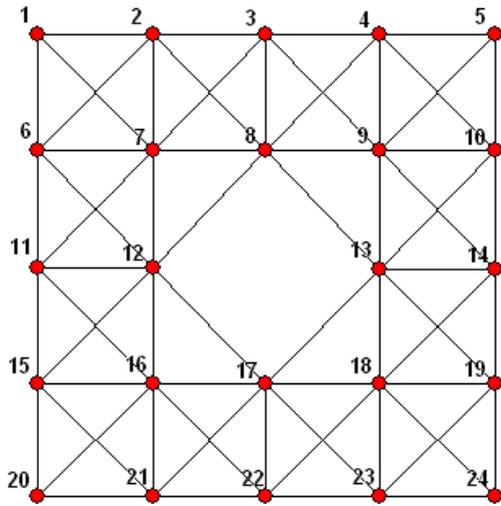


Fig. 4

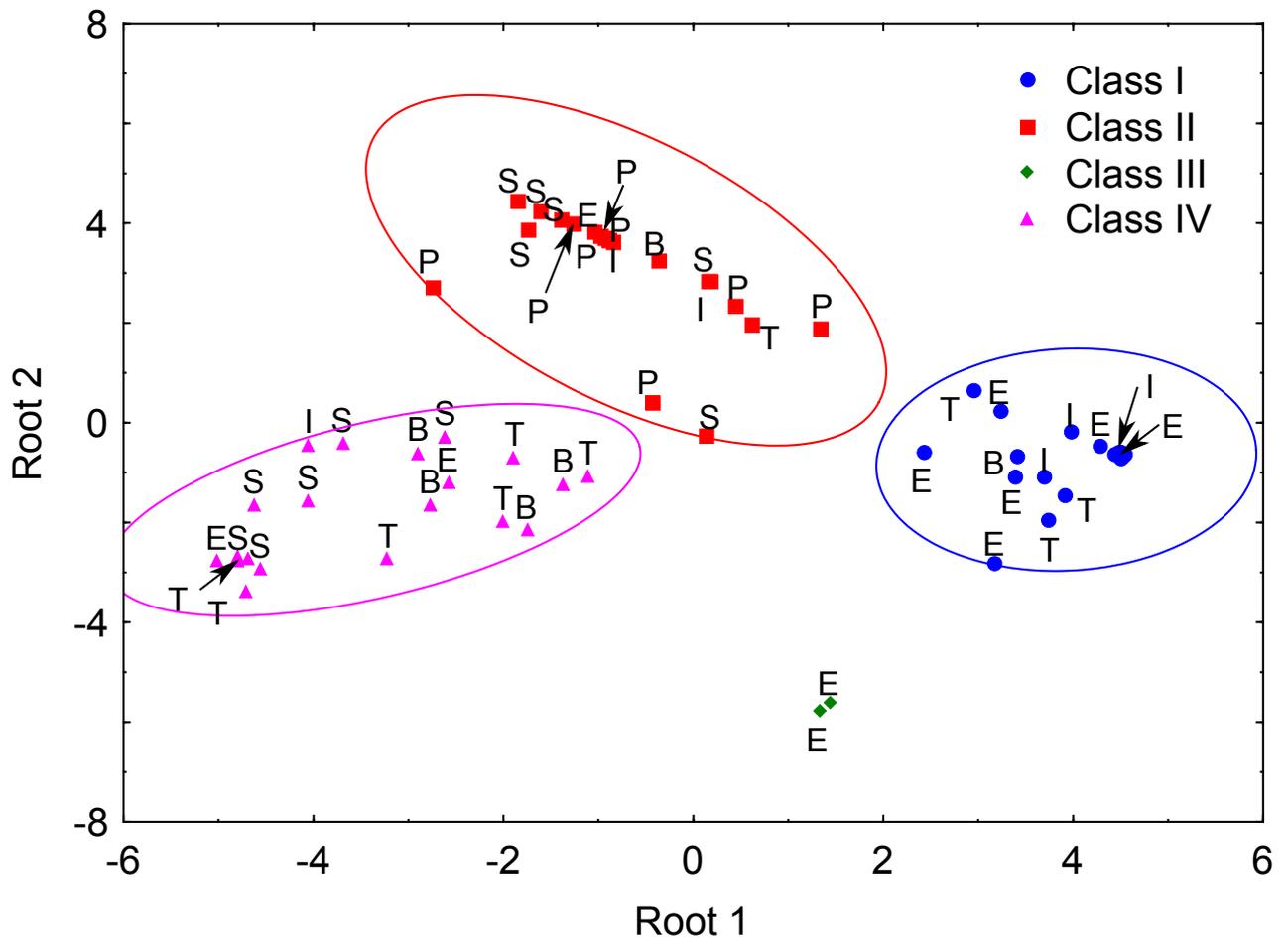
