Gaussianization of the Spectra of Graphs and Networks: Theory and Applications

Alhanouf Ali Alhomaidhi

Department of Mathematics and Statistics
University of Strathclyde, Glasgow

July 14, 2021

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Acknowledgements

First of all, my great thanks go to Allah who made most of my dreams come true.
I am utterly grateful to my supervisor, Professor Ernesto Estrada, who introduced me to the interesting area of Network computations and offered patience, kind supervision and continual support. Over the past six years, he has been a role model from whom I have learnt so much more than mathematical research skills. This work would not have been a Thesis without Prof. Estrada. I also extend my thanks to my internal supervisor at King Saud University professor Fawzi Al-Thukair and Dr Philip Knight who have always been of great support and generous help.
My profound gratitude is due to the Department of Mathematics and Statistics at The University of Strathclyde for providing an excellent research environment and to King Saud University in Saudi Arabia for its generous funding of my studies.
Finally, I am deeply grateful for the love and support I have received from my beloved family, my husband, my lovely children and my great mother.

Abstract

Matrix functions of the adjacency matrix are very useful for understanding important structural properties of graphs and networks, such as communicability, node centrality, bipartivity and many more. Here we propose a new matrix function based on the Gaussianization of the adjacency matrix of a graph. This function gives more weight to a selected reference eigenvalue λ_{ref}, which may be located in any region of the graph spectra. In particular, we study the Gaussian Estrada indices for two reference eigenvalues 0 and -1 separately. In each case, we obtain bounds for this index in simple graphs. We also obtain formulas for the Gaussian Estrada index of Erdős-Rényi random graphs as well as for the Barabási-Albert graphs. Moreover, for $\lambda_{\text{ref}} = 0$, we show that in real-world networks this index is related to the existence of important structural patterns, such as complete bipartite subgraphs (bicliques). Such bicliques appear naturally in many real-world networks as a consequence of evolutionary processes giving rise to them. In addition, we fold the graph spectrum at a given pair of reference eigenvalues, then exponentiate the resulting folded graph spectrum to produce the double Gaussian function of the graph adjacency matrix which give more importance to the reference eigenvalues than to the rest of the spectrum. Based on evidence from mathematical chemistry we focus here our attention on the reference eigenvalues ± 1. They enclose most of the HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) of organic molecular graphs. We prove several results for the trace of the double Gaussian adjacency matrix of simple graphs–the double Gaussian Estrada index– and we apply this index to the classification of polycyclic aromatic hydrocarbons (PAHs) as carcinogenic or non-carcinogenic. We discover that local indices based on the previously developed matrix function allow to classify correctly

100% of the PAHs analyzed. In general, folding the spectrum of the adjacency matrix of networks characterizes important structural information not described in previously used matrix functions of graphs.

Publications and Presentations

The following papers were published in the course of the work contained in this thesis.

- E. Estrada, A.A. Alhomaidhi, F. Al-Thukair, Exploring the "Middle Earth" of network spectra via a Gaussian matrix function. Chaos 27 (2017) 023109.
- A.A. Alhomaidhi, F. Al-Thukair, E. Estrada, Gaussianization of the spectra of graphs and networks. Theory and applications, J. Math. Anal. Appl. 470 (2019) 876-897.
- A. A. Alhomaidhi, F. Al-Thukair, E. Estrada, Double gaussianization of graph spectra, Applied Mathematical Modelling 93 (2021) 134-147,

Some parts of this work have also been presented at the following

- A. A. Alhomaidhi, Exploring the "Middle Earth" of network spectra via a Gaussian matrix function. Talk presented at the fourth Conference on Mathematical Sciences and Applications, Riyadh, Saudi Arabia, 12th April 2018.
- A. A. Alhomaidhi, Folding graph spectra. Extracting structural information hidden in the spectrum. A seminar at the prestigious seminar Rubio de Francia of the University of Zaragoza. 12th November 2019

Contents

List of Figures

List of Tables

Introduction

The study of graph spectra has emerged from applications in Chemistry and Physics. The first mathematical paper on spectral graph theory dates back to 1957 by Collatz and Sinogowitz and was motivated by the membrane vibration problem [1]. Previously, the study of the eigenvalues and eigenvectors of the adjacency matrix of a graph was implicit in the development of the so-called Hückel molecular orbital method since 1930 [2–4]. This method is based on a tight-binding Hamiltonian which is widely used today for the study of certain classes of molecules and solids [5]. In more recent years the study of graph spectra has been widely expanded to many other areas of scientific research. The availability of information about large complex networked systems, ranging from proteomic maps to the Internet has accelerated the application of graph spectral techniques not only in Chemistry and Physics, but in Biology, Computer Science, Economics, Social Sciences, Engineering [6]. Spectral graph theory is nowadays a consolidated area of algebraic graph theory [7]. It is currently used to analyze the structure of graphs and networks through the use of graph invariants that characterize either the nodes, e.g., eigenvector centrality [8,9], pagerank centrality [10], or the global structure of the network, e.g., graph energy [12]. In addition, it is intimately related to many dynamical processes on networks [11], such as epidemic propagation, diffusion, synchronization, among others.

A relatively new approach to study the graph spectra is to consider the analysis of matrix functions, in particular functions of the adjacency matrix of a graph or network [13, 45]. A matrix function is a function that maps a matrix to another matrix [15]. Although the resolvent of the adjacency matrix, $(\alpha I - A)^{-1}$ was used previously for defining the so-called Katz index [16], it was not until the dawn of the XXI century

that the systematic analysis of functions of the adjacency matrix was developed [17–26]. These analyses have given rise to the study of the exponential [26, 27], hyperbolic functions [29], Ψ-functions [28], error function [31], modulus function [32], and sign function [33] of the adjacency matrix of graphs and networks. The well-known concepts of network communicability [27, 34], subgraph centrality [26], Estrada index [22–24], bipartivity index [29, 44], spectral scaling [37, 38], and many more are products of these analyses, which have found many applications in the applied sciences. A large proportion of the matrix functions of the adjacency matrix–$\exp(A)$, $\exp(-A)$, $\sinh(A)$, $\cosh(A)$, $\psi_t(A)$, $\frac{1}{2}\left[\sqrt{2\pi}\mathrm{erf}\left(\frac{A}{\sqrt{2}}\right)+2I\right]\exp\left(A^2/2\right)$, $|A|$– give more weight to the extreme eigenvalues (the largest and/or the smallest) of the adjacency matrix. The problem is that the main structural information contained in other eigenvalues is then hidden by such weighting scheme. For instance, if $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of the adjacency matrix of a connected graph, then if the spectral gap $\lambda_1 - \lambda_2$ is very large we have that $\exp(A)$ is mainly determined by λ_1 and the corresponding eigenvector. The information contained in the rest of the spectrum is almost completely neglected. If we consider matrix functions of the type of $f(A) = \sum_{k=0}^{\infty} c_k A^k$ and consider a simple example, the trace of $f(A) = \exp(A)$ of a simple, connected network, which can be written as $tr\exp(A) = \sum_{j=1}^{n}\exp(\lambda_j)$, where n is the order of the graph and $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of A. It is clear that if the spectral gap of the adjacency matrix, $\lambda_1 - \lambda_2$, is very large, $tr\exp(A)$ depends only on the largest eigenvalue λ_1. This is not a strange situation in real-world networks, where it is typical to find very large spectral gaps for their adjacency matrices. In these cases the use of functions of the type $f(A)$ makes the structural information contained in the smaller eigenvalues and the corresponding eigenvectors of the adjacency matrix not captured by the index. A similar situation happens if we consider $f(-A)$, in this case we give more weight to the smallest eigenvalue/eigenvector of the adjacency matrix and the information contained in the largest ones is again lost. In this Thesis we propose to investigate the information contained in the unexplored part of the spectrum of the adjacency matrix of graphs and networks using a new adjacency matrix function. We propose here the function $f(A) = \exp\left[-\prod_{\lambda_{\mathrm{ref}}}(\lambda_{\mathrm{ref}}I - A)^2\right]$ where λ_{ref} are the

eigenvalues to which we want to give the largest weight in the function.

The first Chapter of this Thesis discusses genaral concepts in network theory and matrix functions that will be used throughout this work.

In the second Chapter we study a Gaussian adjacency matrix function $\exp\left(-A^2\right)$ as a way to characterize the structural information of graphs giving more importance to the eigenvalues/eigenvectors in the middle part of the graph spectrum. Similar Gaussian operators may arise in quantum mechanics of many body systems [41] as well as for the electronic partition function in renormalized tight binding Hamiltonians [42, 43]. We will motivate here the introduction of this function from a quantum-mechanical approach to networks in which a particle is hopping through the nearest neighbor nodes in a graph. We then study here properties of $H = tr \exp\left(-A^2\right)$. We show that although the graph nullity—the multiplicity of the zero eigenvalue of the adjacency matrix of the graph—plays an important role in the values of this index, the H index contains more structural information than the graph nullity even for small simple graphs. We then prove that among the graphs with n nodes, the maximum of the H index is always obtained for the star graph followed by other complete bipartite graphs. Then, we obtain analytic expressions for this index in random graphs with Poisson and power-law degree distribution, showing that the last ones always display larger values of the H index than the first ones. Finally, we study more than 60 real-world networks representing a large variety of complex systems. We found that the networks with the largest index correspond to those having relatively large bicliques—complete bipartite subgraphs, which can be created by different evolutionary mechanisms depending on the kind of complex system considered.

The third chapter is devoted to investigate the matrix function: $\exp\left[-(I+A)^2\right]$. We obtain a few mathematical results for the trace of this matrix function for simple graphs as well as two models of random graphs. Also, we prove that the trace takes its maximum for the complete graph among all connected graphs of same number of nodes.

Motivated by mathematical chemistry, we study the folding of the graph spectrum in the fourth Chapter at a given pair of reference eigenvalues and then exponentiate the

resulting folded graph spectrum. When studying molecules with the so-called tight-binding Hamiltonians, e.g., the Hückel molecular orbital (HMO) approach [4], there are two eigenvalues of the graph spectra which play a fundamental role in understanding molecular properties. They are known as the highest occuppied (HOMO) and the lowest unoccupied molecular orbitals (LUMO), respectively [122]. Fowler and Pisanski [123] have called "normal" the molecular graphs for which $+1 \geq \lambda_{HOMO} \geq \lambda_{LUMO} \geq -1$, while the rest of molecular graphs are called "exceptional". The reason for this is that most of molecular graphs have their HOMO and LUMO within the 'chemical triangle' of an HOMO-LUMO map [123]–a scatterplot of the middle eigenvalues of the graph–, with vertices at $(-1,-1)$, $(+1,-1)$, $(+1,+1)$. They proved that all chemical trees lie within the triangle, as do all chemical graphs with up to 12 vertices [123]. Thereby, we focus our attention on the reference eigenvalues ± 1. They enclose most of the HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) of organic molecular graphs. We prove several results for the trace of the double Gaussian adjacency matrix of simple graphs–the double Gaussian Estrada index. Finally we apply this index to the classification of polycyclic aromatic hydrocarbons (PAHs) as carcinogenic or non-carcinogenic. We discover that local indices based on the double Gaussian matrix function allow to classify correctly 100% of the PAHs analyzed.

Chapter 1

Preliminaries

This chapter collects the basic definitions and concepts that we need in the Thesis. The first part covers the essential material on graph theory and the second presents the definition of matrix functions and relates it with network theory. The last part presents a simple brief about the Hückel Molecular Orbital Theory. From now on, we will use the terms graph and network interchangeably.

1.1 Graph Theory

Networks can always be thought of as a collection of items and the connection between them. In order to give a formal definition we introduce some basic set notation. Let V be a non-empty set and let $E \subseteq V \times V$, whose elements are not necessarily all distinct. E is reflexive if $(v, v) \in E$ for all $v \in V$, it is anti-reflexive if $(v, v) \notin E$ for all $v \in V$ and it is symmetric if $(v_1, v_2) \in E \implies (v_2, v_1) \in E$.

Definition 1.1.1. A network G is a pair (V, E) where $V \neq \emptyset$. Networks are also known as graphs. V is called the vertex set of G; its elements are the vertices of G (also known as nodes) and E is called the set of edges (or links).

- If E is symmetric then G is an undirected network.
- If E is symmetric, anti-reflexive and contains no duplicate edges then G is a simple network.

- If E is nonsymmetric then G is a directed network.

Throughout the Thesis, only simple graphs are considered. In the following we list some features of graphs that will be used in the this Thesis.

Definitions.

- For a graph G, we denote $n = |V(G)|$, $m = |E(G)|$. The number of vertices n is called the order of G, and m is the size of G.
- A walk of length k in G is a set of nodes $v_1, v_2, \ldots, v_k, v_{k+1}$ such that for all $1 \leq l \leq k$, $(v_l, v_l + 1) \in E$. A closed walk is a walk for which $v_1 = v_{k+1}$. A path is a walk with no repeated nodes. A cycle is a closed walk with no repeated nodes.
- The vertices $v_1, v_2 \in V$ are called adjacent if $(v_1, v_2) \in E$.
- A graph is connected if there is a path between every pair of its vertices.
- The degree of a node i is the number of nodes to which it is connected, and is denoted by k_i.
- $G_s = (V_s, E_s)$ is a subgraph of $G = (V, E)$ if $V_s \subseteq V$ and $E_s \subseteq (V_s \times V_s) \cap E$.
- $G_s = (V_s, E_s)$ is an induced subgraph of $G = (V, E)$ if $V_s \subseteq V$ and $E_s = (V_s \times V_s) \cap E$.

The following theorem connects the degrees of the vertices and the number of edges. It is the first theorem of graph theory [40].

Theorem 1.1.1. The sum of the degrees of the vertices of a graph is equal to twice the number of its edges.

1.1.1 Some Special Graphs

The following definitions describe briefly a set of some known graphs that will be mentioned later.

- A graph $G = (V, E)$ is bipartite if the nodes can be divided into nonempty disjoint sets $V_1 \cup V_2$ such that $(u, v) \in E \implies u \in V_i, v \in V_j, i \neq j$.

- The path graph is a graph such that its set of vertices can be labeled in a way all the edges are of the form (v_i, v_{i+1}) where $i = 1, \ldots, n-1$. A path graph with n vertices is denoted by P_n.
- The cycle graph of n vertices is the graph obtained by adding the edge (v_n, v_1) to P_n, and is denoted by C_n.
- The complete graph is a graph in which every two vertices are adjacent. A complete graph with n vertices is denoted by K_n.
- The complete bipartite graph is a bipartite graph $(V_1 \cup V_2, E)$ such that for every two vertices $u \in V_1$ and $v \in V_2$, $(u, v) \in E$. A complete bipartite graph is denoted by K_{n_1,n_1}, where $|V_1| = n_1$, $|V_2| = n_2$.
- An independent set is a set of nodes in the graph in which no pair of them are adjacent.
- A k-partite graph is a graph whose vertices are partitioned into k different independent sets, and a complete k-partite graph is a k-partite graph in which there is an edge between every pair of vertices from different independent sets.
- The star graph $K_{1,n-1}$ is the connected complete bipartite graph in which there is one node connected to $n-1$ nodes.
- A tree is a connected graph with no cycles.

1.1.2 Adjacency Matrix of Graphs

One of the most commonly used representations of graphs is the adjacency matrix. In a network with n vertices we can label each one with an element from the set $V = \{v_1, v_2, \ldots, v_n\}$ then the adjacency matrix associated with the network is defined as follows:

Definition 1.1.2. Let $G = (V, E)$ be a simple network where $V = \{v_1, v_2, \ldots, v_n\}$. For $1 \leq i, j \leq n$ define

$$a_{ij} = \begin{cases} 1, & \text{if } (v_i, v_j) \in E \\ 0 & \text{if } (v_i, v_j) \notin E \end{cases}$$

Then the square matrix $A = (a_{ij})$ is called the adjacency matrix of G. The set of eigenvalues of the adjacency matrix together with their multiplicities is called the spectrum of the graph.

For any graph G, let $\lambda_1, \lambda_2, \ldots, \lambda_s$ be the distinct eigenvalues of G with multiplicities $m_1, m_2, \ldots, m_s$ respectively, then the spectrum of G can be written as

$$Sp(G) = \{[\lambda_1]^{m_1}, [\lambda_2]^{m_2}, \ldots, [\lambda_s]^{m_s}\}. \tag{1.1}$$

In the particular case of an undirected network as the ones studied in this Thesis, the associated adjacency matrix is real and symmetric, thus its eigenvalues are real. We label the eigenvalues of A in a non-increasing order: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Moreover, since A is nonnegative, Perron-Frobenius theorem states that $\lambda_1 \geq |\lambda_i|$ for all $i = 1, \ldots, n-1$ and there exists a corresponding eigenvector that is nonnegative. The following Lemma states the spectrum of some known simple graphs [39].

Theorem 1.1.2.

- $Sp(K_n) = \left\{[-1]^{n-1}, [n-1]^1\right\}$.
- $Sp(K_{n_1,n_2}) = \left\{[-\sqrt{n_1 n_2}]^1, [0]^{n_1+n_2-2}, [\sqrt{n_1 n_2}]^1\right\}$.
- $Sp(P_n) = \left\{2\cos\left(\frac{\pi j}{n+1}\right), j = 1, \ldots, n\right\}$.
- $Sp(C_n) = \left\{2\cos\left(\frac{2\pi j}{n}\right), j = 1, \ldots, n\right\}$.

The following definition is related to the spectrum of graphs

Definition 1.1.3. The nullity of a graph is the multiplicity of the eigenvalue zero in the spectrum of the adjacency matrix of the graph.

We conclude the section by the following important theorem that connects the powers of the adjacency matrix of a graph G to the walks in the graph.

Theorem 1.1.3. [40] For any integer $k \geq 1$, the (i, j)-entry of the matrix A^k is equal to the number of walks from the node v_i to v_j of length k, where A is the adjacency matrix of the graph G.

1.1.3 Random Models of Networks

Random models of networks are used in the study of real-world networks as comparison tools in order to understand the structure of complex networks. We introduce here two simple and general models for generating random networks: the Erdős-Rényi model and Barabási-Albert model.

The Erdős-Rényi model of random network

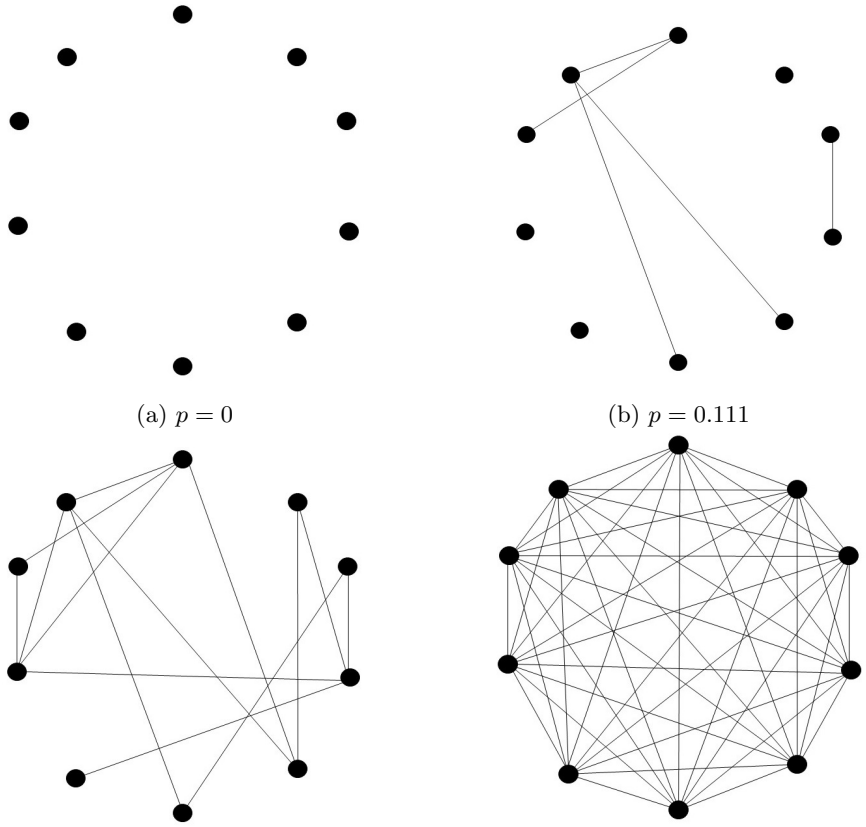
In this model, put forward by Erdős-Rényi in 1959, we start with n isolated nodes. We then pick randomly and independently a pair of nodes and with probability $p > 0$ we add a link between them. In practice we proceed as follow. First, we fix a parameter p. Then, for each pair of nodes we generate a random number, s, uniformly from $[0, 1]$ and if $p > s$ we add a link between them. In Figure 1.1 we illustrate some examples of Erdős-Rényi random networks with 10 nodes and different linking probabilities. The Erdős-Rényi (ER) random network is denoted by $G(n, p)$. The $G(n, p)$ random graphs have a Poisson degree distribution (when $n \to \infty$) and display the Wigner semi-circle distribution [14] of eigenvalues when $n \to \infty$ of the form

$$\rho(\lambda) = \begin{cases} \frac{2\sqrt{r^2-\lambda^2}}{\pi r^2} & , -r \leq \lambda \leq r \\ 0, & otherwise, \end{cases} \tag{1.2}$$

where $r = 2\sqrt{np(1-p)}$. This normalization factor r has been introduced previously by Farkas et al. [35] for studying the spectra of random networks. An example of the density distribution for a network generated with the ER model with parameters $n = 300$ and $p = 0.026$ is illustrated in Figure 1.2 (a).

The Barabási-Albert model

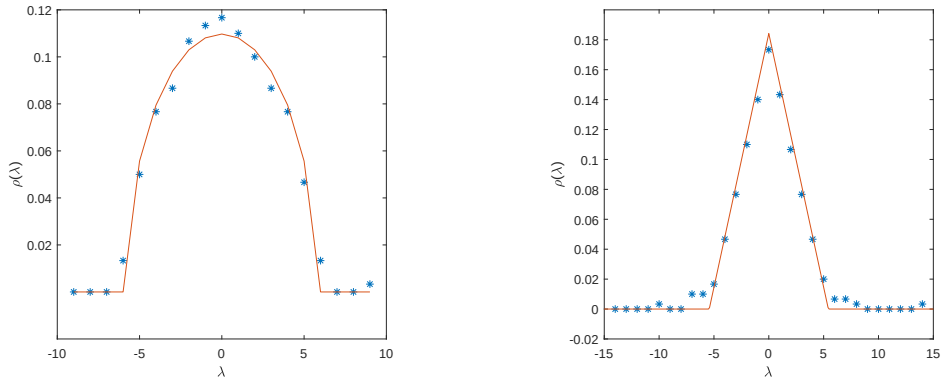
The model was put forward by Barabási and Albert in 1999 [36] to generate networks in which the probability of finding a node of degree k decays as a power law of the degree. We initialize with a small network with m_0 nodes, we can assume that our initial random network is connected and of ER type. At each step we add a new node



(a) $p = 0$ (b) $p = 0.111$

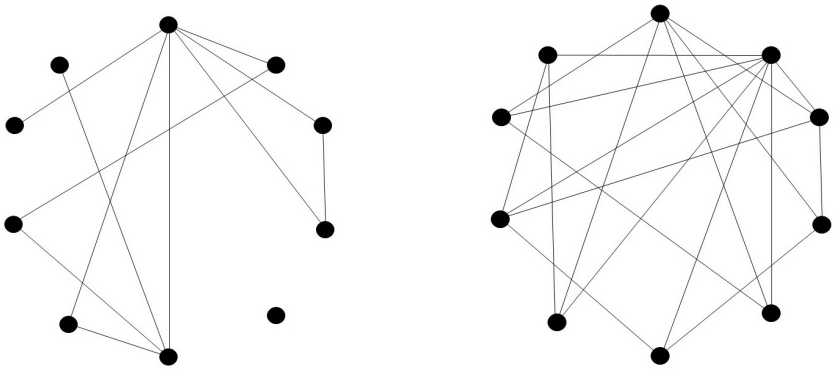
(c) $p = 0.311$ (d) $p = 1$

Figure 1.1: Erdős-Rényi random networks for different probabilities p.



(a) (b)

Figure 1.2: Spectral density for a network generated with the ER model (a) and the BA model (b) with $n = 300$ and $p \approx 0.026$.



(a) $m = 2$ (b) $m = 3$

Figure 1.3: Barabási-Albert networks with $n = 10$ and $m = 2$ (a), $m = 3$ (b).

to the network and connect it to $m \leq m_0$ of the existing nodes with a probability that is proportional to the degree of the existing nodes. This process is known as preferential attachment. In this case, it is known that $p = \dfrac{2m}{n-1}$. Thus, we can construct the normalization factor $r = 2\sqrt{np(1-p)}$ based on m_0. Two examples of Barabási-Albert networks are presented in Figure 1.3. The density of the eigenvalues of BA graphs follow a triangular distribution [14] of the form

$$\rho(\lambda) = \begin{cases} \frac{\lambda+r}{r^2}, & -r \leq \lambda < 0 \\ \frac{r-\lambda}{r^2}, & 0 < \lambda \leq r \\ 0 & otherwise. \end{cases} \tag{1.3}$$

An example of the density distribution for a network generated with the BA model with parameters $n = 300$ and $m = 4$ is illustrated in Figure 1.2 (b).

1.2 Matrix Functions

There are many equivalent ways of defining matrix functions, we present one definition that is related to the Jordan canonical form of a matrix. Other two equivalent definitions based on polynomial interpolation and Cauchy integral theorem can be found in [15].

Any matrix $A \in \mathbb{C}^{n\times n}$ can be expressed in the Jordan canonical form:

$$Z^{-1}AZ = J = diag(J_1, J_2, \ldots, J_p)$$

$$J_k = J_k(\lambda_k) = \begin{bmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{bmatrix} \in \mathbb{C}^{m_k\times m_k}, \tag{1.4}$$

where Z is nonsingular and $m_1 + m_2 + \ldots + m_p = n$.
Denote by $\lambda_1, \ldots, \lambda_s$ the distinct eigenvalues of A and let n_i be the order of the largest Jordan block in which λ_i appears. In order to give the definition of matrix functions we need the following terminology.

Definition 1.2.1. The function f is said to be defined on the spectrum of A if the values $f^j(\lambda_i)$, $j = 0, \ldots, n_i - 1$, where $i = 1, \ldots, s$ exist.

Definition 1.2.2. Let f be defined on the spectrum of $A \in \mathbb{C}^{n\times n}$ and let A have the Jordan canonical form 1.4. Then

$$f(A) := Zf(J)Z^{-1} = Zdiag(f(J_k))Z^{-1}$$

$$f(J_k) := \begin{bmatrix} f(\lambda_k) & f'(\lambda_k) & \cdots & \frac{f^{(m_k-1)}(\lambda_k)}{(m_k-1)!} \\ & f(\lambda_k) & \ddots & \vdots \\ & & \ddots & f'(\lambda_k) \\ & & & f(\lambda_k) \end{bmatrix} \tag{1.5}$$

The Jordan canonical form of a real symmetric matrix, which is the case of the adjacency matrix of a simple network, is of the form $A = Q\Lambda Q^T$ where Q is a real orthonormal matrix with columns of which are eigenvectors of A , and Λ is real and diagonal having the eigenvalues of A on the diagonal. Thus, definition (1.2.2) yields $f(A) = Qf(\Lambda)Q^T = Qdiag(f(\lambda_i))Q^T$. Therefore $f(A)$ has the same eigenvectors as A and its eigenvalues are obtained by applying f to those of A. The following theorem guarantees the validity of a matrix Taylor series if the eigenvalues of the "increment" lie within the radius of convergence of the associated scalar Taylor series.

Theorem 1.2.1. [15] Suppose f has a Taylor series expansion

$$f(z) = \sum_{k=0}^{\infty} \frac{f^{(k)}(\alpha)}{k!}(z-\alpha)^k \tag{1.6}$$

with radius of convergence r. If $A \in \mathbb{C}^{n\times n}$ then $f(A)$ is defined and is given by

$$f(A) = \sum_{k=0}^{\infty} \frac{f^{(k)}(\alpha)}{k!}(A-\alpha I)^k \tag{1.7}$$

if and only if each of the distinct eigenvalues $\lambda_1, \ldots, \lambda_s$ of A satisfies one of the following conditions

(a) $|\lambda_i - \alpha| < r$,

(b) $|\lambda_i - \alpha| = r$ and the series for $f^{(n_i-1)}(\lambda)$ is convergent at the point $\lambda = \lambda_i$, $i = 1, \ldots, s$.

1.3 Functions of the Adjacency Matrix of Graphs

Matrix functions have emerged as an important mathematical tool for studying networks [45]. The greatest appeal of the use of functions of the adjacency matrix for studying graphs is that when representing them in terms of a Taylor series expansion: $f(A) = \sum_{k=0}^{\infty} c_k A^k$, the entries of the kth power of the adjacency matrix provides information about the number of walks of length k between the corresponding pair of (not necessarily different) nodes. Functions of adjacency matrices are widely used as centrality and communicability measures of a network. The notion of centrality measure of a node is used in the determination of the most important nodes in a network [14], it quantifies the relative importance of each node in the graph [13]. The communicability measure between nodes in a network quantifies how well different parts of a network are connected to one another [13]. In this section we provide a short review about the use of matrix functions in the analysis of graphs and networks.

1.3.1 Katz Centrality

A classical node centrality is the degree centrality which measures the ability of a node to communicate directly with others. So with degree centrality, i is more central than j if $k_i > k_j$. The degree of the node i counts the number of walks of length one from i to every other node of the network. In 1953, Katz [16] extended this idea to count not only the walks of length one, but those of any length starting at node i giving more weight to shorter walks than to longer ones [14]. The Katz index is given by

$$\begin{aligned} K_i &= \left[\left(\alpha^0 A^0 + \alpha A + \alpha^2 A^2 + \ldots + \alpha^k A^k + \ldots\right)\mathbf{e}\right]_i \\ &= \left[\sum_{k=0}^{\infty}\left(\alpha^k A^k\right)\mathbf{e}\right]_i \end{aligned} \tag{1.8}$$

where $\mathbf{e}$ is a column vector of entries 1 and A is the adjacency matrix of the graph. The value of the attenuation factor α has to be chosen such that it is smaller than the reciprocal of the largest eigenvalue of A, i.e $\alpha < \frac{1}{\lambda_1}$, in order to get a convergence series. In this case we have

$$K_i = \left[(I - \alpha A)^{-1}\mathbf{e}\right]_i. \tag{1.9}$$

When deriving the index, we can ignore the contribution from $A^0 = I$ and instead use

$$\overline{K}_i = \left[\left((I - \alpha A)^{-1} - I\right)\mathbf{e}\right]_i. \tag{1.10}$$

1.3.2 The Communicability Function

Katz centrality is computed from the entries of the matrix

$$K = \sum_{k=0}^{\infty}\alpha^k A^k. \tag{1.11}$$

The idea can be generalized to work with other weighted sums of powers of the adjacency matrix, namely [14],

$$f(A) = \sum_{k=0}^{\infty} c_k A^k. \tag{1.12}$$

The coefficients c_k need to be selected in such a way that the series (1.12) converges. Also, they should give more weight to small powers of the adjacency matrix than to the larger ones. By selecting $c_k = \frac{1}{k!}$ we have

$$f(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k = e^A. \tag{1.13}$$

Subgraph centrality is the term used for measures that characterize the participation of a node in the different subgraphs of a network, giving more importance to the smaller than to the bigger ones. Due to the nature of closed walks in graphs, subgraph centralities can be defined by using the diagonal entries of the matrix function of the form $f(A) = \sum_{k=0}^{\infty} c_k A^k$. Estrada and Rodríguez-Velásquez [26] introduced the subgraph centrality based on the exponential function of the adjacency matrix and is defined by

$$EE_i = \left(\sum_{k=0}^{\infty} \frac{1}{k!} A^k \right)_{ii} = \left(e^A \right)_{ii}. \tag{1.14}$$

Using the spectral decomposition of the adjacency matrix, the subgraph centrality can be presented in terms of the eigenvalues and eigenvectors of the adjacency matrix as follows:

$$EE_i = \sum_{k=1}^{n} \mathbf{q}_k(i)^2 \exp(\lambda_k), \tag{1.15}$$

where $\mathbf{q}_k$ is the eigenvector corresponding to the eigenvalue λ_k. The communicability between two nodes p and q in a network is defined by

$$G_{pq} = \sum_{k=0}^{\infty} \frac{1}{k!} \left(A^k \right)_{pq} = \left(\exp(A) \right)_{pq} \tag{1.16}$$

$$= \sum_{k=1}^{n} \mathbf{q}_k(p) \mathbf{q}_k(q) \exp(\lambda_k) \tag{1.17}$$

which was first introduced by Estrada and Hatano [27]. It provides an effective measure of the connectivity between pairs of nodes. The underlying intuition is that two nodes p and q are well-connected if there are many walks in the graph G connecting p and q with shorter walks being given more weight than longer ones.

The Estrada index of a network is defined to be the trace of $f(A) = \exp(A)$ and was first proposed by Estrada in 2000 [30] as a way of characterizing the degree of folding of proteins and in 2005 Estrada and Rodríguez-Velázquez [26] introduced this index for studying complex networks. The index was then renamed by de la Peña et al. [23] as the 'Estrada index' of a graph. The index can be expressed in terms of the eigenvalues of the adjacency matrix as follows:

$$EE = \sum_{j=1}^{n} e^{\lambda_j}. \tag{1.18}$$

Then, the Estrada index can be defined as the sum of the subgraph centralities of the network. The communicability function $f(A) = e^A$ has proved very useful in a number of applications. Some recent examples include:

- Detection of communities in networks [52];
- Detection of brain strokes in humans [46];
- Cancer Therapeutics [47];
- Detection of human genetic diseases [48];
- Analysis of economic and financial systems [49];
- Deformation of granular materials [50];
- Navigation of autonomous vehicles [51].

The spectral moments of a graph G are defined as:

$$\mu_k = \sum_{i=1}^{n} \lambda_i^k, \tag{1.19}$$

where μ_k is the kth spectral moment of G. Using spectral moments of the graph, the Estrada index can be defined by a Taylor expansion of the form

$$EE(G) = \mu_0 + \mu_1 + \frac{\mu_2}{2!} + \frac{\mu_3}{3!} + \cdots + \frac{\mu_k}{k!} + \cdots . \tag{1.20}$$

Then, by moving forward one step the spectral moments respect to the factorial denominators we get

$$EE^1(G) = \mu_0 + \frac{\mu_1}{2!} + \frac{\mu_2}{3!} + \frac{\mu_3}{4!} + \cdots + \frac{\mu_k}{(k+1)!} + \cdots . \tag{1.21}$$

which have the following spectral formula only in the case when no eigenvalue is equal to zero

$$EE^1(G) = \sum_{j=1}^{n} \frac{e^{\lambda_j} - 1}{\lambda_j}. \tag{1.22}$$

Then $EE^1(G)$ can be obtained as the trace of the $\psi_1(A)$ matrix function [15]

$$EE^1(G) = tr\psi_1(A) = tr\left[A^{-1}\left(e^A - I\right)\right], \tag{1.23}$$

where A is a nonsingular matrix. The positive rescaling approach can be extended to generate a series of indices characterizing a graph in terms of the spectral moments of the adjacency matrix weighted by inverse factorials. That is

$$EE^t(G) = \sum_{k=0}^{\infty} \frac{\mu_k}{(k+t)!} \tag{1.24}$$

which are related to matrix functions through the trace formula:

$$EE^t(G) = tr\psi_t(A), \tag{1.25}$$

where $\psi_t(A)$ matrix functions have the following integral formula:

$$\psi_t(A) = \frac{1}{(t-1)!} \int_0^1 e^{(1-x)A} x^{t-1} dx. \tag{1.26}$$

A negative rescaling can be defined by

$$EE^{-t}(G) = \sum_{s=0}^{t-1} \mu_s + \sum_{k=t}^{\infty} \frac{\mu_k}{(k-t)!}, \tag{1.27}$$

which have the following spectral realization:

$$EE^{-t}(G) = \sum_{j=0}^{n} \left(\sum_{s=0}^{t-1} \lambda_j^s + \lambda_j^t e^{\lambda_j} \right). \tag{1.28}$$

And the generalized subgraph centrality indices defined by using positive and negative rescaling [28] are:

$$EE^{t}(i) = [\psi_t(A)]_{ii}\,; \tag{1.29}$$

$$EE^{-t}(i) = \left(\sum_{s=0}^{t-1} A^s + A^t e^A \right)_{ii}. \tag{1.30}$$

By penalizing more the longest walks (positive rescaling) in the original definition of the subgraph centrality, the concentration is more in the local environment of the corresponding node. On the other hand, the negative rescaling decreases the penalization imposed to the spectral moments in the original definition of the subgraph centrality which corresponds to a zooming out of the surrounds of the corresponding node. In this case we allow long walks to contribute to the index in such a way that we obtain more global information about the environment of the node under study [28]. The zooming out strategy (negative rescaling) gives a more global picture of the topological surrounds of a node while the zooming in strategy (positive rescaling) focuses more on the local topological environment of a node. The generalized Subgraph centrality indices defined earlier were applied to the study of protein–protein interaction (PPI) networks and have been able to identify more essential proteins in the yeast PPI network than any of the other centrality measures studied in [28].

Moreover, a distance measure defined on the basis of the communicability function was proposed by Estrada [53]. For any pair of nodes p and q in a network, the communicability function $(\exp(A))_{pq}$ accounts for a successive travel of information

between the nodes p and q while $(\exp(A))_{pp}$ and $(\exp(A))_{qq}$ accounts for disrupted information sent from p and q respectively, then the communicability distance between the nodes p and q accounts for the amount of information disrupted minus the amount of information that arrives at its destination. That is

$$\xi_{pq} = \sqrt{(e^A)_{pp} + (e^A)_{qq} - 2\,(e^A)_{pq}}. \tag{1.31}$$

The communicability distance was proven to be a Euclidean distance [53]. Moreover, it induces an embedding of the graph G of size n into a hypersphere in an n-dimensional space [54] which gives rise to the definition of the communicability angle [55] which is defined by

$$\cos\theta_{pq} = \frac{\left(e^A\right)_{pq}}{\sqrt{(e^A)_{pp}\,(e^A)_{qq}}}. \tag{1.32}$$

The average communicability angle accounts for the spatial efficiency of networks and it was used to identify the critical links in consensus dynamics problems [56].

1.3.3 Hyperbolic Matrix Functions

The communicability function of a graph G can be presented by separating the contributions of odd and even walks, giving rise to expressions for the communicability function in terms of hyperbolic matrix functions, that is

$$\begin{aligned} e^A &= \sum_{k=0}^{\infty} \frac{1}{k!} A^k \\ &= \sum_{k=0}^{\infty} \frac{1}{(2k)!} A^{2k} + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} A^{2k+1} \\ &= \cosh(A) + \sinh(A). \end{aligned} \tag{1.33}$$

The subgraph centralities based on hyperbolic communicability functions are defined by

$$EE_i^{even} = (\cosh(A))_{ii} = \left(\sum_{k=0}^{\infty} \frac{1}{(2k)!} A^{2k} \right)_{ii}; \tag{1.34}$$

$$EE_i^{odd} = (\sinh(A))_{ii} = \left(\sum_{k=0}^{\infty} \frac{1}{(2k+1)!} A^{2k+1} \right)_{ii}. \tag{1.35}$$

And the Estrada index can be written as

$$\begin{aligned} EE &= tr\left[\cosh(A)\right] + tr\left[\sinh(A)\right] = \sum_{j=1}^{n} \cosh(\lambda_j) + \sum_{j=1}^{n} \sinh(\lambda_j) \\ &= EE_{even}(G) + EE_{odd}(G). \end{aligned} \tag{1.36}$$

Then, $EE_{even}(G)$ and $EE_{odd}(G)$ represent weighted sums of all odd and even closed walks in the network, respectively, giving more weight to the small ones. It is easy to see that every closed walk of odd length visits the nodes of at least one cycle of odd length. That is, the subgraph associated with an odd closed walk contains at least one odd cycle. However, closed walks of even length can be trivial in the sense that the subgraphs associated with them do not necessarily contain cycles. This simple fact is the foundation of the bipartivity measures based on hyperbolic matrix functions knowing that a bipartite network does not contain any odd-length cycle. Estrada and Rodríguez-Velázquez [44] proposed the following bipartivity measure

$$b_s = \frac{tr\left(\cosh(A)\right)}{tr(e^A)} = \frac{\sum_{j=1}^{n} \cosh(\lambda_j)}{\sum_{j=1}^{n} \exp(\lambda_j)}. \tag{1.37}$$

Another way of accounting for the global bipartivity of a network is to consider the difference of the number of closed walks of even and odd length, and then to normalize the index by the sum of closed walks. That is,

$$b_e = \frac{\sum_{j=1}^{n} \cosh(\lambda_j) - \sum_{j=1}^{n} \sinh(\lambda_j)}{\sum_{j=1}^{n} \cosh(\lambda_j) + \sum_{j=1}^{n} \sinh(\lambda_j)} = \frac{tr\left(\exp(-A)\right)}{tr\left(\exp(A)\right)}. \tag{1.38}$$

In addition, $EE_{odd}(G)$ is proposed by Estrada [37] as a spectral scale method that classifies complex networks into four theoretically possible topological structures.

1.3.4 The Double-Factorial Communicability Function

The new matrix function was proposed by Estrada and Silver [31] by suggesting double-factorial penalization of walks between nodes in a graph as a way to increase the contribution of longer walks in communicability-based functions for graphs and real-world networks. The double factorial of a non-negative integer n is defined by

$$n!! = \begin{cases} n.(n-2).(n-4).\ldots.5.3.1 & n > 0 \text{ odd} \\ n.(n-2).(n-4).\ldots.6.4.2 & n > 0 \text{ even} \\ 1 & n = 0 \end{cases} \tag{1.39}$$

Then, the double-factorial communicability function is defined as

$$f(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!!}, \tag{1.40}$$

which can be approximated by the matrix function:

$$D(A) = \frac{1}{2}\left[\sqrt{2\pi}\tanh\left(\frac{kA}{\sqrt{2}}\right) + 2I\right] e^{\frac{A^2}{2}}, \tag{1.41}$$

where $k = \sqrt{2}\ln(2)$. The study of the double-factorial subgraph centrality, $[D(A)]_{ii}$, showed that in networks where there are no structural holes, the double factorial indices produce similar results as the single-factorial ones in ranking the nodes of the networks while in networks containing significant chordless cycles or holes in their structures, the centrality index based on single- as well as on the double-factorial penalization of the walks produce significant differences in the ranking of the nodes.