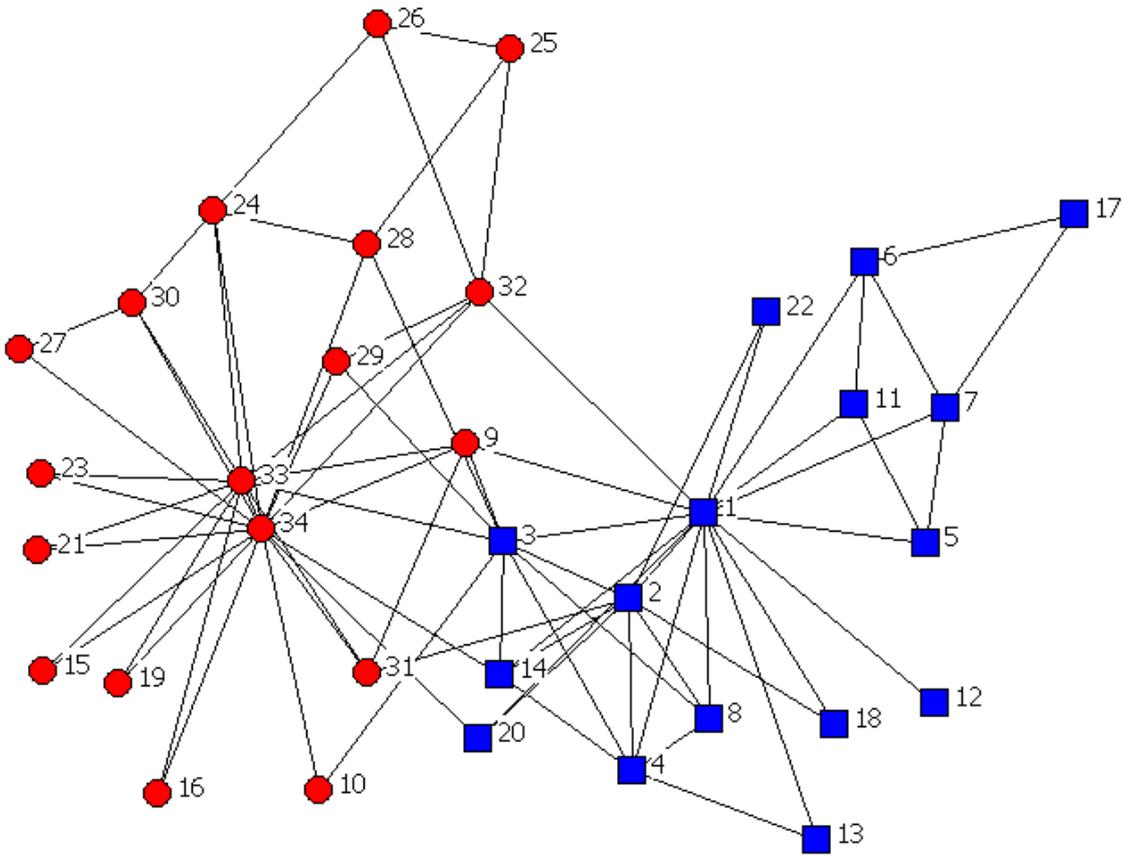


Fig. 5

(a)



(b)