1.3.5 More on Adjacency Matrix Functions

The graph energy of a simple undirected graph G is defined as

$$E(G) = \sum_{j=1}^{n} |\lambda_j|. \tag{1.42}$$

Estrada and Benzi [32] provided a structural interpretation of the graph energy using the modulus matrix function. In particular, they prove that the graph energy can be represented as a weighted sum of the traces of even powers of the adjacency matrix. The main result is the following

$$E(G) = tr\,|A| = \lambda_1 \left[\sum_{k=0}^{\infty} \binom{2k}{k} \frac{(-1)^{k+1}}{2^{2k}(2k-1)} \left(\frac{A^2}{\lambda_1^2} - I \right)^k \right]. \tag{1.43}$$

New bounds for the energy of graph obtained in terms of subgraphs contributing to it were obtained using the previous expression of $E(G)$.

Not only the modulus matrix function is used in the literature to obtain new expressions of physical or chemical quantities. The matrix sign function is used to describe the electron density matrix of the Hückel (tight-binding) molecular orbital (HMO) method. Estrada [33] provided an analytical expression of the HMO charge-density matrix as

$$P = P_0 = \frac{1}{2} \left(I + \mathrm{sgn}(A) \right), \tag{1.44}$$

which can be expressed in terms of powers of the adjacency matrix A as

$$P = \frac{1}{2} \left(I + \frac{A}{\lambda_1} \sum_{k=0}^{\infty} (-1)^k \frac{(2k-1)!!}{(2k)!!} \left(\frac{A^2}{\lambda_1^2} - I \right)^k \right). \tag{1.45}$$

In addition, a Euclidean distance based on the HMO charge-density matrix is defined by

$$\delta_{rs}^2 = P_{rr} + P_{ss} - 2P_{rs}, \tag{1.46}$$

which induces an embedding of a molecule into a high-dimensional Euclidean space in which the separation between the atoms scales very well with the bond lengths of

PAHs. The approach is extended to describe a quasi-correlated tight-binding model, which quantifies the number of unpaired electrons and the distribution of effectively unpaired electrons.

1.4 Hückel Molecular Orbital Theory

In the field of chemistry, one way to form molecules is when two or more atoms fused together by sharing electrons and forming covalent bonds between them. The Hückel molecular orbital theory (HMO) provides a simple but neat tool to describe the molecular electronic structures and energies of such molecules based on quantum mechanical models. Our aim here is to use HMO method as a metaphor that allows us to interpret some of the mathematical concepts of graph theory in a chemical way. HMO is only applied to conjugated hydrocarbons, thus we start by giving a brief description of conjugated hydrocarbons and π-electrons which are of a special interest in HMO.

1.4.1 PAHs and π-electrons

Hydrocarbons are organic compounds that consist entirely of hydrogen and carbon atoms. A hydrogen atom has one valence electron (electrons in the outer shell that can form covalent bonds with other atoms) while the carbon atom has four. Thus, in hydrocarbons, hydrogen atoms are always connected to carbon atoms through single bonds while each carbon atom is capable to form four bonds to either other carbon atoms or to hydrogen atoms. Atomic orbitals can be defined to be the space in the atom where the presence of an electron can be predicted. A hydrogen atom has one atomic orbital of 1s-type while a carbon atom has s,p-type atomic orbitals. The outer shell of a carbon atom contains 4 electrons in 2s, 2p atomic orbitals. In order for a carbon atom to make four covalent bonds, the 2s orbital and the three 2p orbitals combine to form equivalent hybrid atomic orbitals: hybrid sp, hybrid sp^2 or hybrid sp^3. What we call a σ-bond is a covalent bond resulted from end-to-end overlap of the s and hybrid atomic orbitals in two different atoms. π-bonds are the result of the side-to-side overlap of p_z-orbitals which are perpendicular to other atomic orbitals which are in the same plane. Figure 1.4 shows examples of end-to-end and side-by-side overlaps. Single

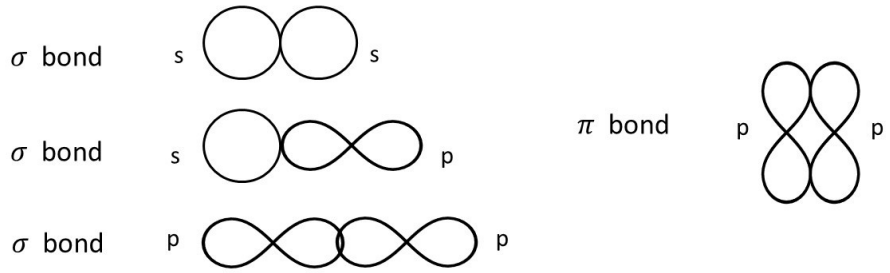


Figure 1.4: Examples of end-to-end bond (left) and side-by-side bond (right)

bonds in hydrocarbon molecules are σ-bonds while double bonds between two carbon atoms are the result of one σ-bond and one π-bond. A conjugated hydrocarbon is a molecule where the single and double bonds between carbon atoms alternates. The benzene molecule C_6H_6 is an example of a conjugated hydrocarbon molecule where the hybrid sp^2 carbon atoms connect to form a planar hexagonal ring with six π-electrons form three π-bonds. Polycyclic aromatic hydrocarbns (PAHs) are planar compounds formed by fusing two or more benzene rings.

1.4.2 Hückel Method

Quantum mechanics is the study of the mechanical properties of very small particles like electrons. The spatial distribution of electrons are expressed as wave functions and the quantum mechanical equations that are used to calculate the properties of such functions are based on a set of fundamental postulates that are given bellow [4, 14]:

1. Any state of a system of n particles (such as molecule) is described by a wave function Ψ, which is a function of the spatial coordinates of the particle and the time. The quantity $\Psi\Psi^* d\tau$ gives the probability of finding the particle in the volume $d\tau$, where Ψ^* is the complex conjugate of Ψ. In this work, wave functions do not contain complex parts, thus $\Psi\Psi^*$ reduces to Ψ^2. Ψ must be everywhere finite, single-valued and Ψ^2 is an integrable function. When

$$\int_0^\infty \Psi^2 d\tau = 1, \tag{1.47}$$

the function Ψ is said to be normalized.

2. For every observable property of a system that is described by some Ψ, there exists a linear Hermitian operator where the physical properties we are interested in can be obtained from the operator and the wave function.

3. In any measurement of the observable associated with an operator $\hat{P}$, the only values that will be observed are the eigenvalues λ, which satisfy the eigenvalue equation

$$\hat{P}\Psi = \lambda\Psi. \tag{1.48}$$

4. The expectation value of the physical quantity associated with the observable $\hat{P}$ is given by

$$\langle\lambda\rangle = \frac{\int \Psi\hat{P}\Psi d\tau}{\int \Psi^2 d\tau}, \tag{1.49}$$

where the integrals are taken over all space.

5. If we consider the total energy E of a system and the associated total energy operator, or the Hamiltonian operator $\hat{\mathcal{H}}$, then the time evolution of a state is determined by the time-dependent Schrödinger equation

$$i\hbar\frac{d\Psi}{dt} = \hat{\mathcal{H}}\Psi, \tag{1.50}$$

where, $\hbar$ is the reduced Planck constant. If the Hamiltonian is independent of time, the energy levels are obtained from the eigenvalue equation

$$\hat{\mathcal{H}}\Psi = E\Psi. \tag{1.51}$$

Since the energies and electron densities are time-independent, we will deal only with equation (1.51).

Molecular orbital theory states that in a molecule, valence electrons are delocalized and they can move in the whole molecule. Each electron is assigned a wave function that determines its properties, Ψ_{MO}, the molecular orbital which according to the LCAO

method is the linear combination of a set of atomic orbitals Φ_{AO}s that are similar in definition of Ψ_{MO} but describe the properties of electrons in the atom. That is, for a system of n electrons, each molecular orbital is expressed as

$$\Psi_j = C_{j1}\Phi_1 + C_{j2}\Phi_2 + \cdots + C_{jn}\Phi_n,\ (j = 1, 2, \cdots, n) \tag{1.52}$$

In a simple molecule like benzene (C_6H_6) we get thirty molecular orbitals expressed in terms of thirty atomic orbitals and it gets more complicated for larger molecules. The HMO method simplifies the equations by considering only the π-electrons in conjugated hydrocarbons. Then, each Ψ_{MO} is the linear combination of $2p_z$-type atomic orbitals. For the Hamiltonian operator of the π-system, the energy levels satisfy the set of equations (Postulate 5)

$$\hat{\mathcal{H}}\psi_j = E_j\psi_j. \tag{1.53}$$

The expected value of E_j is

$$\langle E_j \rangle = \varepsilon_j = \frac{\int \psi_j \hat{\mathcal{H}} \psi_j d\tau}{\int \psi_j^2 d\tau},\ \text{where}\ \psi_j = \sum_{i=1}^{n} C_{ji}\phi_i \tag{1.54}$$

and it is required to minimize $\langle E_j \rangle$ for each j. This is done by minimizing with respect to each coefficient C_{ji} for each value of j. Now, the problem is to find sets of C_{ji} such that

$$\partial \varepsilon_j / \partial C_{ji} = 0 \tag{1.55}$$

for each value of i up to n. Substituting equation 1.52 in equation 1.54 yields

$$\begin{aligned} \varepsilon =& \frac{\int \psi \hat{\mathcal{H}} \psi d\tau}{\int \psi^2 d\tau} \\ =& \frac{\int \left(\sum_{i=1}^{n} C_i\phi\right) \hat{\mathcal{H}} \left(\sum_{i=1}^{n} C_i\phi\right) d\tau}{\int \left(\sum_{i=1}^{n} C_i\phi\right)^2 d\tau}. \end{aligned} \tag{1.56}$$

Recall that the Hamiltonian is a linear operator, thus we have

$$\begin{aligned}\varepsilon =& \frac{\sum_{i=1}^{n} C_i \int \phi_i \hat{\mathcal{H}} \phi_i d\tau}{\sum_{i=1}^{n} C_i \int (\phi_i)^2 d\tau} \\ =& \frac{\sum_{i=1}^{n} \sum_{k=1}^{n} C_i C_k \int \phi_i \hat{\mathcal{H}} \phi_k d\tau}{\sum_{i=1}^{n} \sum_{k=1}^{n} C_i C_k \int \phi_i \phi_k d\tau}.\end{aligned} \tag{1.57}$$

Putting $H_{ik} = \int \phi_i \hat{\mathcal{H}} \phi_k d\tau$ and $S_{ik} = \int \phi_i \phi_k d\tau$ and arranging equation 1.57

$$\sum_{i=1}^{n} \sum_{k=1}^{n} C_i C_k H_{ik} - \left(\sum_{i=1}^{n} \sum_{k=1}^{n} C_i C_k S_{ik} \right) \varepsilon = 0. \tag{1.58}$$

Now, differentiating equation 1.58 with respect to each C_i produces the following linear system of n equations

$$\sum_{i=1}^{n} C_i \left(H_{ji} - \varepsilon S_{ji} \right) = 0,\ j = 1, \cdots, n, \tag{1.59}$$

which can be represented in matrix notation

$$\left(H - \varepsilon S \right) C = 0. \tag{1.60}$$

The matrix elements H_{ii} are called coulomb integrals and for all carbon atoms that are part of the π-system are set equal to some particular numerical value α. The off-diagonal matrix elements H_{ij} are called bond (or resonance) integrals. These are set equal to some particular numerical value β if atoms i and j are directly σ-bonded to each other and set equal to zero otherwise. The terms S_{ij} are called overlap integrals and are set equal to unity if $i = j$, otherwise they are set equal to zero. equation 1.60 becomes

$$\left(\beta A + \alpha I \right) C = \varepsilon C, \tag{1.61}$$

where A is the adjacency matrix of the corresponding molecular graph where the nodes are the carbon atoms and the edges are the σ-bonds between them.

1.4.3 HOMO and LUMO

As seen in the previous section, for a conjugated hydrocarbon with n carbon atoms, we can construct n molecular orbitals containing n π-electrons. we distribute the electrons over the molecular orbitals starting with the lowest-energy (least eigenvalue) orbital and obeying the Pauli principle; that is, placing only two electrons with opposite spin in each orbital. The highest occupied molecular orbital (HOMO) is the highest energy orbital that has electrons in it and the lowest unoccupied molecular orbital (LUMO) is the lowest energy orbital that is empty. These frontier orbitals are the most important orbitals with respect to reactivity.

Chapter 2

Gaussian Communicability Function e^{-A^2}

We consider a network $G = (V, E)$ whose adjacency matrix is designated by A. Then, the spectrum of the graph satisfies the following equation:

$$A\mathbf{q}_j = \lambda_j \mathbf{q}_j, \tag{2.1}$$

where $\mathbf{q}_j$ is the eigenvector associated to the eigenvalue λ_j.
Then, we consider an approach known as the *Folded Spectrum Method* (FSM) [57] whose main idea is the following. An eigensolution $(\lambda_j, \mathbf{q}_j)$ of (2.1) also satisfies

$$(A - \lambda_{\text{ref}} I)^2 \mathbf{q}_j(\mathbf{r}) = (\lambda_j - \lambda_{\text{ref}})^2 \mathbf{q}_j(\mathbf{r}), \tag{2.2}$$

where λ_{ref} is a given eigenvalue. If we select $\lambda_{\text{ref}} = 0$, we have

$$A^2 \mathbf{q}_j(\mathbf{r}) = \lambda_j^2 \mathbf{q}_j(\mathbf{r}). \tag{2.3}$$

This process of 'folding' the spectrum of a network is illustrated in Figure 2.1, which is inspired by a similar one displayed by Canning et al [58]. In the left-hand side of the Figure we show the standard spectrum of the network, in which the eigenvalues are represented on the line as dots. We then squared the eigenvalues and represent them

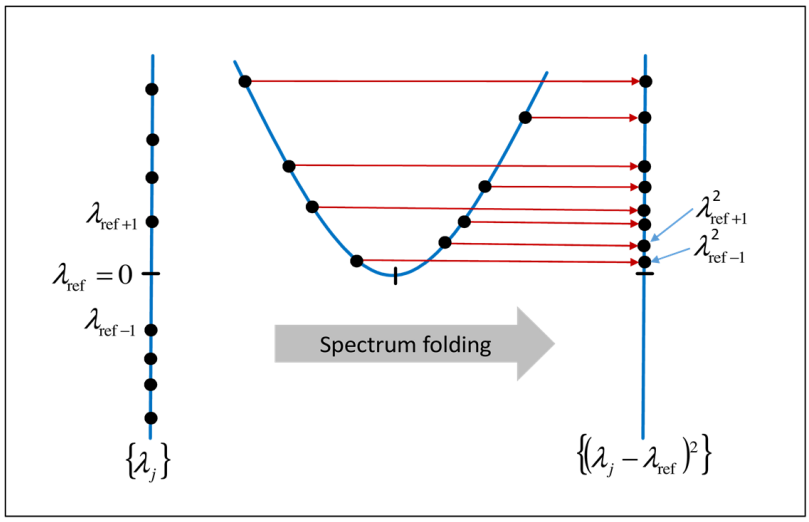


Figure 2.1: Illustration of the folded spectrum method. The eigenvalues of the adjacency matrix of the network are folded at λ_{ref} into the spectrum of $(A - \lambda_{\text{ref}})^2$. The eigenvalues closest to zero are now the lowest states in the spectrum of $(A - \lambda_{\text{ref}})^2$.

in a new line on the right-hand side of the Figure. The first consequence of this folding process is that now all the eigenvalues are nonnegative, with the smallest ones being the zeros of the network spectrum. As can be seen in the Figure, the eigenvalues closest to zero, here designated as $\lambda_{\text{ref}\pm1}$, becomes the lowest state in the spectrum of A^2. As we will see later this will imply that we can obtain important structural information about the network which is encoded by these eigenvalues of the adjacency matrix. In this Thesis we will call

$$\tilde{G} = \sum_{k=0}^{\infty} \frac{\left(-A^2\right)^k}{k!} = \exp(-A^2) \tag{2.4}$$

the Gaussian matrix function of A, $\tilde{G}_{pq}$ the Gaussian communicability function between the nodes p and q based on $-A^2$ and

$$H = \text{Tr}(\exp(-A^2)) \tag{2.5}$$

the Gaussian Estrada index of the graph. The term $\tilde{G}_{pp}$ is the Gaussian subgraph centrality based on the same matrix function.

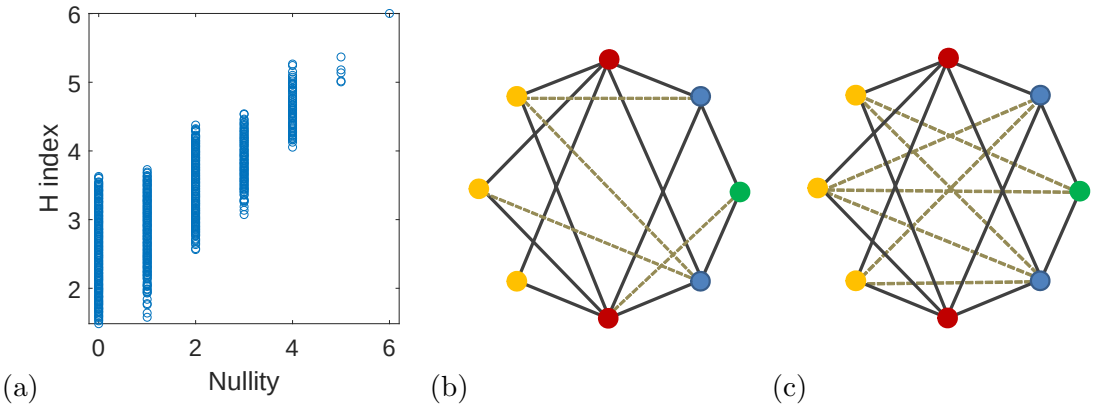


Figure 2.2: (a) Plot of graph nullity versus H index for all connected graphs with 8 nodes. (b) Graph with the largest H index among all the connected graphs with 8 nodes and nullity zero. (c) The same as in (b) for all connected graphs with nullity one.

2.1 Mathematical Analysis

Using the spectral decomposition of the adjacency matrix we can express the Gaussian communicability function and the Gaussian Estrada index as

$$\tilde{G}_{pq} = \sum_{j=1}^{n} \psi_{j,p} \psi_{j,q} \exp\left(-\lambda_j^2\right), \tag{2.6}$$

$$H = \sum_{j=1}^{n} \exp\left(-\lambda_j^2\right). \tag{2.7}$$

Let $\eta(A)$ be the nullity of the adjacency matrix A. Then, it is obvious that the H index is related to η as follows:

$$H \geq \eta, \tag{2.8}$$

with both indices identical if and only if $\lambda_j = 0$, for all j, which is attained only for the trivial graph, i.e., the graph with n nodes and no edges. Indeed,

$$H = \eta + \sum_{\lambda_j \neq 0} \exp\left(-\lambda_j^2\right). \tag{2.9}$$

Then, it is interesting at least empirically, to explore the relation between H and η for simple graphs. We investigate all the connected graphs with $n \leq 8$ for which we obtain both H and η. The correlation between both indices for the 11,117 connected

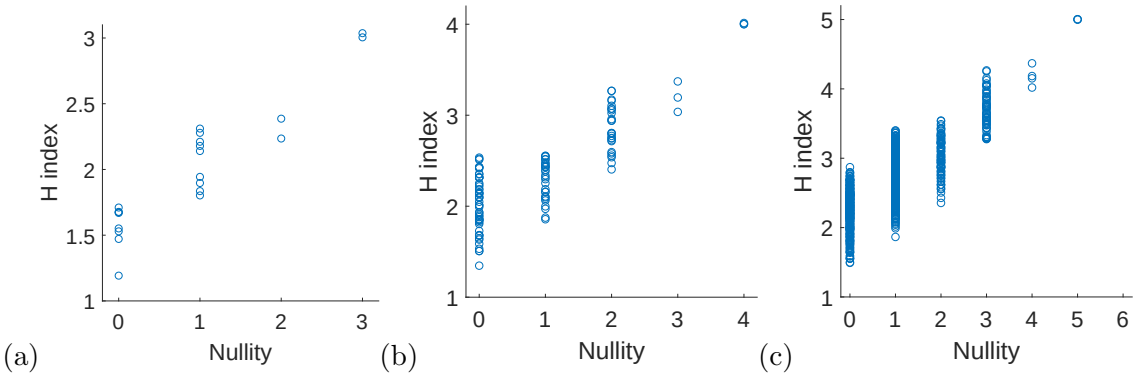


Figure 2.3: Plot of graph nullity versus H index for all connected graphs with (a) 5 nodes, (b) 6 nodes, (c) 7 nodes.

graphs with 8 nodes is illustrated in Figure 2.2. Although the correlation is statistically significant—the Pearson correlation coefficient is 0.74—it hides the important differences between the two indices. For instance, there are 5,724 graphs with zero nullity among all the connected graphs with 8 nodes. For these graphs $1.484 \leq H \leq 3.629$, which represents a wide range of values taking into account that the minimum and maximum values of H for all connected graphs with 8 nodes are 1.484 and 6, respectively. It is also easy to see that there are graphs having nullity zero which have larger H indices than some graphs having nullity one, two or three. The results are very similar for $n < 8$ and they are shown in Figure 2.3. In Figure 2.2 we show the graphs with the largest H indices among all connected graphs with 8 nodes and nullity zero or one. These graphs show a common pattern containing several complete bipartite subgraphs. For instance, every yellow node in the Figure 2.2 is connected to every red ones, every red is connected to every blue and every blue is connected to the green one, while there is no yellow-yellow, red-red or blue-blue connections. This pattern will be revealed when we study the mathematical properties of this index and its importance will be analyzed for real-world networks.

2.2 H Index of Graphs

2.2.1 Elementary Properties

In the following we show some results about $\tilde{G}_{pq}$ of some elementary graphs which will help us to interpret this measure when applied to more complex structures. In particular, we study the path P_n, the cycle C_n, the star graph $K_{1,n-1}$, the complete graph K_n and the complete bipartite graph K_{n_1,n_2}. Here we give expressions for the $H(G)$ index of the before mentioned graphs in the form of Lemmas.

Lemma 2.2.1. Let K_n be the complete graph of n nodes. Then

$$H(K_n) = e^{-(n-1)^2} + \frac{n-1}{e}. \tag{2.10}$$

Proof. The spectrum of K_n is $\sigma(K_n) = \left\{[n-1]^1, [-1]^{n-1}\right\}$ with the eigenvector $\mathbf{q}_1 = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)$ so we have

$$\tilde{G}_{pq}(K_n) = \mathbf{q}_1(p)\mathbf{q}_1(q)e^{-(n-1)^2} + \sum_{j=2}^{n} \mathbf{q}_j(p)\mathbf{q}_j(q)e^{-1}, \tag{2.11}$$

and since the eigenvector matrix has orthonormal rows and columns we have $\sum_{j=2}^{n} \mathbf{q}_j(p)\mathbf{q}_j(q) = -\frac{1}{n}$ if $p \neq q$ and $\frac{n-1}{n}$ if $p = q$. Thus,

$$\tilde{G}_{pq}(K_n) = \frac{e^{-(n-1)^2}}{n} - \frac{1}{ne} \tag{2.12}$$

when $p \neq q$. Now, if $p = q$ then $\tilde{G}_{pp}(K_n) = \mathbf{q}_1^2(p)e^{-(n-1)^2} + \sum_{j=2}^{n} \mathbf{q}_j^2(p)e^{-1} = \frac{e^{-(n-1)^2}}{n} + \frac{n-1}{ne}$. Then, it is straightforward to realize that

$$\begin{aligned} H(K_n) &= \sum_{j=1}^{n} \left(\frac{e^{-(n-1)^2}}{n} + \frac{n-1}{ne}\right) \\ &= e^{-(n-1)^2} + \frac{n-1}{e}. \end{aligned} \tag{2.13}$$

□

Let

$$I_\gamma(x) = \frac{1}{\pi}\int_0^\pi \cos(\gamma\theta)\exp(x\cos\theta)d\theta - \frac{\sin(\gamma\pi)}{\pi}\int_0^\infty \exp(-x\cosh t - \gamma t)dt, \quad (2.14)$$

be the modified Bessel function of the first kind. Then, we have the following result.

Lemma 2.2.2. Let P_n be a path having n nodes. Then, asymptotically as $n \to \infty$

$$H\left(P_n\right) \sim \frac{I_0(2)}{e^2}(n+1) - e^{-4}. \quad (2.15)$$

Proof. The eigenvalues of P_n are: $2\cos\frac{\pi j}{n+1}$, $j = 1, \ldots, n$. By substituting the eigenvalues and eigenvectors of the path graph into the expression for $\tilde{G}_{pp}\left(P_n\right)$ we obtain

$$\begin{aligned}\tilde{G}_{pp}\left(P_n\right) &= \frac{2}{n+1}\sum_{j=1}^{n}\sin^2\left(\frac{j\pi p}{n+1}\right)\exp\left(-4\cos^2\left(\frac{j\pi}{n+1}\right)\right) \\ &= \frac{e^{-2}}{n+1}\sum_{j=1}^{n}\left[1-\cos\left(\frac{2j\pi p}{n+1}\right)\right]\exp\left(-2\cos\left(\frac{2j\pi}{n+1}\right)\right). \end{aligned} \quad (2.16)$$

Now, when $n \to \infty$ the summation in 2.16 can be approached by the following integral

$$\tilde{G}_{pp}\left(P_n\right) = \frac{e^{-2}}{\pi}\int_0^\pi \exp(-2\cos\theta)d\theta - \frac{e^{-2}}{\pi}\int_0^\pi \cos\left(p\theta\right)\exp(-2\cos\theta)d\theta, \quad (2.17)$$

where $\theta = \frac{2j\pi}{n+1}$. Thus, when $n \to \infty$ we have

$$\tilde{G}_{pp}\left(P_n\right) \sim e^{-2}\left(I_0(-2) - I_p(-2)\right), \quad (2.18)$$

which by using $I_\gamma(-x) = (-1)^\gamma I_\gamma(x)$ [59] gives

$$\tilde{G}_{pp}\left(P_n\right) \sim e^{-2}\left(I_0(2) - (-1)^p I_p(2)\right).$$

Let n be even. Then due to the symmetry of the path we have

$$\begin{aligned} H\left(P_n\right) &= 2\sum_{p=1}^{n/2} \tilde{G}_{pp}\left(P_n\right) \\ &= 2\sum_{p=1}^{n/2} e^{-2}\left[I_0(2) - (-1)^p I_p(2)\right] \\ &= \frac{nI_0(2)}{e^2} - \frac{2}{e^2}\sum_{p=1}^{n/2}(-1)^p I_p(2). \end{aligned} \tag{2.19}$$

For $n \to \infty$ we have

$$\sum_{\gamma=1}^{\infty}(-1)^\gamma I_\gamma(x) = \frac{1}{2}\left(e^{-x} - I_0(x)\right). \tag{2.20}$$

Then, we can write for $n \to \infty$

$$\begin{aligned} H\left(P_n\right) &\sim \frac{nI_0(2)}{e^2} - \frac{1}{e^2}\left(e^{-2} - I_0(2)\right) \\ &= \frac{I_0(2)}{e^2}(n+1) - e^{-4}. \end{aligned} \tag{2.21}$$

Now, when n is odd we can split the path into two paths of lengths $\frac{n+1}{2}$ and $\frac{n-1}{2}$, respectively. Then, we write

$$\begin{aligned} H\left(P_n\right) &= \sum_{p=1}^{\frac{n+1}{2}} \tilde{G}_{pp}\left(P_n\right) + \sum_{p=\frac{n-1}{2}}^{n} \tilde{G}_{pp}\left(P_n\right) \\ &= \frac{(n+1)I_0(2)}{2e^2} - \frac{1}{e^2}\sum_{p=1}^{\frac{n+1}{2}}(-1)^p I_p(2) + \frac{(n-1)I_0(2)}{2e^2} - \frac{1}{e^2}\sum_{p=\frac{n-1}{2}}^{n}(-1)^p I_p(2). \end{aligned} \tag{2.22}$$

When $n \to \infty$ we can consider that the summation in the second and fourth terms of 2.22 are both equal to $\left(e^{-2} - I_0(2)\right)/2$, which then gives the final result. □

Lemma 2.2.3. Let C_n be a cycle having n nodes. Then, asymptotically as $n \to \infty$

$$H(C_n) \sim \frac{nI_0(-2)}{e^2}. \tag{2.23}$$

Proof. The eigenvalues of C_n are: $2\cos\frac{2\pi j}{n}$, $j = 1, \ldots, n$. By substituting the eigenvalues of the cycle graph into the expression for $H(C_n)$ we obtain

$$\begin{aligned} H(C_n) &= tr\left(e^{-A^2}\right) \\ &= n\left(\frac{tr(e^{-A^2})}{n}\right) \\ &= n\left(\frac{1}{n}\sum_{j=1}^{n} e^{-4\cos^2(\frac{2\pi j}{n})}\right) \\ &= ne^{-2}\left(\sum_{j=1}^{n}\frac{1}{n}e^{-2\cos\frac{4\pi j}{n}}\right). \end{aligned} \tag{2.24}$$

Now, when $n \to \infty$ the summation in 2.24 can be approached by the following integral

$$H(C_n) \sim ne^{-2}\frac{1}{\pi}\int_0^{\pi} e^{-2cos\theta}d\theta, \tag{2.25}$$

where $\theta = \frac{2j\pi}{n}$. Thus, when $n \to \infty$ we have

$$H(C_n) \sim ne^{-2}I_0(-2). \tag{2.26}$$

□

Lemma 2.2.4. Let K_{n_1,n_2} be the complete bipartite graph of $n_1 + n_2$ nodes. Then

$$H\left(K_{n_1,n_2}\right) = 2e^{-n_1n_2} + n_1 + n_2 - 2. \tag{2.27}$$

Proof. The spectrum of K_{n_1,n_2} is $\sigma(K_{n_1,2}) = \left\{\left[\sqrt{n_1n_2}\right]^1, [0]^{n_1+n_2-2}, \left[\sqrt{n_1n_2}\right]^1\right\}$ and the eigenvectors associated with the largest and the smallest eigenvalues are, respectively:

$$\varphi_1 = \frac{1}{\sqrt{2n_2}}\left(\frac{\sqrt{n_1n_2}}{n_1}, \frac{\sqrt{n_1n_2}}{n_1}, \ldots, \frac{\sqrt{n_1n_2}}{n_1}, 1, \ldots, 1\right) \tag{2.28}$$

where the first n_1 components are $\frac{\sqrt{n_1n_2}}{n_1}$ and the last n_2 components are 1. similarily:

$$\varphi_{n_1+n_2} = \frac{1}{\sqrt{2n_2}}\left(\frac{-\sqrt{n_1n_2}}{n_1}, \ldots, \frac{-\sqrt{n_1n_2}}{n_1}, , 1, \ldots, 1\right). \tag{2.29}$$

From the orthonormality of the eigenvectors of the adjacency matrix we have:

$$\sum_{j=2}^{n_1+n_2-1} \left[\mathbf{q}_j\left(p\right)\right]^2 = 1 - \frac{1}{n_1},\ p \in V_1, \tag{2.30}$$

$$\sum_{j=2}^{n_1+n_2-1} \left[\mathbf{q}_j\left(p\right)\right]^2 = 1 - \frac{1}{n_2},\ p \in V_2. \tag{2.31}$$

Hence, if $p \in V_1$

$$\begin{aligned}
\tilde{G}_{pp}\left(K_{n_1,n_2}\right) &= \sum_{j=1}^{n_1+n_2} \left[\mathbf{q}_j\left(p\right)\right]^2 \exp(-\lambda_j^2) \\
&= e^{-n_1 n_2}\left(\frac{n_1 n_2}{2n_1 n_2^2} + \frac{n_1 n_2}{2n_2 n_1^2}\right) + \sum_{j=2}^{n_1+n_2-1} \left[\mathbf{q}_j\left(p\right)\right]^2 \\
&= e^{-n_1 n_2}\left(\frac{1}{n_1}\right) + 1 - \frac{1}{n_1} = \frac{1}{n_1}\left(e^{-n_1 n_2} - 1\right) + 1,
\end{aligned} \tag{2.32}$$

and similarly we have $\tilde{G}_{pp}\left(K_{n_1,n_2}\right) = \frac{1}{n_2}(e^{-n_1 n_2} - 1) + 1$ when $p \in V_2$. Then

$$\begin{aligned}
H\left(K_{n_1,n_2}\right) &= \sum_{j=1}^{n_1+n_2} \tilde{G}_{pp} \\
&= \sum_{j=1}^{n_1} \tilde{G}_{pp} + \sum_{j=m+1}^{n_1+n_2} \tilde{G}_{pp} \\
&= n_1\left(\frac{1}{n_1}(e^{-n_1 n_2} - 1) + 1\right) + n_2\left(\frac{1}{n_2}(e^{-n_1 n_2} - 1) + 1\right) \\
&= 2e^{-n_1 n_2} + n_1 + n_2 - 2.
\end{aligned} \tag{2.33}$$

□

The following corollary will be of importance in the following section of this work.

Corollary 2.2.1. Let $K_{1,n-1}$ be the star graph of n nodes. Then

$$H\left(K_{1,n-1}\right) = 2e^{1-n} + n - 2. \tag{2.34}$$

2.2.2 Graphs with Maximum H Index

We are mainly interested in understanding why certain networks display large values of the H index. Then, we prove that among the connected graphs with n nodes, the maximum value of the H index is always obtained for the star graph $K_{1,n-1}$. We start this section by obtaining an upper bound of H, then we prove a general result for trees, which is needed to prove the upper bound. Michele Benzi, Gene H. Golub and Paola Boito [60, 61] obtained bounds for the entries of matrix function $f(A)$ by representing them in terms of Riemann–Stieltjes integrals and by approximating such integrals by Gaussian quadrature rules. Consider an interval $[a, b]$ contains the eigenvalues of A, then if $\varphi(x, y) = \frac{(A)_{ii}(f(x)-f(y))+xf(y)-yf(x)}{x-y}$ and f is strictly completely monotonic function on the interval $[a, b]$, that is, $f^{(2j)}(x) > 0$ and $f^{(2j+1)} < 0$ on $[a, b]$ where $f^{(k)}$ denotes the kth derivative of f and $f^{(0)} \equiv f$ then [60]

$$\begin{aligned}(f(A))_{ii} &\leq \varphi(a, b) \\ &\leq \frac{(A)_{ii}\,(f(a) - f(b)) + af(b) - bf(a)}{a - b}.\end{aligned} \tag{2.35}$$

The following theorem gives an upper bound of the H index.

Theorem 2.2.1. Let G be a graph with n nodes and m edges and let $H = tr \exp\left(-A^2\right)$. Then,

$$H\,(G) \leq \sum_{i=1}^{n} \left[\frac{k_i(e^{-b} - 1)}{b} + 1\right] = 2m\frac{(e^{-b} - 1)}{b} + n. \tag{2.36}$$

where k_i is the degree of the node i in the graph G and $b \in \mathbb{R}$ such that $\lambda_1^2 \leq b$.

Proof. If we consider the matrix A^2, where A is the adjacency matrix of a graph G and $f(x) = e^{-x}$ we have $\omega_1 = k_i$ and $[a, b] = [0, b]$ since A^2 has nonnegative eigenvalues. Hence,

$$(e^{-A^2})_{ii} \leq \frac{k_i(1 - e^{-b}) - b}{-b} = \frac{k_i(e^{-b} - 1)}{b} + 1.$$

To find the bound of the trace of e^{-A^2} we take the summation from 1 to n on the previous inequality, which by Lemma 1.1.1 gives the final result. □

Lemma 2.2.5. Let T_n be a tree of n nodes, then

$$H(T_n) \leq H(K_{1,n-1}). \tag{2.37}$$

Proof. We have the following upper bound

$$H(G) \leq 2m\frac{(e^{-\lambda_1^2}-1)}{\lambda_1^2} + n, \tag{2.38}$$

Perron-Frobenius theorem states that $\lambda_1 \geq |\lambda_i|$ for all $i = 1, \ldots, n-1$ where m is the number of edges and $[0, b]$ is the interval that contains all the eigenvalues of A^2. Let λ_1 be the eigenvalue with the maximum absolute value among all eigenvalues (Perron-Frobenius theorem). Now, Let T_n be a tree with $n \geq 2$, then Collatz and Sinogowitz [1] have proved that

$$\lambda_1(T_n) \leq \lambda_1(K_{1,n-1}) = \sqrt{n-1}, \tag{2.39}$$

where the equality holds if T_n is the star graph. Thus, the interval $[0, n-1]$ contains all the eigenvalues of any tree T_n. Now, substituting in (2.38)

$$H(T_n) \leq 2(n-1)\frac{(e^{1-n}-1)}{n-1} + n = n - 2 + 2e^{1-n}. \tag{2.40}$$

Thus, for any tree of n nodes $H(T_n) \leq H(K_{1,n-1})$. □

Now we prove an important result for general connected graphs, which also allow us to understand the nature of the index H when studying real-world networks.

Theorem 2.2.2. Let G be connected graph of n nodes, then

$$H(G) \leq H(K_{1,n-1}). \tag{2.41}$$

Proof. The largest eigenvalue of any graph G is less than or equal the maximum degree. Thus the interval $[0, (n-1)^2]$ contains all the eigenvalues of A^2 and we get from the quadrature-rule bound

$$H(G) \leq n - 2m\frac{(1-e^{-(n-1)^2})}{(n-1)^2}. \tag{2.42}$$

Now, $H(G)$ is maximum when m is the lowest possible for a connected graph. That is,

$$H(G) \leq n - 2\frac{(1 - e^{-(n-1)^2})}{n-1}. \tag{2.43}$$

A connected graph with $n-1$ edges is a tree. Then, because of Lemma 2.2.5 we have

$$H(G) \leq H(K_{1,n-1}). \tag{2.44}$$

□

The previous result is true for any connected graph with any (finite) number of nodes n. In addition, when $n \to \infty$, it is easy to see that $H\left(K_{1,n-1}\right) \sim n-2$. In a similar way, when $n \to \infty$

$$H\left(K_{n_1,n_2}\right) \sim n_1 + n_2 - 2 = n - 2. \tag{2.45}$$

Among the connected graphs with n nodes, as proved here, the maximum value is always reached for the star graph $K_{1,n-1}$. Moreover, for all connected graphs of $n \leq 8$ nodes, it is followed by the complete bipartite graph $K_{2,n-2}$, then $K_{3,n-3}$, and so forth. For instance, in the case $n = 8$ we have $H\left(K_{1,7}\right) \approx 6.001824$; $H\left(K_{2,6}\right) \approx 6.000012$; $H\left(K_{3,5}\right) \approx 6.000001$; $H\left(K_{4,4}\right) \approx 6.000000$. This observation will play a fundamental role in the analysis of random graphs and real-world networks in the next sections of this Chapter.

2.2.3 Graphs with Minimum H Index

As we have seen before, an important contribution to the H index is made by the graph nullity η and by the eigenvalues which are relatively close to zero. Let $x > 0$ be a real number such that $\exp\left(-x^2\right) \sim 0$. Then,

$$H \approx \sum_{\lambda_j \geq -x}^{\lambda_j \leq x} \exp\left(-\lambda_j^2\right). \tag{2.46}$$

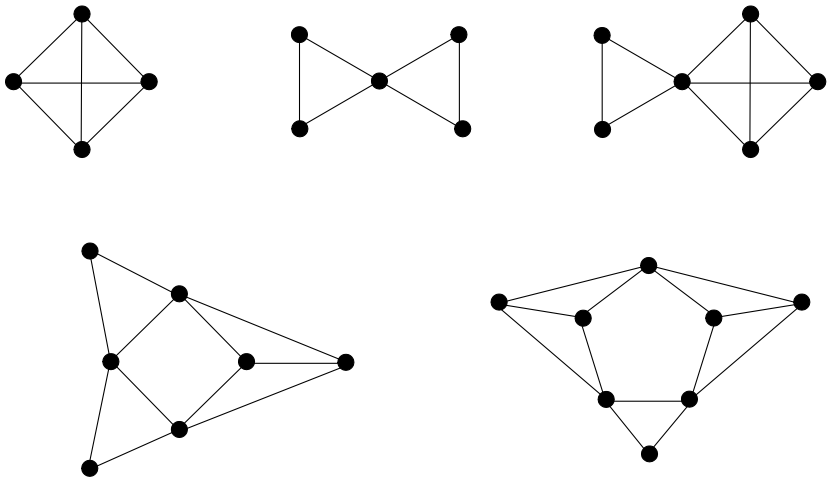


Figure 2.4: Illustration of the graphs having minimum H index among all connected graphs with $n = 4, 5, 6, 7, 8$.

Consequently, the graphs with minimum H index are those having very small density of eigenvalues in the interval $(-x, x)$. For instance, the graph having the smallest H index among all connected graphs with 8 nodes has eigenvalues: -2.0000, -1.7321, -1.0000, -1.0000, -0.8136, 1.4707, 1.7321, 3.3429, which produces $H \approx 1.4845$, which is well approximated if we consider only the eigenvalues in the interval $(-1.5, 1.5)$. The graphs with minimum H index among all connected graphs with $n = 4, 5, 6, 7, 8$ are illustrated in the Figure 2.4. As can be seen, there is no specific description of the graphs but it calls the attention the existence of bow-tie subgraphs in most of these graphs.

2.2.4 H Index of Random Networks

In this section we study two different models of random graphs, the Erdős-Rényi model of random network and the Barabási-Albert model. Using the previous mentioned distributions of the spectrum of these models we obtain the following results.

Theorem 2.2.3. For an Erdős-Rényi random graph $G(n, p)$ with $\frac{\ln n}{n} \ll p$ we have

$$H\left(ER\right) \sim ne^{\frac{-r^2}{2}}(I_0(\frac{r^2}{2}) + I_1(\frac{r^2}{2})) \tag{2.47}$$

almost surely, as $n \to \infty$, where $r = 2\sqrt{np(1-p)}$ and I_n is the modified Bessel function

of the first kind.

Proof. We know that the spectral density of $G(n,p)$ converges to the semicircular distribution (1.2) as $n \to \infty$. Also, Krivelevich and Sudakov [62] showed that the largest eigenvalue λ_1 of $G(n,p)$ is almost surely $(1+o(1))np$ provided that $np \gg \ln n$. Then,

$$\begin{aligned} H\left(ER\right) &= \exp(-\lambda_1^2) + \sum_{i=2}^{n} \exp(-\lambda_i^2) \\ &= e^{-\lambda_1^2} + n\left(\frac{1}{n}\sum_{i=2}^{n} e^{-\lambda_i^2}\rho(\lambda)\right). \end{aligned} \tag{2.48}$$

When $n \to \infty$ we have

$$\begin{aligned} H\left(ER\right) &\sim n\int_{-r}^{r} \rho(\lambda)e^{-\lambda^2}d\lambda \\ &= \frac{4n}{\pi r^2}\int_0^r \sqrt{r^2-\lambda^2}e^{-\lambda^2}d\lambda \\ &= \frac{4n}{\pi r^2}\int_0^{\frac{\pi}{2}} r^2\cos^2\theta e^{-r^2\sin^2\theta}d\theta \\ &= \frac{4n}{\pi}\int_0^{\frac{\pi}{2}} \frac{1}{2}(1+\cos 2\theta)e^{\frac{-r^2}{2}(1-\cos 2\theta)}d\theta \\ &= 2ne^{\frac{-r^2}{2}}(\frac{1}{\pi}\int_0^{\frac{\pi}{2}} e^{\frac{r^2}{2}\cos 2\theta}d\theta + \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \cos 2\theta e^{\frac{r^2}{2}\cos 2\theta}d\theta) \\ &= ne^{\frac{-r^2}{2}}(\frac{1}{\pi}\int_0^{\pi} e^{\frac{r^2}{2}\cos u}du + \frac{1}{\pi}\int_0^{\pi} \cos u e^{\frac{r^2}{2}\cos u}du) \\ &= ne^{\frac{-r^2}{2}}(I_0(\frac{r^2}{2}) + I_1(\frac{r^2}{2})). \end{aligned} \tag{2.49}$$

□

We now consider the case of the Barabási-Albert (BA) model. We prove the following result.