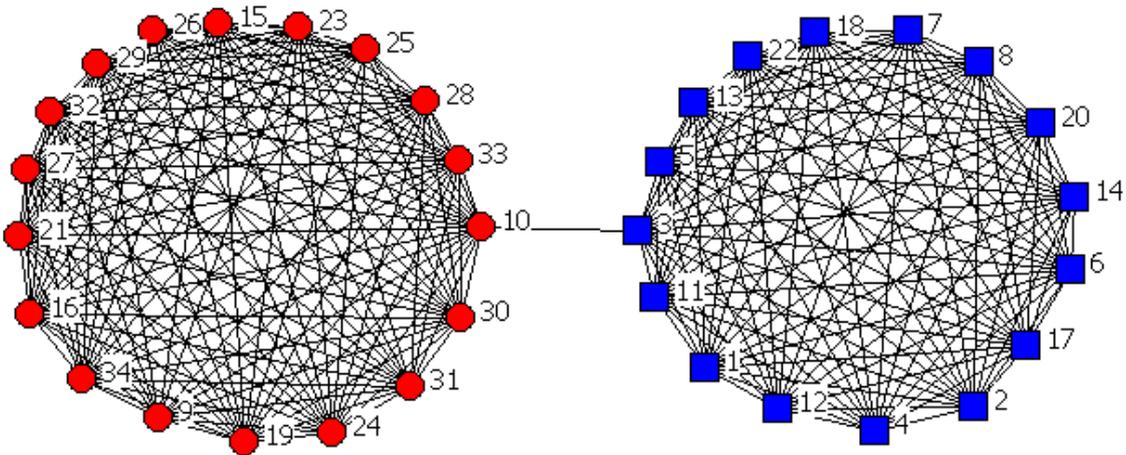
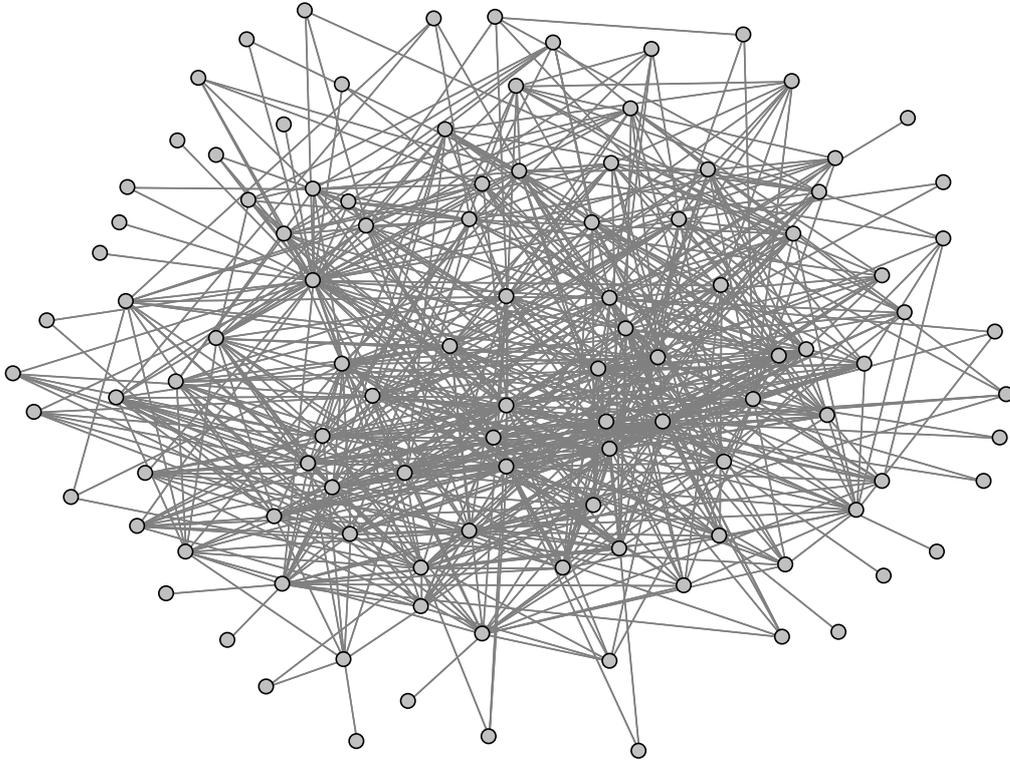


Fig. 7

(a)



(b)