Theorem 2.2.4. Let G be a BA random network. Then, when $n \to \infty$, the H index of a BA network is bounded as

$$H\left(BA\right) \sim \frac{n}{r^2}\left(\sqrt{\pi}r\mathrm{erf}\left(r\right) + e^{-r^2} - 1\right), \tag{2.50}$$

where $r = 2\sqrt{np(1-p)}$ and $\mathrm{erf}\,(\cdots)$ is the error function.

Proof. We know that the density of BA graphs follows a triangular distribution (1.3). Thus

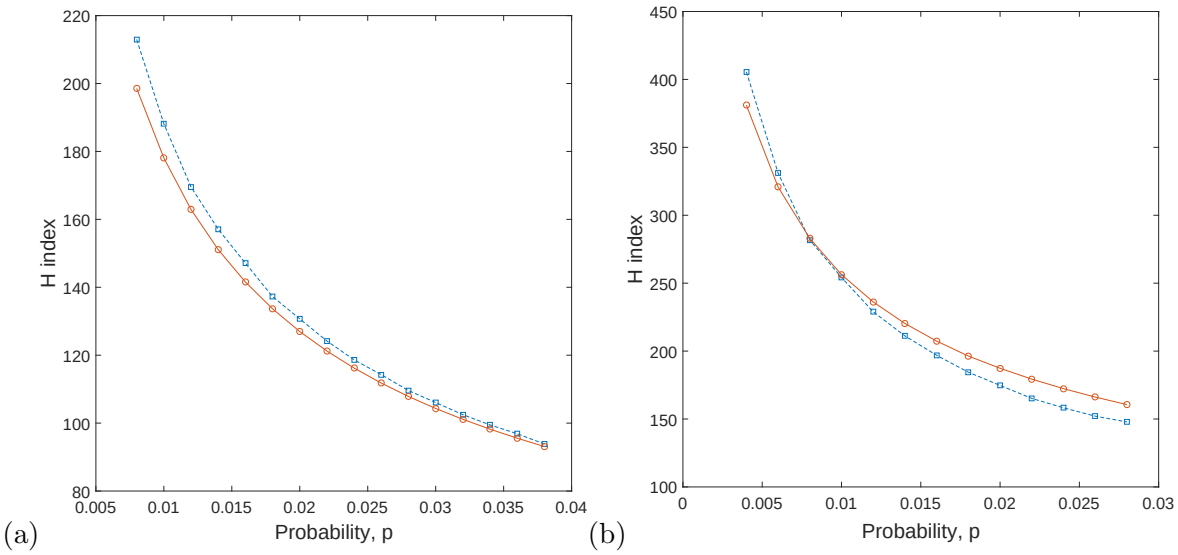
$$\begin{aligned} H\,(BA) &= \sum_{j=1}^{n} \rho(\lambda_j) e^{-\lambda_j^2} \\ &= n \left(\frac{1}{n} \sum_{j=1}^{n} \rho(\lambda_j) e^{-\lambda_j^2} \right). \end{aligned} \tag{2.51}$$

When $n \to \infty$ we have

$$\begin{aligned} H\,(BA) &\sim n \left(\int_{-r}^{r} \rho(\lambda) e^{-\lambda^2} d\lambda,\ as\, n \to \infty \right) \\ &= n \left(\int_{-r}^{0} \frac{\lambda + r}{r^2} e^{-\lambda^2} d\lambda + \int_{0}^{r} \frac{r - \lambda}{r^2} e^{-\lambda^2} d\lambda \right) \\ &= \frac{n}{r^2} \left(\sqrt{\pi} r \mathrm{erf}\,(r) + e^{-r^2} - 1 \right). \end{aligned} \tag{2.52}$$

□

In Figure 2.5(a) we illustrate the results obtained for the H index of ER random graphs $G_{ER}\,(1000, p)$ in which p is systematically changed from 0.008 to 0.04. The results are shown for both, the formula (2.47) and the calculation using the function 'expm' implemented in Matlab®. As can be seen for ER networks, as soon as the probability increases, such that $np \gg \ln n$, the two results quickly converge to a common value, i.e., the error decays quickly with the increase of p. In Figure (2.5(b)) we also plot similar results for the BA model using $G_{BA}\,(1000, m) = G_{BA}\,(1000, (n-1)\,p/2)$ in which m is systematically varied from 2 to 14. In this case we plot the results using the values of p instead of m to make the plot comparable to the one of the ER networks. In this case the behavior is more complex as there is a crossing point between the two curves. This difference between the behavior of the theoretical function (2.50) for low and large densities of the graphs may be due to the fact that the eigenvalue distribution of the BA networks is different at these two density regimes. According to our computational experiments, it is only true that the BA networks display triangular



(a) (b)

Figure 2.5: (a) Change of the H index with the increase of the probability p in ER random graphs $G_{ER}(1000, p)$ obtained using the formula 2.47 (empty circles and solid line) and using the function 'expm' in Matlab (squares and broken line). (b) Change of the H index with the increase of m in BA random graphs $G_{BA}(1000, m)$ obtained using the formula (2.50) (empty circles and solid line) and using the function 'expm' in Matlab (squares and broken line). All the calculations are the average of 100 random realizations.

eigenvalue distributions for relatively small edge densities and deformations of it occurs for larger densities, which may produce the observed deviations from the theoretical and computational results. In Appendix A, we show the eigenvalue distribution of samples of networks generated using the BA model for different values of m and it shows that the distribution does not follow equation (2.50) as m grows. More theoretical work is needed to understand completely the eigenvalue distribution of these networks at different density regimes. Such studies are out of the scope of the Thesis.

It is easy to show that for a given value of r, $H(BA) > H(ER)$. That is, for the same network density the network having power-law degree distribution has larger value of the H index than the analogous one with Poisson degree distribution. This result is somehow expected from the qualitative analysis of the eigenvalues distributions of these two classes of random networks. While the ER networks display a semicircle distribution of eigenvalues, the BA networks for small values of r displays a triangular distribution peaked at $\lambda_j = 0$. In other words, the nullity of the BA graphs is larger than that of the ER ones, and the concentration of eigenvalues close to zero is also

larger for the BA networks than for the ER. Both characteristics give rise to larger values of the H index in the BA networks. The question that arises here is what this difference implies from the structural point of view. We will analyze this question in the remaining part of this section.

We have already seen that the largest values of the H index occurs in graphs having complete bipartite structures. Then, in order to understand the main structural differences giving rise to the larger H index in BA networks than in ER ones we consider the existence of such subgraphs in both networks. In particular, we will consider the existence of complete bipartite subgraphs, known as bicliques, in both kind of networks. Let us start by the analysis of the BA networks. These networks are created from an initial seed of m_0 nodes connected randomly and independently according to the ER model. Then, at each stage of the evolution of the network, a new node is connected preferentially to $m \leq m_0$ nodes. The connection probability is proportional to the degree of the existing nodes. Because an ER network is uncorrelated the probability that the highest degree nodes are connected to each other is relatively low. Then, when a new node is added and connected to m of the highest degree existing nodes there is a high probability that a biclique is formed. Such a process is continued as more nodes are added to the graph, resulting a large bicliques with high probability (see Figure 2.6). The creation of an ER network follows a completely different process in which pairs of nodes are connected randomly and independently, which does not generates any preferred subgraphs, thus not producing a large number of bicliques. This qualitative analysis explaining structurally the existence of networks with high values of the H index will be very useful in the next section of this thesis where we will analyze real-world networks.

In order to corroborate the hypothesis that the presence of complete bipartite structures on networks play a fundamental role on the high values of the H index we developed the following experiment. We consider complete bipartite graphs K_{n_1,n_2} with $n = n_1 + n_2$ equal to 100, and 200, respectively. Then, we rewire randomly and independently a small percentage of the edges of these graphs and compute the average value of the H index after 100 random rewirings. These rewirings are produced by keeping the

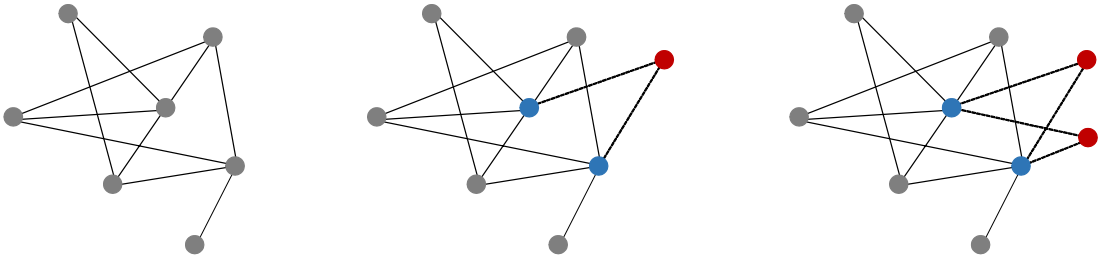


Figure 2.6: Illustration of the evolution of a graph under the BA model to sketch how bicliques are formed in such kind of networks. (a) Seed of $m_0 = 7$ nodes created with a Poissonian degree distribution to start the BA evolution process. (b) Given $m = 2$ the new node (red one) is preferentially attached to those with the highest degree among the existing m_0 ones (marked as blue). (c) Second iteration of the process, which creates a biclique $K_{2,2}$ (red and blue nodes joined by dotted lines).

same degree of each of the nodes in the original graphs. In Figure 2.7 we illustrate graphically the results of this experiment. We keep the rewiring probability low to avoid the appearance of other effects produced by the total negletion of the original complete bipartite structure. As can be seen for each of the complete bipartite graphs studied, the rewiring of the edges, which necessarily implies the destruction of the complete bipartivity, decays significantly the values of H index. We have studied two sizes of the graphs and observed that there are no significant differences in the results when the size of the graphs is doubled from 100 to 200. However, there is a significant change in the shape of the decaying function of the H index as a function of the rewiring probability for different kinds of complete bipartite structures. That is, when $n_1 = n_2$ there is a much faster decay to the H index as a consequence of the rewiring than when $n_1 \gg n_2$. In the last case, the number of nodes in one of the two disjoint sets of the bipartite graphs is relatively small in relation to the number of edges rewired. Consequently, these nodes can be totally saturated by a few rewirings that connect almost every pair of the nodes in this set. Thus, such small set is transformed into a random subgraph more quickly than a larger set of nodes, which is the case when $n_1 = n_2$.

In closing, we have provided theoretical and empirical evidence showing that the H index accounts for the existence of complete bipartite subgraphs in a network. Such subgraphs can appear naturally in preferential attachment processes, like the Barabási-Albert model, but they may also emerge as a consequence of other natural or man-made

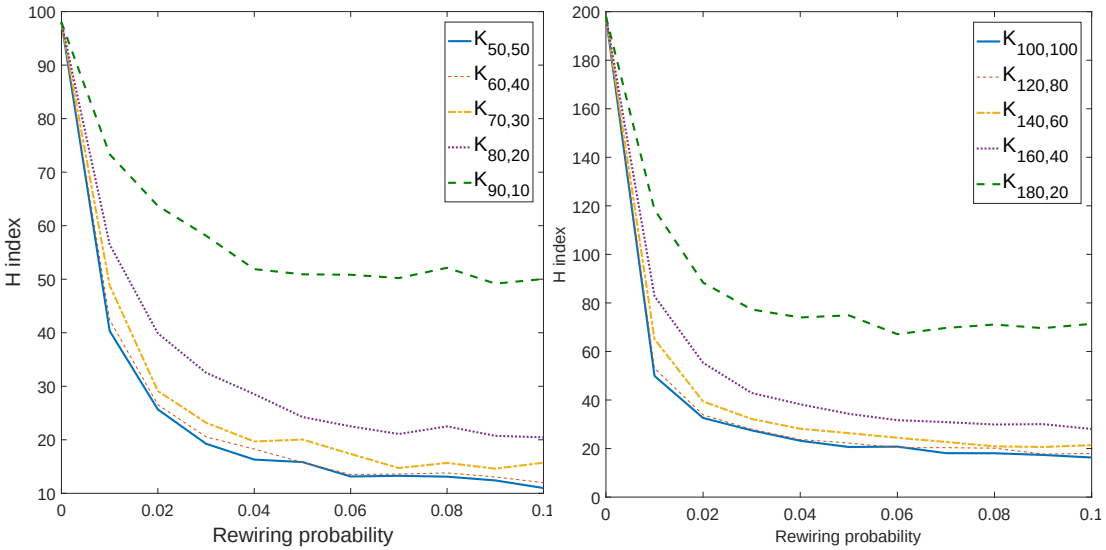


Figure 2.7: Effects of rewiring the edges of complete bipartite graphs K_{n_1,n_2} by keeping the degree of the nodes on the H index. In the left panel we illustrate graphs with 100 nodes and in the right panel we shown results for graphs with 200 nodes. Every point in the curves is obtained from the average of 100 random rewirings of the edges of each graph that keeps the same degree of the nodes.

processes giving rise to the actual structure of networks.

2.3 Studies of Real-World Networks

In this section we study a group of real-world networks representing a variety of social, environmental, technological, infrastructural and biological complex systems. A description of the networks and their main characteristics are given in Appendix B. The sizes of the networks studied here range from 29 to 4,941 nodes. Together with the H index we have also calculated the Estrada index $EE = tr\left(\exp\left(A\right)\right)$, the graph nullity η, and the Newman degree assortativity coefficient r [112] then we obtained the linear correlation coefficient among every pair of variables for the group of networks studied and the results are given in Table 2.2. The assortativity coefficient r is given by:

$$r = \frac{\frac{1}{m}\sum_{(i,j)\in E} k_i k_j - \left(\frac{1}{2m}\sum_{(i,j)\in E}(k_i + k_j)\right)^2}{\frac{1}{2m}\sum_{(i,j)\in E}(k_i^2 + k_j^2) - \left(\frac{1}{2m}\sum_{(i,j)\in E}(k_i + k_j)\right)^2}. \tag{2.53}$$

	n	H	EE	η	r
n	1	0.942	0.283	0.734	0.203
H		1	0.202	0.915	0.064
EE			1	0.030	-0.027
η				1	-0.077
r					1

Table 2.1: Pearson correlation coefficients among every pair of variables studied in this work (see (2.2)).

It measures the degree-degree Pearson correlation coefficient. The networks where $r > 0$ are known as assortative and those for which $r < 0$ are known as disassortative. Networks with $r = 0$ are known as neutral. The index indicates the tendency of high-degree nodes to be connected to each other (assortative networks) or to low degree nodes (disassortative networks).

As can be seen, the H index displays a significant correlation with the network size. Then, in order to avoid any size influence, we normalize the H index by dividing it by the number of nodes of the network. We will call $\hat{H}$ the normalized index. The normalized index $\hat{H}$ ranges from about 0.14 to about 0.75 for the studied networks, indicating that real-world networks cover most of the values that this index can take (see table 2.2). The scatterplot of the normalized nullity versus the normalized H index for the 61 real-world networks studied here (see Fig 2.8(b)) reveals that although both indices follow the same trend, there are important differences among them. In particular, we can observe that there are 9 networks with zero nullity which display values of $\hat{H}$ ranging from about 0.14 (the lowest $\hat{H}$ index) to about 0.36 (ranked 25th in increasing order of $\hat{H}$ index).

Name	H	$\hat{H}$	EE	η	r
Ants	30.998	0.4189	2.64E+02	14	-0.102
Benguela	9.573	0.3301	4.11E+06	0	0.021
BridgeBrook	56.018	0.7469	9.20E+08	48	-0.668
Canton	40.333	0.3735	3.12E+08	24	-0.226
CatCortex	12.636	0.2430	8.95E+09	0	-0.044
Centrality_literature	42.976	0.3642	2.44E+08	9	-0.202

Chesapeake	13.240	0.4012	4.71E+02	3	-0.196
Coachella	10.984	0.3661	7.61E+07	0	0.035
ColoSpg	182.077	0.5620	1.15E+03	142	-0.295
CorporatePeople	228.395	0.1440	1.27E+10	0	0.268
Dolphins	20.845	0.3362	2.06E+03	2	-0.044
Drugs	279.467	0.4537	6.91E+07	131	-0.117
Electronic1	37.694	0.3090	4.84E+02	0	-0.002
Electronic2	77.982	0.3095	1.04E+03	8	-0.006
Electronic3	158.658	0.3099	2.17E+03	24	-0.030
ElVerde	51.696	0.3314	4.76E+13	5	-0.174
Galesburg	9.519	0.3071	4.36E+02	1	-0.135
GD	90.440	0.3632	1.60E+04	15	0.098
Geom	1462.396	0.4039	4.04E+12	537	0.168
Hi_tech	10.975	0.3326	2.95E+03	1	-0.087
Internet1997	2148.635	0.7126	6.17E+13	1883	-0.229
Internet1998	2473.122	0.7022	1.42E+15	2158	-0.210
LittleRockA	117.772	0.6507	5.32E+17	93	-0.234
MacaqueVisualCortex	9.665	0.3020	1.26E+06	1	0.008
Neurons	69.083	0.2467	1.31E+10	3	-0.069
ODLIS	1131.046	0.3903	1.54E+19	270	-0.173
PIN_Afulgidus	16.366	0.5114	9.91E+01	12	-0.472
PIN_Bsubtilis	53.144	0.6327	3.52E+02	46	-0.486
PIN_Ecoli	102.189	0.4443	8.30E+06	57	-0.015
PIN_Hpyroli	397.649	0.5601	4.60E+04	316	-0.243
PIN_KSHV	18.119	0.3624	1.82E+03	2	-0.058
PIN_Malaria	83.377	0.3641	2.25E+04	13	-0.083
PIN_Yeast	1135.731	0.5107	1.94E+08	754	-0.105
Power_grid	1907.307	0.3860	2.13E+04	593	0.003
PRISON	20.325	0.3034	7.08E+02	0	0.103
ReefSmall	12.888	0.2578	2.07E+10	0	-0.193

Roget	264.570	0.2662	2.38E+05	2	0.174
Sawmill	12.307	0.3419	2.57E+02	2	-0.071
ScotchBroom	103.975	0.6752	2.46E+06	90	-0.311
Shelf	20.724	0.2559	1.60E+18	2	-0.094
Skipwith	15.023	0.4292	3.87E+09	7	-0.319
SmallWorld	115.730	0.4967	1.27E+09	70	-0.303
College	8.049	0.2515	5.36E+02	0	-0.119
Software_Abi	575.133	0.5557	1.65E+05	418	-0.086
Software_Digital	82.277	0.5485	1.31E+03	63	-0.228
Software_Mysql	648.971	0.4385	2.70E+09	282	-0.083
Software_VTK	440.251	0.5710	1.11E+05	324	-0.195
Software_XMMS	478.168	0.4924	4.64E+04	294	-0.114
StMarks	13.607	0.2835	1.43E+05	0	0.111
StMartin	14.438	0.3281	2.78E+05	2	-0.153
Stony	41.359	0.3693	7.23E+09	30	-0.222
Termite_1	206.581	0.4075	1.92E+03	75	-0.046
Termite_2	116.912	0.4497	7.32E+02	58	-0.150
Termite_3	100.975	0.3768	1.89E+03	23	0.045
Trans_Ecoli	214.517	0.6540	1.06E+04	184	-0.265
Trans_urchin	22.218	0.4937	9.12E+02	13	-0.207
Transc_yeast	478.315	0.7225	3.59E+04	440	-0.410
USAir97	142.765	0.4300	8.08E+17	58	-0.208
Ythan1	58.374	0.4356	1.86E+07	23	-0.263
Ythan2	41.326	0.4492	7.07E+06	22	-0.322
Zackar	15.994	0.4704	1.04E+03	10	-0.476

Table 2.2: Dataset of real-world networks studied in this thesis, their size n, Gaussian Estrada index H, exponential Estrada index EE, graph nullity η, and degree assortativity r.

The largest value of $\hat{H}$ corresponds to the food web of Bridge Brook, which displays the second highest normalized nullity, i.e., the nullity divided by n. It is followed by the

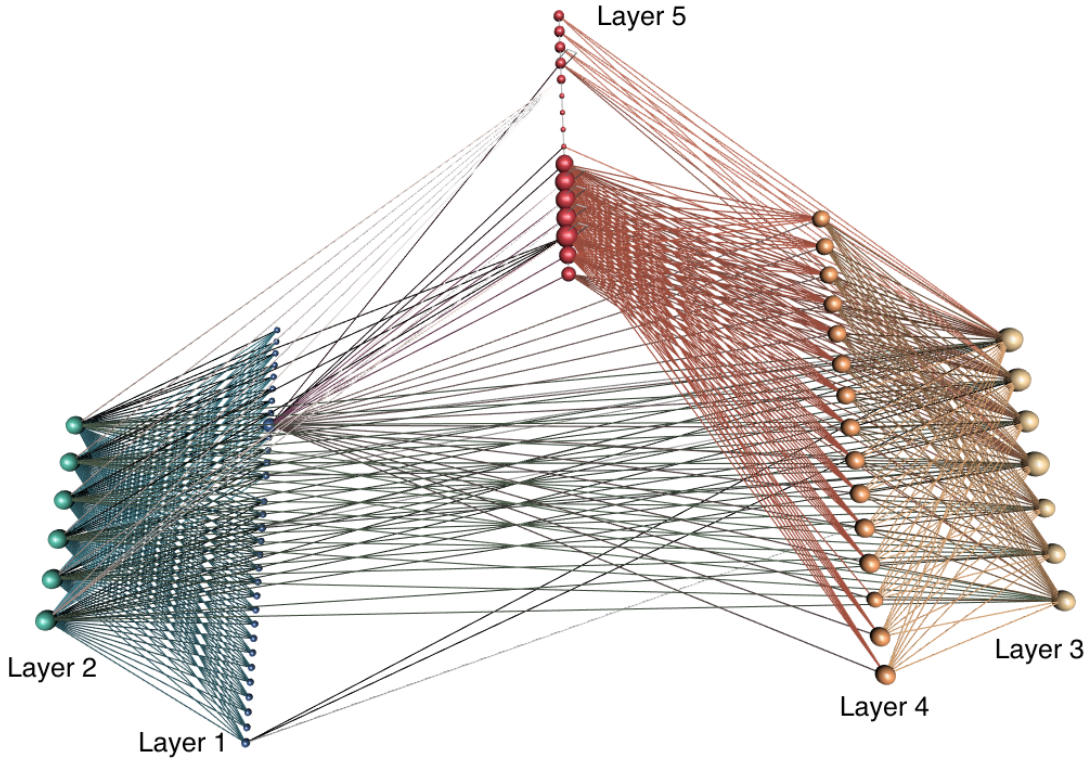
transcription network of yeast (displaying the highest value of the normalized nullity) and the versions of Internet at Autonomous System (AS) of 1997 and 1998. The three networks display triangular eigenvalue distributions peaked at the zero eigenvalue which explains their large values of the $\hat{H}$ index. However, while the yeast transcription network and the Internet at AS have fat-tailed degree distributions, the Bridge Brook food web displays a uniform one. Thus, the existence of large values of the $\hat{H}$ index is not tied up to the existence of fat-tailed degree distributions. Most of the networks (75.4%) have values of the $\hat{H}$ index below 0.5. That is, only 15 networks out of 61 have $\hat{H} \geq 0.5$. Among these 15 networks there are 4 of the 7 protein-protein interaction networks (PINs) studied and two of the three transcription networks studied. Thus, almost half of the networks with $\hat{H} \geq 0.5$ represent biological systems containing proteomic or transcriptomic information. The other transcription network studied has $\hat{H} \approx 0.494$ and the other 3 PINs have values of $\hat{H}$ ranging between 0.36 and 0.44. It is interesting to explore the main structural causes for these high values of the $\hat{H}$ index. In previous sections we have found that the main structural characteristic determining the high values of this index is the presence of bicliques, e.g. the highest value of $\hat{H}$ is obtained for complete bipartite graphs, also the BA networks display larger $\hat{H}$ index than the ER ones due to the presence of complete bipartite subgraphs created during the evolution of the preferential attachment mechanism. Consequently, we should expect that such kind of subgraphs appear in those real-world networks having the largest $\hat{H}$ index. In the case of the food web of Bridge Brook we have found a biclique consisting of two sets of nodes V_1 and V_2 with cardinalities of 6 and 35 nodes, respectively (see Figure 2.8(a)). This subgraph represents a biclique $K_{6,35}$ which contains 55% of the total number of nodes in the network. There are also other smaller bicliques in this network (one of them is a biclique $K_{7,15}$), which together with the $K_{6,35}$ contribute to the large $\hat{H}$ value observed. We have corroborated the differences between these two kind of systems by plotting the decay of the H index as a function of the rewiring probability. As shown in figure 2.8(c) we observed that the network of Bridge Brook displays a significant decay of the H index for relatively small rewiring probabilities, while the H index of the cat visual cortex network remains almost constant after random

rewiring.

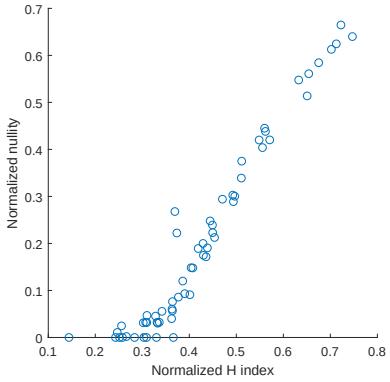
In the cases of the yeast transcription network and the Internet at AS, the networks are characterized by having a few hubs connected to many nodes of degree one, then producing bicliques of the type K_{1,n_2}. In fact, almost 90% of the induced subgraphs of size 5 in both Internet networks studied here are $K_{1,4}$ and almost 70% of the induced subgraphs of size 6 are $K_{1,5}$. In general these findings can be understood on the basis of different mechanisms which give rise to the existence of bicliques in real-world networks. For instance, in some food webs there are top predators which compete for a group of preys. If for this group of species there are no prey-prey nor predator-predator trophic interactions, the corresponding subgraph is a biclique as the one observed for the Bridge Brook network previously considered. In the cases of transcription and PINs the bicliques can be formed as a consequence of lock-and-key kind of interaction. That is, a group of proteins (genes) can act as locks (activators) that physically interact with other proteins (activate other genes) acting as keys. Such kind of interactions is prone to produce relatively large bicliques in the structure of the networks resulting from them. On the other hand, among the networks with $\hat{H} \leq 0.3$ we find the network of corporate directors, the three neuronal networks studied, i.e., macaque and cat visual cortex and the neuronal network of *C. elegans*, as well as some social networks and food webs. Also, the three electronic circuits studied here also display values of $\hat{H}$ index around 0.3. These networks are characterized by the lack of complete bipartite subgraphs and they may represent a variety of topologies difficult to be reproduced by a single mechanism.

2.4 Further Researches

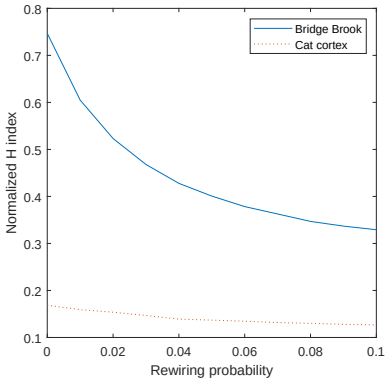
We would like to remark a few important characteristics of the Gaussian matrix function of a network that point out to the necessity of further studies of it for real-world networks and simple graphs in general. The first, is our observation that although networks with fat-tailed degree distribution may give rise to high values of the $\hat{H}$ index, it is not a necessary condition for a network to display such a characteristic. We have seen that networks with exponential and even uniform degree distributions



(a)



(b)



(c)

Figure 2.8: (a) Representation of the food web of Bridge Brook. Layer 1 and layer 2 form a biclique $K_{6,35}$. Layers 3 and 4 form a biclique $K_{7,15}$. (b) scatterplot of the normalized nullity versus the normalized H index for the 61 real-world networks studied here. (c) Effect of rewiring the edges of the Bridge Brook and cat cortex networks by keeping the degree of the nodes on the H index. Every point in the curves is obtained from the average of 100 random rewirings of the edges of each graph that keeps the same degree of the nodes.

display large values of the $\hat{H}$ index. Another structural parameter that could be related to the $\hat{H}$ index is the degree assortativity, i.e., the Pearson correlation coefficient of the degree-degree distribution of a network. We have explored such relation between the $\hat{H}$ index and the assortativity for the 61 networks studied here. We have found that the two parameters are negatively correlated. That is, high values of the $\hat{H}$ index in general implies that the networks are disassortative, i.e., there is a trend of high degree nodes to be connected to low degree ones. This is understandable on the basis of our findings that bicliques of the type K_{1,n_2} play a fundamental role in the value of the $\hat{H}$ index. However, the correlation is very weak and displays a Pearson correlation coefficient of -0.68. Thus, further explorations—both theoretical and computational—of the relation of the $\hat{H}$ index and other network parameters are necessary for a complete understanding of this index and its application in network theory.

Chapter 3

Gaussian Communicability Function $e^{-(A+I)^2}$

The previous chapter was devoted to the reference eigenvalue $\lambda_{\text{ref}} = 0$. In this Chapter we generalize the concept of folding the spectrum around any chosen reference eigenvalue. Thus for any reference eigenvalue λ_{ref}, an eigensolution $(\lambda_j, \mathbf{q}_j)$ of (2.1) also satisfies

$$(\lambda_{\text{ref}} I - A)^2 \mathbf{q}_j = (\lambda_{\text{ref}} - \lambda_j)^2 \mathbf{q}_j, \tag{3.1}$$

where λ_{ref} is a given reference eigenvalue. This process of 'folding' the spectrum of a network is illustrated in Figure 3.1. In the left-hand side of the Figure we shown the normal spectrum of the network, in which the eigenvalues are represented on the line as dots. The reference eigenvalue λ_{ref} is marked with a star in the axis. We then consider $(\lambda_{\text{ref}} - \lambda_j)^2$ and represent the squared differences in a new line on the right-hand side of the Figure. The first consequence of this folding process is that now all the eigenvalues are nonnegative, with the smallest one corresponding to λ_{ref}, which obviously takes now the value of zero. Now, as we want to give more weight to λ_{ref} in the matrix function, we transform the spectrum in a way that for every eigenvalue we obtain a Gaussian of the form $\exp\left[-(\lambda_{\text{ref}} - \lambda_j)^2\right]$. The result of this Gaussian spectrum of the adjacency matrix is given in the right-hand sided panel of the Figure 3.1. As can be seen the entry corresponding to λ_{ref} receives the largest weight, which is equal to one, and the

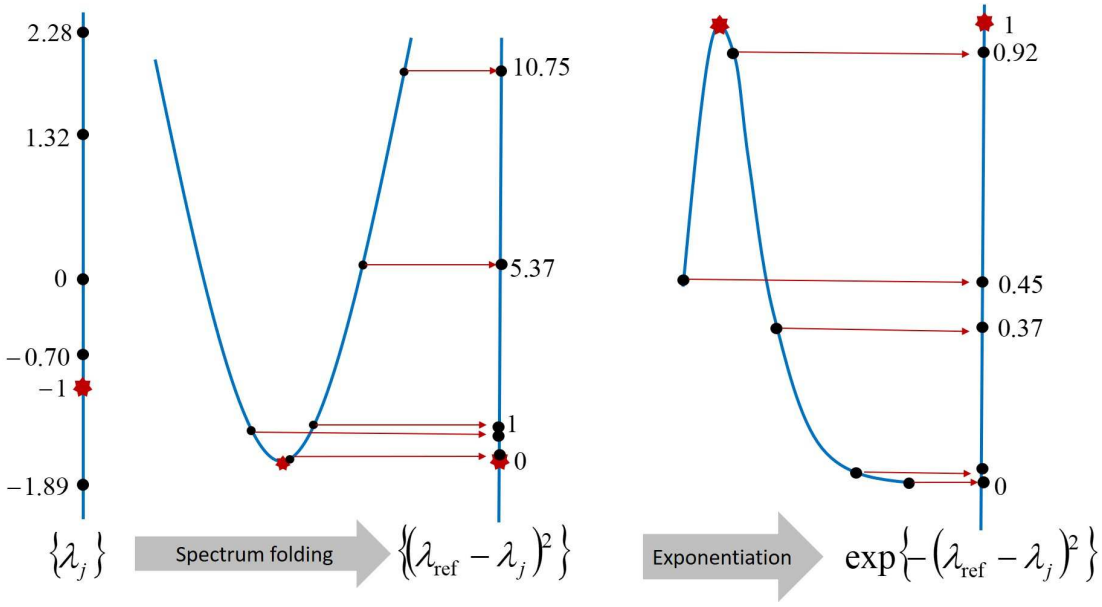


Figure 3.1: Illustration of the folded spectrum method. The eigenvalues of the adjacency matrix of the network are folded at λ_{ref} into the spectrum of $(\lambda_{\text{ref}}I - A)^2$. Then we exponentiate them to give more weight to λ_{ref}.

rest of the eigenvalues receive weights smaller than one. We will call

$$\tilde{G}\left(\lambda_{\text{ref}}\right) = \sum_{k=0}^{\infty} \frac{\left(-\left(\lambda_{\text{ref}}I - A\right)^2\right)^k}{k!} = \exp\left[-\left(\lambda_{\text{ref}}I - A\right)^2\right] \tag{3.2}$$

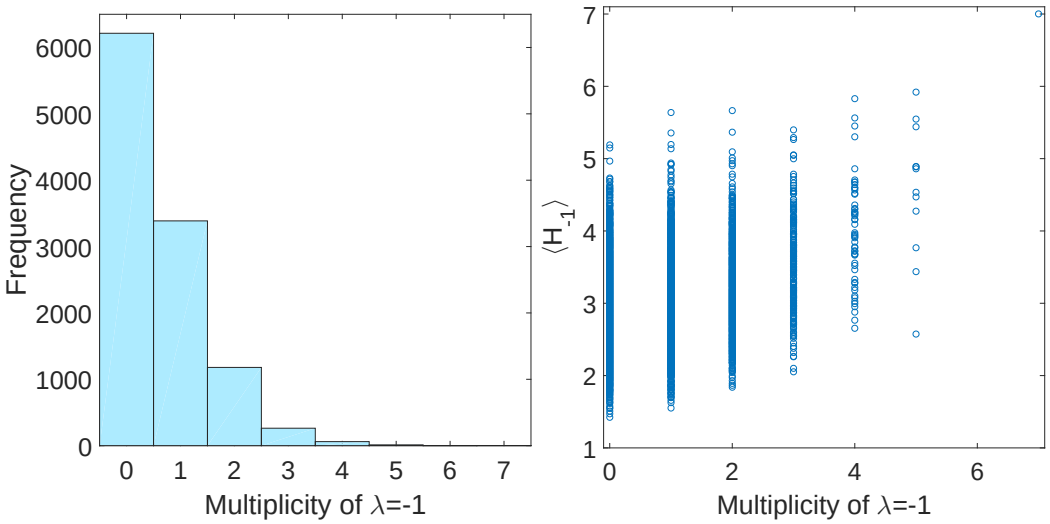
the Gaussian adjacency matrix, $\tilde{G}_{pq}$ the Gaussian communicability function between the nodes p and q and

$$H_{\lambda_{\text{ref}}} = tr\tilde{G}\left(\lambda_{\text{ref}}\right) \tag{3.3}$$

the Gaussian Estrada index of the graph. The term $\tilde{G}_{pp}$ is the *Gaussian subgraph centrality* based on the same matrix function. This chapter is dedicated to $\lambda_{\text{ref}} = -1$ motivated by the importance of it in graphs. It is well known that the complete graph K_n has an eigenvalue -1 with multiplicity $n-1$. The next results characterize other extreme cases.

Proposition 3.0.1. [116] Let G be a graph on n vertices having an eigenvalue -1 with multiplicity $n-2$. Then G is the disjoint union of two complete graphs.

The following result is attributed in [116] to an unpublished result by Van Dam,



(a) (b)

Figure 3.2: (a)Histogram of the multiplicity of the eigenvalue $\lambda_{\text{ref}} = -1$ of the adjacency matrix of all 11,117 connected graphs with 8 nodes. (b) Scatterplot of the multiplicity of the eigenvalue $\lambda = -1$ vs. the index H_{-1} for the 11,117 connected graphs with 8 nodes.

Haemers and Stevanović.

Proposition 3.0.2. Let G be a graph with n vertices having an eigenvalue -1 with multiplicity $n-3$. Then $G = K_n \setminus K_{\ell,m}$, where $\ell, m \geq 1$, $\ell + m \leq n-1$, or $G = K_k + K_\ell + K_m$, where $k, \ell, m \geq 1$, $k + \ell + m = n$.

Proposition 3.0.3. [117] Let G be any graph. If G includes n_1 mutually adjacent vertices sharing the same closed neighborhood, then the spectrum of the adjacency matrix of G contains the eigenvalue -1 with multiplicity $n_1 - 1$.

In order to illustrate how frequently $\lambda = -1$ appears in graphs we study all the connected graphs having $n = 8$ vertices. Among the 11,117 connected graphs with 8 nodes, 55.9% have no eigenvalue $\lambda = -1$, 30.1% has multiplicity 1 for this eigenvalue, 10.6% have multiplicity 2, 2.4% have multiplicity 3, 0.5% have multiplicity 4, and only 12 graphs have multiplicity 5 and the complete graph that has multiplicity 7 (see the histogram in Figure 3.2 (a). Another type of graphs in which the eigenvalue $\lambda = -1$ also appears is in the study of benzenoid systems [114]. Benzenoid systems are formed by fusing together hexagonal cycles forming different shapes. They may represent

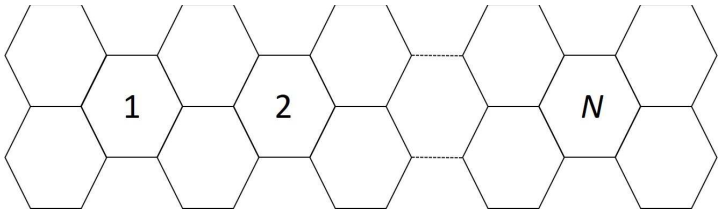


Figure 3.3: A benzenoid system in which the multiplicity of the eigenvalue $\lambda = -1$ depends on the size of the system and the number of perylene units.

molecules, like polycyclic aromatic compounds, or radicaloids such as triangulenes [113, 115]. It is known that linear benzenoid systems have $\lambda = -1$ with multiplicity 1 or 2, for even or odd number of hexagons, i.e., the single hexagon (benzene) has multiplicity 2 for this eigenvalue. There are systems like triphenylene where the multiplicity of this eigenvalue is zero, while others have multiplicity 3 (coronene), 4 (perylene) or a multiplicity that depends on the size of the system. For instance, the benzenoid system [114] illustrated in Figure 3.3 where the number of perylene units are designated by N has multiplicity of the eigenvalue $\lambda = -1$ equal to $3j + 4$, where j is an integer determining the number N as $N = 3j + 1$. Although the multiplicity of the eigenvalue $\lambda = -1$ makes an important global contribution to the index H_{-1}, it is important to remark that this index accounts for the global influence of the eigenvalues of the graph giving more weight to those close to the eigenvalue $\lambda = -1$. This is clearly observed on the plot of multiplicity of the eigenvalue $\lambda = -1$ vs. the index H_{-1} for the 11,117 connected graphs with 8 nodes in Figure 3.2 (b). For instance, notice that among the graphs with zero multiplicity of the eigenvalue $\lambda = -1$ there are some graphs having values of the index H_{-1} which range between 1.5 to more than 5. Consequently, there are graphs with no eigenvalues $\lambda = -1$ which have larger values of H_{-1} than others having multiplicity 5 for this eigenvalue.

3.1 H_{-1} Index of Graphs

3.1.1 Elementary Properties

Here we will investigate some mathematical properties of the index H_{-1}, in particular some analytical formulas for specific types of graphs.

Lemma 3.1.1. Let K_n be the complete graph of n nodes. Then

$$H_{-1}\left(K_n\right) = e^{-n^2} + n - 1. \tag{3.4}$$

Proof. The spectrum of K_n is $\sigma(K_n) = \left\{[n-1]^1, [-1]^{n-1}\right\}$, thus we have

$$\begin{aligned} H_{-1}\left(K_n\right) &= \sum_{j=1}^{n} e^{-(1+\lambda_j)^2} \\ &= (n-1)e^0 + e^{-n^2} \\ &= e^{-n^2} + n - 1. \end{aligned} \tag{3.5}$$

□

Lemma 3.1.2. Let K_{n_1,n_2} be the complete bipartite graph of $n_1 + n_2$ nodes. Then

$$H_{-1}\left(K_{n_1,n_2}\right) = e^{-1}\left(2e^{-n_1 n_2}\cosh(2\sqrt{n_1 n_2}) + n_1 + n_2 - 2\right). \tag{3.6}$$

Proof. The spectrum of K_{n_1,n_2} is $\sigma\left(K_{n_1,n_2}\right) = \left\{\left[\sqrt{n_1 n_2}\right]^1, \left[-\sqrt{n_1 n_2}\right]^1, [0]^{n_1+n_2-2}\right\}$, thus we have

$$\begin{aligned} H_{-1}\left(K_{n_1,n_2}\right) &= \sum_{j=1}^{n_1+n_2} e^{-(1+\lambda_j)^2} \\ &= (n_1+n_2-2)e^{-1} + e^{-(1+\sqrt{n_1 n_2})^2} + e^{-(1-\sqrt{n_1 n_2})^2} \\ &= \frac{n_1+n_2-2}{e} + \frac{e^{-n_1 n_2}}{e}\left(e^{-2\sqrt{n_1 n_2}} + e^{2\sqrt{n_1 n_2}}\right) \\ &= e^{-1}\left(2e^{-n_1 n_2}\cosh(2\sqrt{n_1 n_2}) + n_1 + n_2 - 2\right). \end{aligned} \tag{3.7}$$

□

Corollary 3.1.1. Let $K_{1,n-1}$ be the star graph of n nodes. Then

$$H_{-1}\left(K_{1,n-1}\right) = e^{-1}\left(e^{1-n}\cosh(2\sqrt{n-1}) + n - 2\right). \tag{3.8}$$

In order to give the formula for the path and cycle graph we use the following known

Lemma.

Lemma 3.1.3. Let f and g be defined on the interval $[a, b]$ with g continuous, $f \geq 0$, and f is integrable. Then there is a point $x_0 \in (a, b)$ such that $\int_a^b f(x)g(x)dx = g(x_0)\int_a^b f(x)dx$.

Proof. We can assume that $g([a, b]) = [m, M]$ because g is a continuous function. Since $f \geq 0$ on $[a, b]$ we can write:

$$\int_a^b g(x)f(x)dx - m\int_a^b f(x)dx = \int_a^b (g(x) - m)\, f(x)dx \geq 0. \tag{3.9}$$

In a similar way, we get

$$M\int_a^b f(x)dx - \int_a^b g(x)f(x)dx = \int_a^b (M - g(x))\, f(x)dx \geq 0. \tag{3.10}$$

Thus, the inequality holds

$$m\int_a^b f(x)dx \leq \int_a^b g(x)f(x)dx \leq M\int_a^b f(x)dx. \tag{3.11}$$

That yields to the following:

$$m \leq \frac{1}{\int_a^b f(x)dx}\int_a^b g(x)f(x)dx \leq M. \tag{3.12}$$

Then, by the intermediate value theorem there exists $x_0 \in (a, b)$ such that

$$g(x_0) = \frac{1}{\int_a^b f(x)dx}\int_a^b g(x)f(x)dx.$$

□

Lemma 3.1.4. Let C_n be a cycle having n nodes. Then, asymptotically as $n \to \infty$

$$H_{-1}(C_n) \sim ne^{-3}e^{-4\cos c}I_0(-2), \text{ for some } c \in (0, \pi). \tag{3.13}$$

Proof.

$$\begin{aligned} H_{-1}(C_n) &= n\left(\frac{tr(e^{-(I+A)^2})}{n}\right) \\ &= n\left(\frac{1}{n}\sum_{j=1}^{n} e^{-(1+2\cos(\frac{2\pi j}{n}))^2}\right) \\ &= e^{-3}n\left(\sum_{j=1}^{n}\frac{1}{n}e^{4\cos(\frac{2\pi j}{n})}e^{-2\cos(\frac{4\pi j}{n})}\right). \end{aligned} \tag{3.14}$$

Now, when $n \to \infty$ the summation can be approximated by the following integral

$$\begin{aligned} H_{-1}(C_n) &\sim \frac{e^{-3}n}{2\pi}\int_0^{2\pi} e^{-4\cos\frac{\theta}{2}}e^{-2\cos\theta}d\theta \\ &= \frac{e^{-3}n}{\pi}e^{-4\cos c}\int_0^{\pi} e^{-2\cos\theta}d\theta, \text{ for some } c \in (0,\pi) \\ &= ne^{-3}e^{-4\cos c}I_0(-2), \text{ for some } c \in (0,\pi). \end{aligned} \tag{3.15}$$

□

Lemma 3.1.5. Let P_n be a path having n nodes. Then, asymptotically as $n \to \infty$

$$H_{-1}(P_n) \sim e^{-3}e^{-4\cos c}\left((n+1)I_0(2) - e^{-2}\right), \text{ for some } c \in (0,\frac{\pi}{2}). \tag{3.16}$$

Proof. By substituting the eigenvalues and eigenvectors of the path graph into the expression for $\tilde{G}_{pp}(P_n)$ we obtain

$$\begin{aligned} \tilde{G}_{pp}(P_n) &= \frac{2}{n+1}\sum_{j=1}^{n}\sin^2\left(\frac{j\pi p}{n+1}\right)\exp\left[-\left(1+2\cos\left(\frac{j\pi}{n+1}\right)\right)\right]^2 \\ &= \frac{e^{-3}}{n+1}\sum_{j=1}^{n}\left[1-\cos\left(\frac{2j\pi p}{n+1}\right)\right]e^{\left(-4\cos\left(\frac{j\pi}{n+1}\right)-2\cos\left(\frac{2j\pi}{n+1}\right)\right)}. \end{aligned} \tag{3.17}$$

Now, when $n \to \infty$ the summation in (3.17) can be approached by the following integral

$$\begin{aligned}\tilde{G}_{pp}(P_n) &\sim \frac{e^{-3}}{\pi}\int_0^{\pi}\left(e^{(-4\cos\frac{\theta}{2}-2\cos\theta)} - \cos(p\theta)\, e^{(-4\cos\frac{\theta}{2}-2\cos\theta)}\right)d\theta \\ &= \frac{e^{-3}}{\pi}e^{\left(-4\cos\frac{c}{2}\right)}\int_0^{\pi}\left(e^{(-2\cos\theta)} - \cos(p\theta)\, e^{(-2\cos\theta)}\right)d\theta, \text{ for some } c\in(0,\pi)\end{aligned}$$

where $\theta = \frac{2j\pi}{n+1}$. Thus, when $n \to \infty$ we have

$$\tilde{G}_{pp}(P_n) \sim e^{-3}\exp(-4\cos c)\left(I_0(-2) - I_p(-2)\right), \text{for some } c\in(0,\frac{\pi}{2}). \tag{3.18}$$

Completing the proof as in the proof of Lemma 2.2.2 will produce the final result. □

The most important thing about the formulas for $H_{-1}(C_n)$ and $H_{-1}(P_n)$ is that they show a linear increase of the index with the number of nodes in the graph. That is, $H_{-1}(C_n) \approx 0.3595n$ and $H_{-1}(P_n) \approx 0.3593n + 0.2239$.
Another analytical expression of the H_{-1} index for the paths and cycles is in the following Lemma.

Lemma 3.1.6. Let P_n be a path having n nodes. Then, asymptotically as $n \to \infty$

$$H_{-1}(P_n) \sim \frac{n+1}{\pi}\int_0^{\pi} e^{-(2\cos\theta+1)^2}d\theta - \frac{1}{2}\left(e^{-1}+e^{-9}\right). \tag{3.19}$$

Proof. The spectrum of P_n consists of the numbers $2\cos\frac{j\pi}{n+1}$, $j = 1, 2, \ldots, n$. The angles $\frac{j\pi}{n+1}$ do not cover the entire interval $[0, \pi]$. Therefore when employing an integral approximation we need to compensate for the missing near-zero and near-π contributions [118]. It is done as follows:

$$\begin{aligned}H_{-1}(P_n) &= \sum_{j=1}^{n} e^{-(\lambda_j+1)^2} \\ &= \sum_{j=1}^{n} e^{-\left(2\cos\left(\frac{j\pi}{n+1}\right)+1\right)^2} \\ &= \frac{1}{2}\sum_{j=0}^{n} e^{-\left(1+2\cos\left(\frac{j\pi}{n+1}\right)\right)^2} + \frac{1}{2}\sum_{j=1}^{n+1} e^{-\left(1+2\cos\left(\frac{j\pi}{n+1}\right)\right)^2} - \frac{1}{2}e^{-9} - \frac{1}{2}e^{-1}.\end{aligned} \tag{3.20}$$

Now, when $n \to \infty$ the summation can be approached by the following integral

$$H_{-1}(P_n) \sim \frac{n+1}{\pi} \int_0^{\pi} e^{-(2\cos\theta+1)^2} d\theta - \frac{1}{2}\left(e^{-1} + e^{-9}\right). \tag{3.21}$$

□

Lemma 3.1.7. Let C_n be a cycle having n nodes. Then, asymptotically as $n \to \infty$

$$H_{-1}(C_n) \sim \frac{n}{\pi} \int_0^{\pi} e^{-(2\cos\theta+1)^2} d\theta. \tag{3.22}$$

Proof.

$$\begin{aligned} H_{-1}(C_n) &= n\left(\frac{tr\left(e^{-(A+I)^2}\right)}{n}\right) \\ &= n\left(\frac{1}{n}\sum_{j=1}^{n} e^{-\left(1+2\cos\left(\frac{2\pi j}{n}\right)\right)^2}\right) \end{aligned} \tag{3.23}$$

Now, when $n \to \infty$ the summation in 3.23 can be approached by the following integral

$$\begin{aligned} H_{-1}(C_n) &\sim n\frac{1}{2\pi}\int_0^{2\pi} e^{-(1+2\cos\theta)^2} d\theta \\ &= n\frac{1}{2\pi}(2)\int_0^{\pi} e^{-(1+2\cos\theta)^2} d\theta \end{aligned} \tag{3.24}$$

where $\theta = \frac{2j\pi}{n}$. Thus, when $n \to \infty$ we have

$$H_{-1}(C_n) \sim \frac{n}{\pi} \int_0^{\pi} e^{-(2\cos\theta+1)^2} d\theta. \tag{3.25}$$

□

3.1.2 Graphs with Maximum H_{-1} Index

In this section we will prove that among all connected graphs with n nodes, the index takes its maximum for the complete graph. First we list two results from spectral graph theory that will be used to prove the main theorem in this section.

Theorem 3.1.1. [120] A graph has exactly one positive eigenvalue if and only if its non-isolated vertices form a complete multipartite graph.

Theorem 3.1.2. [121]

1. The $p-1$ negative eigenvalues $\lambda_{n-p+1}, \ldots, \lambda_n$ of a complete p-partite graph $K(p_i)$ with p partition numbers p_i and t distinct partition numbers $\bar{p_i}$ satisfy the inequalities $p_1 \leq -\lambda_{n-p+2} \leq p_2 \leq \lambda_{n-p+3} \leq \ldots \leq p_{p-1} \leq -\lambda_n \leq p_p$.

2. Moreover, for $t-1$ of the negative eigenvalues $\lambda_i^*(i = 2, \ldots, t)$, the strict inequalities $\bar{p_1} < -\lambda_2^* < \bar{p_2} < -\lambda_3^* < \ldots < \bar{p}_{t-1} < -\lambda_t^* < \bar{p}_t$ hold, and together with λ_1 they are the roots of the equation $1 - \sum_{i=1}^{t} \frac{t_i \bar{p}_i}{x+\bar{p}_i} = 0$.

It is straighforward to realize that the nullity of the complete p-partite graph, i.e., the multiplicity of the zero eigenvalue, is $\eta(K(p_1^{t_1}, \ldots, p_r^{t_r})) = n - p$, where n is the number of nodes and p is the number of partitions. Now we give the main result of this section.

Theorem 3.1.3. Let G be a simple connected graph with n nodes. Then,

$$H_{-1}(G) \leq e^{-n^2} + n - 1. \tag{3.26}$$

where the equality holds if and only if $G = K_n$.

Proof. In order to prove the inequality (3.26), it is enough to find two eigenvalues λ, μ such that $e^{-(1+\lambda)^2} + e^{-(1+\mu)^2} \leq 1$ and that implies $\sum_{j=1}^{n} e^{-(1+\lambda_j)^2} \leq (n-2) + 1 = n - 1 \leq n - 1 + e^{-n^2}$. According to Eq. (3.4) this bound corresponds to the index of the complete graph. First, it is known that $\lambda_1 \geq \chi(G) - 1 \geq 1$ where $\chi(G)$ is the chromatic number of G (the number of the least coloring needed to color the graph G) [39]. Hence, $e^{-(1+\lambda_1)^2} \leq e^{-4} < e^{-1}$. Now, if $\lambda_2 > 0$, then $e^{-(1+\lambda_1)^2} < e^{-1}$ and $e^{-(1+\lambda_1)^2} + e^{-(1+\lambda_2)^2} \leq 2e^{-1} < 1$. If $\lambda_2 \leq 0$, then by theorem 3.1.1, G is a complete multipartite graph and from the interlacing Theorem for complete multipartite graphs (Theorem 3.1.2,1) we get $p_{p-1} \leq -\lambda_n$ so $\lambda_n \leq -p_{p-1}$. If $p_{p-1} \geq 2$ then $e^{-(1+\lambda_n)^2} \leq e^{-1}$. If $p_{p-1} = 1$ then $G = K(1^k, s)$ which has nullity $s - 1$. Now, if $s \geq 2$ then there exists

an eigenvalue $\lambda = 0$ and $e^{-(1+\lambda)^2} = e^{-(1)^2} = e^{-1}$. If $s = 1$ then G is the complete graph and we get the equality. □

3.1.3 H_{-1} of Random Graphs

we have the following results for the $H_{-1}(ER)$ index.

Theorem 3.1.4. For an Erdős-Rényi random graph $G(n,p)$ with $\frac{\ln n}{n} \ll p$ then, for significantly large r, we have

$$H_{-1}(ER) \sim \frac{2n\sqrt{r^2-1}}{r} e^{r^2} \mathrm{erfc}(r). \tag{3.27}$$

almost surely as $n \to \infty$, where $r = 2\sqrt{np(1-p)}$ and $\mathrm{erfc}\,(r) = 1 - \mathrm{erf}(r)$ is the complimentary error function of r.

Note that $e^{r^2}\mathrm{erfc}(r) \sim \frac{1}{\sqrt{\pi}r}$, thus we can write $H_{-1}(ER) \sim \frac{2n\sqrt{r^2-1}}{\sqrt{\pi}r^2}$. Before we prove this result we need the following two auxiliary Lemma.

Lemma 3.1.8. Let $x > 0$, $x \neq 1$. As $x \to \infty$ we have

$$\sqrt{x^2-a^2}e^{-(1+a)^2} \sim x^2\sqrt{x^2-1}\frac{e^{-(1+a)^2}}{(1+a)^2+x^2} \tag{3.28}$$

for any constant a satisfying $|a| < x$.

Proof. By taking the limit of $\frac{\sqrt{x^2-a^2}e^{-(1+a)^2}}{x^2\sqrt{1-x^2}\frac{e^{-(1+a)^2}}{(1+a)^2+x^2}}$ as $x \to \infty$ and using L'Hôpital rule we get

$$\begin{aligned}
\lim_{x\to\infty} \frac{\sqrt{x^2-a^2}e^{-(1+a)^2}}{x^2\sqrt{x^2-1}\frac{e^{-(1+a)^2}}{(1+a)^2+x^2}} &= \lim_{x\to\infty} \frac{\sqrt{x^2-a^2}}{\left(\frac{x^2\sqrt{x^2-1}}{(1+a)^2+x^2}\right)} \\
&= \lim_{x\to\infty} \frac{x^2-2a^2-(1+a)^2}{x^2-2}\sqrt{\frac{x^2-1}{x^2-a^2}} \\
&= 1
\end{aligned} \tag{3.29}$$

□

Lemma 3.1.9. [119]

$$\begin{aligned}\int_0^\infty \frac{e^{-a^2x^2}}{x^2+b^2}dx &= \frac{\pi}{2b}\left(1-\operatorname{erf}(ab)\right)e^{a^2b^2},\ b>0\\ &= \frac{\pi}{2b}\left(\operatorname{erfc}(ab)\right)e^{a^2b^2}.\end{aligned} \tag{3.30}$$

Proof. Let $x = bt$, then

$$\int_0^\infty \frac{e^{-a^2x^2}}{x^2+b^2}dx = \frac{1}{b}\int_0^\infty \frac{e^{-a^2b^2t^2}}{t^2+1}dt. \tag{3.31}$$

Let $f(c) = \int_0^\infty \frac{e^{-c^2\left(t^2+1\right)}}{t^2+1}dt$, where $c = ab$. Differentiating $f(c)$ with respect to c yields

$$\begin{aligned}f'(c) &= \int_0^\infty \frac{\left(t^2+1\right)}{\left(t^2+1\right)}(-2ce^{-c^2\left(t^2+1\right)}dt\\ &= -2ce^{-c^2}\int_0^\infty e^{-(ct)^2}dt\\ &= -2ce^{-c^2}\frac{\sqrt{\pi}}{2c}\\ &= -e^{-c^2}\sqrt{\pi}.\end{aligned} \tag{3.32}$$

By integrating equation 3.32 with respect to c we get

$$f(c) = -\sqrt{\pi}\left(\frac{1}{2}\sqrt{\pi}\operatorname{erf}(c)\right) + c_0 = -\frac{\pi}{2}\operatorname{erf}(c) + c_0. \tag{3.33}$$

Now, $f(0) = \int_0^\infty \frac{1}{t^2+1}dt = \frac{\pi}{2}$ which implies $c_0 = \frac{\pi}{2}$. substitute $f(c)$ in equation 3.31 gives the wanted result. □

Now we proceed with the proof of the result about the ER graphs.

Proof. We know that the spectral density of $G(n,p)$ converges to the semicircular distribution as $n \to \infty$ and the largest eigenvalue λ_1 of $G(n,p)$ is almost surely

$(1+o(1))np$ provided that $np \gg \ln n$. Then,

$$\begin{aligned} H_{-1}(ER) &= \exp(-(1+\lambda_1)^2) + \sum_{i=2}^{n} \exp(-(1+\lambda_i)^2) \\ &= e^{-(1+\lambda_1)^2} + n\left(\frac{1}{n}\sum_{i=2}^{n} e^{-(1+\lambda_i)^2}\rho(\lambda_i)\right). \end{aligned} \tag{3.34}$$

When $n \to \infty$ we have

$$\begin{aligned} H_{-1}(G(n,p)) &\sim n\int_{-r}^{r} \rho(\lambda) e^{-(1+\lambda)^2} d\lambda \\ &= \frac{2n}{\pi r^2}\int_{-r}^{r} \sqrt{r^2-\lambda^2} e^{-(1+\lambda)^2} d\lambda \end{aligned} \tag{3.35}$$

which can be rewritten using Lemma 3.1.8 as

$$\begin{aligned} H_{-1}(G(n,p)) &\sim \frac{2n}{\pi r^2}\int_{-r}^{r} r^2\sqrt{r^2-1}\frac{e^{-(1+\lambda)^2}}{(1+\lambda)^2+r^2} d\lambda, \ as\, r \to \infty \\ &= \frac{2n}{\pi}\sqrt{r^2-1}\int_{1-r}^{1+r} \frac{e^{-x^2}}{x^2+r^2} dx. \end{aligned} \tag{3.36}$$

By applying Lemma 3.1.9 on the equation (3.36) we have

$$\begin{aligned} H_{-1}(G(n,p)) &\sim \frac{4n}{\pi}\sqrt{r^2-1}\int_{0}^{\infty} \frac{e^{-x^2}}{x^2+r^2} dx \\ &= \frac{2n}{r}\sqrt{r^2-1} e^{r^2} \mathrm{erfc}(r). \end{aligned} \tag{3.37}$$

□

We now obtain the formula for the H_{-1} index for the Barabási-Albert (BA) random graphs.

Theorem 3.1.5. Let G be a BA random network. Then, when $n \to \infty$, the H_{-1} index of a BA network is

$$\begin{aligned} H_{-1}(G) \sim \frac{n}{r^2}\Bigg(& e^{-1-r^2}\cosh(2r) + \frac{\sqrt{\pi}}{2}\left((1-r)\mathrm{erf}(1-r) + (1+r)\mathrm{erf}(1+r)\right) \\ & - \sqrt{\pi}\mathrm{erf}(1) - e^{-1}\Bigg) \end{aligned} \tag{3.38}$$

where $r = 2\sqrt{np(1-p)}$ and $\mathrm{erf}\,(\cdots)$ is the error function.

Proof. Let G be a BA random network, then

$$H_{-1}(G) = \sum_{j=1}^{n} e^{-(1+\lambda_j)^2} \tag{3.39}$$

as $n \to \infty$, we have

$$\begin{aligned} H_{-1}(G) &\sim n \int_{-r}^{r} \rho(\lambda) e^{-(1+\lambda)^2} d\lambda \\ &= n \left(\int_{-r}^{0} \frac{r+\lambda}{r^2} e^{-(1+\lambda)^2} d\lambda + \int_{0}^{r} \frac{r-\lambda}{r^2} e^{-(1+\lambda)^2} d\lambda \right) \end{aligned} \tag{3.40}$$

Making the substitution $x = 1 + \lambda$ and put $a = 1 - r$, $b = 1 + r$, the last equation becomes

$$\begin{aligned} H_{-1}(G) \sim \frac{n}{r^2} \Bigg(&\int_{a}^{0} (x-a) e^{-x^2} dx + \int_{0}^{b} (b-x) e^{-x^2} dx \\ &+ \int_{0}^{1} (x-a) e^{-x^2} dx - \int_{0}^{1} (b-x) e^{-x^2} dx \Bigg) \\ = \frac{n}{r^2} &\left(S_1 + S_2 + S_3 + S_4 \right). \end{aligned} \tag{3.41}$$

Now we have:

$$\begin{aligned} S_1 &= \int_{a}^{0} x e^{-x^2} dx - a \int_{a}^{0} e^{-x^2} dx \\ &= \left[-\frac{1}{2} e^{-x^2} \right]_{a}^{0} + a \int_{0}^{a} e^{-x^2} dx \\ &= -\frac{1}{2} + \frac{1}{2} e^{-a^2} + \frac{\sqrt{\pi} a}{2} \left[\mathrm{erf}(x) \right]_{0}^{a} \\ &= -\frac{1}{2} + \frac{1}{2} e^{-a^2} + \frac{\sqrt{\pi} a}{2} \mathrm{erf}(a). \end{aligned} \tag{3.42}$$

Moreover,

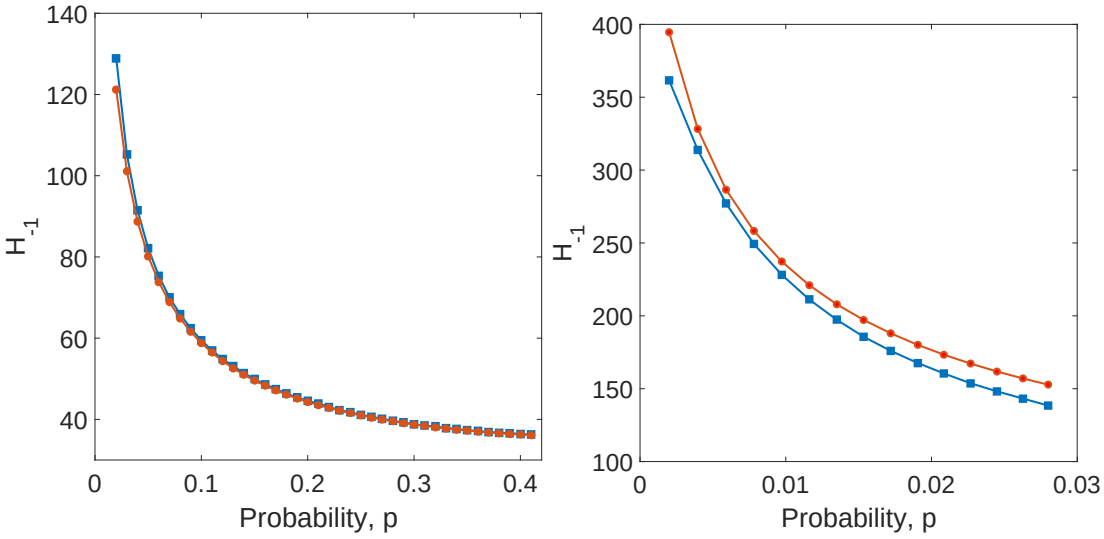
$$S_2 = -\frac{1}{2} + \frac{1}{2}e^{-b^2} + \frac{\sqrt{\pi}b}{2}\text{erf}(b). \tag{3.43}$$

$$S_3 = \frac{1}{2} - \frac{1}{2}e^{-1} - \frac{\sqrt{\pi}a}{2}\text{erf}(1). \tag{3.44}$$

$$S_4 = \frac{1}{2} - \frac{1}{2}e^{-1} - \frac{\sqrt{\pi}b}{2}\text{erf}(1). \tag{3.45}$$

By substituting the values into equation (3.41) we have the final result. □

In Figure 3.4 we compare the values of the H_{-1} index calculated by using the function "expm" of Matlab with those obtained by the formulas 3.27 and 3.38 for the ER and BA random graphs, respectively. In this case we fixed the number of nodes to $n = 1000$ and change the values of p. As can be seen the results for the ER are very good with a complete correspondence between the "exact" values and those obtained by the formula 3.27. In the case of the BA graphs the fit between both approaches is not so good. As mentioned previously, we are assuming that the eigenvalues of a BA network follows the triangular distribution found by [35] for BA networks constructed by using $2 \leq m \leq 5$. Thus, outside this regime there are deviations from the triangular distribution of eigenvalues that obviously affect our prediction.



(a) (b)

Figure 3.4: Comparison of the index H_{-1} obtained by using the formulas 3.27 and 3.38 for the ER (a) and BA (b) graphs, respectively, with the "exact" values obtained by using the function 'expm' from Matlab. The results from the formulas 3.27 and 3.38 are represented by filled circles and those from the function 'expm' are represented as filled squares. All the results are the average of 100 random realizations.

Chapter 4

The Double Gaussian Communicability Function $e^{-(A^2-I)^2}$

In several scenarios, for instance in the study of conjugated molecules, the importance of pairs of specific eigenvalues may be relevant for understanding the structure and function of these systems. When studying molecules with the tight-binding Hamiltonians, e.g., the Hückel molecular orbital (HMO) approach [4], there are two eigenvalues of the graph spectra which play a fundamental role in understanding molecular properties. The highest occupied (HOMO) and the lowest unoccupied (LUMO) molecular orbitals, respectively [122]. These "molecular orbitals" are schematically illustrated in Figure 4.1 (left panel) where we indicate their importance as electron donor and acceptor, respectively. Our goal in this Thesis is then to "fold" the graph spectra such that two eigenvalues, like for instance the HOMO and LUMO, have the largest contribution to the corresponding matrix function. Therefore, we define double Gaussian functions of the graph spectra:

$$\tilde{G}\left(\lambda_{\mathrm{ref}_1}, \lambda_{\mathrm{ref}_2}\right) = \exp\left[-\left(\lambda_{\mathrm{ref}_1} I - A\right)^2 \left(\lambda_{\mathrm{ref}_2} I - A\right)^2\right]. \tag{4.1}$$

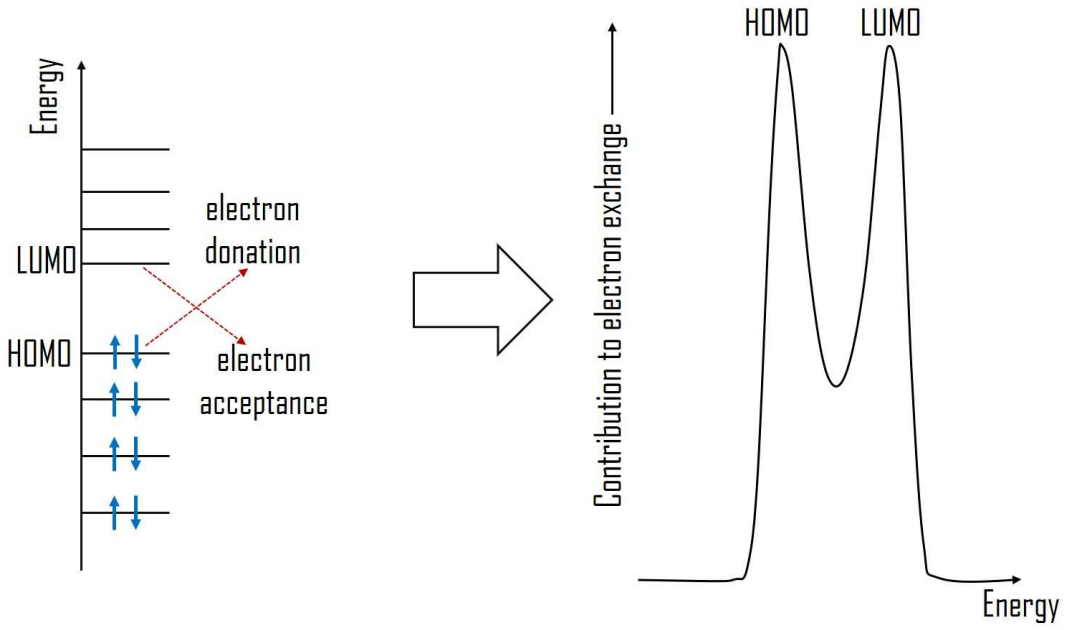


Figure 4.1: Scheme illustrating the double-Gaussian transformation of the spectrum of a graph representing the energy levels of a molecule.

The schematic process of the double Gaussianization of the graph spectra is illustrated in Figure 4.2. In the case of alternant conjugated molecules with n atoms and graph eigenvalues $\lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$, the HOMO/LUMO correspond to the eigenvalue $\mp\lambda_{n/2}$, respectively. Therefore, here we will focus on the case in which $\lambda_{\text{ref}_1} = -\lambda_{\text{ref}_2}$, but the formulation is general enough as to consider any further case. Additionally, Fowler and Pisanski [123] has called "normal" the molecular graphs for which $+1 \geq \lambda_{HOMO} \geq \lambda_{LUMO} \geq -1$, while the rest of molecular graphs are called "exceptional". The reason for this is that most of molecular graphs have their HOMO and LUMO within the 'chemical triangle' of an HOMO-LUMO map [123]–a scatterplot of the middle eigenvalues of the graph–, with vertices at $(-1,-1)$, $(+1,-1)$, $(+1,+1)$. They proved that all chemical trees lie within the triangle, as do all chemical graphs with up to 12 vertices [123]. Therefore, we will focus here on the case $\lambda_{\text{ref}_1} = 1$, $\lambda_{\text{ref}_2} = -1$.

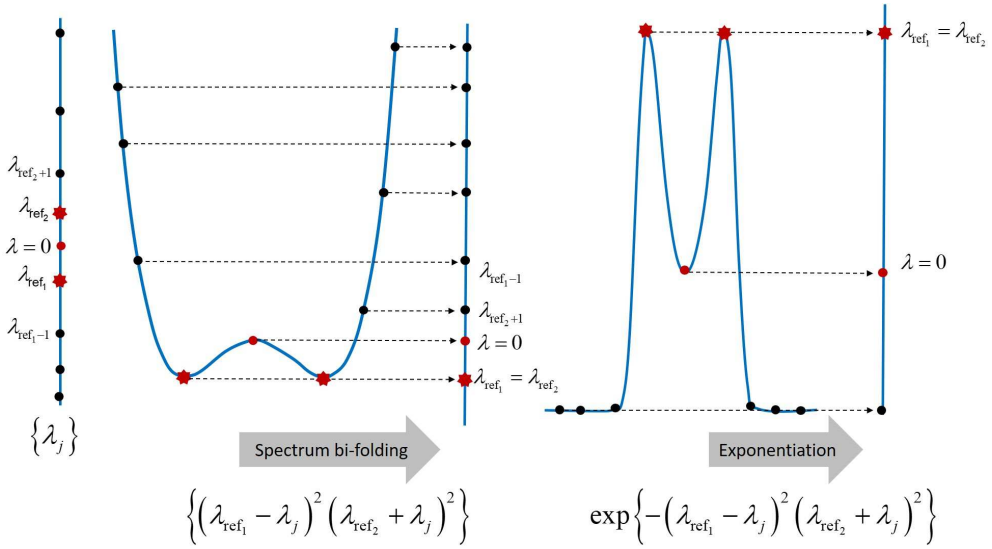


Figure 4.2: Schematic illustration of the double Gaussianization of thge graph spectra. In the extreme left we illustrate the graph spectra where the eigenvalues are represented as dots in a vertical line. In the central panel we illustrate the bifolding of the spectrum where the reference eigenvalues occupy the lowest position in a vertical line. Finally (extreme right) we exponentiate the bifolded spectrum and the reference eigenvalues make the highest contribution to the matrix function.

Then, we study the function

$$\begin{aligned}\tilde{G}(-1,1) &= \exp\left[-\left((-1)I-A\right)^2(I-A)^2\right] \\ &= \exp\left[-\left(A^2-I\right)^2\right],\end{aligned} \tag{4.2}$$

and in particular the corresponding double Gaussian Estrada index of this function:

$$\begin{aligned}H_{-1,1} &= tr\tilde{G}(-1,1) \\ &= \sum_{j=1}^{n} e^{-\left(\lambda_j^2-1\right)^2},\end{aligned} \tag{4.3}$$

4.1 $H_{-1,1}$ Index of Graphs

4.1.1 $H_{-1,1}$ Index for Graphs with All but Two Eigenvalues Equal to ± 1

Three infinite families of connected graphs have been reported [124] to have eigenvalues $r > 1$ and $s < -1$, and all other eigenvalues equal to ± 1. The adjacency matrix and spectra of these families are as follows. Let O be an all-zeros matrix, J an all-ones matrix and I_h the $h \times h$ identity matrix. Let R_{2k} be the adjacency matrix of k copies of K_2, i.e., the disjoint union of k edges.

Theorem 4.1.1. [124] The infinite families of graphs having the following adjacency matrices and spectra are the only ones having all but two eigenvalues equal to ± 1:

(i) $\begin{bmatrix} O & J-I_m \\ J-I_m & O \end{bmatrix}$ $(m \geq 3)$ with spectrum $\left\{\pm(m-1), 1^{m-1}, -1^{m-1}\right\}$;

(ii) $\begin{bmatrix} J-I_a & J \\ J & R_{2k} \end{bmatrix}$ $(a \geq 1,\, k \geq 2)$ with spectrum $\left\{\frac{a}{2} \pm \frac{1}{2}\sqrt{a^2+8ak-4a+4}, 1^{k-1}, -1^{a+k-1}\right\}$.
When $a = 1$, the resulted family is the frindship graphs;

(iii) $\begin{bmatrix} R_{2\ell} & J \\ J & R_{2m} \end{bmatrix}$ $(\ell \geq m \geq 2)$ with spectrum $\left\{1 \pm 2\sqrt{\ell m}, 1^{\ell+m-2}, -1^{\ell+m}\right\}$.

Then we have the following result.

Theorem 4.1.2. The $H_{-1,1}$ index of the previous families is presented as follows

(i)
$$H_{-1,1} = 2e^{-m^2(m-2)^2} + 2m - 2. \tag{4.4}$$

(ii)
$$H_{-1,1} = e^{-\frac{a^2}{4}(a+4k+b-2)^2} + e^{-\frac{a^2}{4}(a+4k-b-2)^2} + (a+2k) - 2, \tag{4.5}$$

where $b = \sqrt{a^2+8ak-4a+4}$.

(iii)
$$H_{-1,1} = e^{-16\left(\ell m+\sqrt{\ell m}\right)^2} + e^{-16\left(\ell m-\sqrt{\ell m}\right)^2} + 2\left(\ell + m\right) - 2. \tag{4.6}$$

Proof. (i)

$$\begin{aligned} H_{-1,1} &= \sum_{j=1}^{n} e^{-\left(\lambda_j^2-1\right)^2} \\ &= e^{-\left((m-1)^2-1\right)^2} + e^{-\left((1-m)^2-1\right)^2} + m - 1 + m - 1 \\ &= e^{-m^2(m-2)^2} + e^{-m^2(m-2)^2} + 2m - 2 \\ &= 2e^{-m^2(m-2)^2} + 2m - 2. \end{aligned} \tag{4.7}$$

(ii) let $b = \sqrt{a^2 + 8ak - 4a + 4}$ for more simplification, then

$$\begin{aligned} H_{-1,1} &= \sum_{j=1}^{n} e^{-\left(\lambda_j^2-1\right)^2} \\ &= e^{-\left(\left(\frac{a}{2}+\frac{1}{2}b\right)^2-1\right)^2} + e^{-\left(\left(\frac{a}{2}-\frac{1}{2}b\right)^2-1\right)^2} + k - 1 + a + k - 1 \\ &= e^{-\frac{a^2}{4}(a+4k+b-2)^2} + e^{-\frac{a^2}{4}(a+4k-b-2)^2} + (a + 2k) - 2. \end{aligned} \tag{4.8}$$

(iii)

$$\begin{aligned} H_{-1,1} &= \sum_{j=1}^{n} e^{-\left(\lambda_j^2-1\right)^2} \\ &= e^{-\left(\left(1+2\sqrt{\ell m}\right)^2-1\right)^2} + e^{-\left(\left(1-2\sqrt{\ell m}\right)^2-1\right)^2} + \ell + m - 2 + \ell + m \\ &= e^{-16\left(\ell m+\sqrt{\ell m}\right)^2} + e^{-16\left(\ell m-\sqrt{\ell m}\right)^2} + 2\left(\ell + m\right) - 2. \end{aligned}$$

□

4.1.2 $H_{-1,1}$ Index of Simple Graphs

Here we prove some results for simple graphs which may be useful in understanding further structure-spectra relations in general graphs.

Lemma 4.1.1. Let K_n be the complete graph of n nodes. Then

$$H_{-1,1}\left(K_n\right) = n - 1 + e^{-n^2(n-2)^2}. \tag{4.9}$$

Proof. The spectrum of K_n is $\sigma(K_n) = \left\{[n-1]^1, [-1]^{n-1}\right\}$ so we have

$$\begin{aligned} H_{-1,1}(K_n) &= \sum_{j=1}^{n} e^{-\left(\lambda_j^2-1\right)^2} \\ &= (n-1)e^0 + e^{-\left((n-1)^2-\right)1^2} \\ &= n-1+e^{-n^2(n-2)^2}. \end{aligned}$$

□

Lemma 4.1.2. Let K_{n_1,n_2} be the complete bipartite graph of n_1+n_2 nodes. Then

$$H_{-1,1}(K_{n_1,n_2}) = \frac{n_1+n_2-2}{e} + 2e^{-(n_1n_2-1)^2}. \tag{4.10}$$

Proof. The spectrum of K_{n_1,n_2} is $\sigma(K_{n_1,n_2}) = \left\{\left[\sqrt{n_1n_2}\right]^1, \left[-\sqrt{n_1n_2}\right]^1, [0]^{n_1+n_2-2}\right\}$ so we have

$$\begin{aligned} H_{-1,1}(K_{n_1,n_2}) &= \sum_{j=1}^{n_1+n_2} e^{-\left(\lambda_j^2-1\right)^2} \\ &= e^{-(n_1n_2-1)^2} + e^{-(n_1n_2-1)^2} + (n_1+n_2-2)\,e^{-1} \\ &= \frac{n_1+n_2-2}{e} + 2e^{-(n_1n_2-1)^2}. \end{aligned} \tag{4.11}$$

□

Corollary 4.1.1. Let $K_{1,n-1}$ be the star graph of n nodes. Then

$$H_{-1,1}(K_{1,n-1}) = \frac{n-2}{e} + 2e^{-(n-2)^2}. \tag{4.12}$$

Lemma 4.1.3. Let P_n be a path having n nodes. Then, asymptotically as $n \to \infty$

$$H_{-1,1}(P_n) \sim \frac{n+1}{\pi}\int_0^{\pi} e^{-(2\cos\theta+1)^2} d\theta - e^{-9}. \tag{4.13}$$

Proof. By following the same approximating method in the proof of Lemma 3.1.6 we

have

$$\begin{aligned} H_{-1,1}(P_n) &= \sum_{j=1}^{n} e^{-\left(\lambda_j^2-1\right)^2} \\ &= \sum_{j=1}^{n} e^{-\left(4\cos^2\left(\frac{j\pi}{n+1}\right)-1\right)^2} \\ &= \sum_{j=1}^{n} e^{-\left(1+2\cos\left(\frac{2j\pi}{n+1}\right)\right)^2} \\ &= \frac{1}{2}\sum_{j=0}^{n} e^{-\left(1+2\cos\left(\frac{2j\pi}{n+1}\right)\right)^2} + \frac{1}{2}\sum_{j=1}^{n+1} e^{-\left(1+2\cos\left(\frac{2j\pi}{n+1}\right)\right)^2} - \frac{1}{2}e^{-9} - \frac{1}{2}e^{-9} \\ &= \frac{1}{2}\sum_{j=0}^{n} e^{-\left(1+2\cos\left(\frac{2j\pi}{n+1}\right)\right)^2} + \frac{1}{2}\sum_{j=1}^{n+1} e^{-\left(1+2\cos\left(\frac{2j\pi}{n+1}\right)\right)^2} - e^{-9}. \end{aligned} \tag{4.14}$$

Now, when $n \to \infty$ the summation in (4.14) can be approached by the following integral

$$\begin{aligned} H_{-1,1}(P_n) &\sim \frac{1}{2}\frac{n+1}{\pi}\int_0^{\pi} e^{-(2\cos\theta+1)^2}d\theta + \frac{1}{2}\frac{n+1}{\pi}\int_0^{\pi} e^{-(2\cos\theta+1)^2}d\theta - e^{-9} \\ &= \frac{n+1}{\pi}\int_0^{\pi} e^{-(2\cos\theta+1)^2}d\theta - e^{-9}. \end{aligned} \tag{4.15}$$

□

Lemma 4.1.4. Let C_n be a cycle having n nodes. Then, asymptotically as $n \to \infty$

$$H_{-1,1}(C_n) \sim \frac{n}{\pi}\int_0^{\pi} e^{-(2\cos\theta+1)^2}d\theta. \tag{4.16}$$

Proof. By following the same approximating method in the proof of Lemma 3.1.7 we have

$$\begin{aligned} H_{-1,1}(C_n) &= n\left(\frac{tr\left(e^{-\left(A^2-I\right)^2}\right)}{n}\right) \\ &= n\left(\frac{1}{n}\sum_{j=1}^{n} e^{-\left(4\cos^2\left(\frac{2\pi j}{n}\right)-1\right)^2}\right) \end{aligned}$$

$$= n\left(\sum_{j=1}^{n}\frac{1}{n}e^{-\left(2+2\cos\left(\frac{4\pi j}{n}\right)-1\right)^2}\right)$$
$$= n\left(\sum_{j=1}^{n}\frac{1}{n}e^{-\left(1+2\cos\frac{4\pi j}{n}\right)^2}\right). \tag{4.17}$$

Now, when $n\to\infty$ the summation in 4.17 can be approached by the following integral

$$\begin{aligned} H_{-1,1}(C_n) &\sim n\frac{1}{4\pi}\int_0^{4\pi} e^{-(1+2\cos\theta)^2}d\theta \\ &= n\frac{1}{4\pi}(2)\int_0^{2\pi} e^{-(1+2\cos\theta)^2}d\theta \\ &= n\frac{1}{4\pi}(2)(2)\int_0^{\pi} e^{-(1+2\cos\theta)^2}d\theta \end{aligned}$$

where $\theta=\frac{4j\pi}{n}$. Thus, when $n\to\infty$ we have

$$H_{-1,1}(C_n)\sim\frac{n}{\pi}\int_0^{\pi} e^{-(1+2\cos\theta)^2}d\theta. \tag{4.18}$$

□

Corollary 4.1.2. Asymptotically, as $n\to\infty$,

1. $H_{-1,1}(P_n)\sim H_{-1}(P_n)$.
2. $H_{-1,1}(C_n)\sim H_{-1}(C_n)$.
3. $H_{-1,1}(K_n)\sim H_{-1}(K_n)$.
4. $H_{-1,1}(K_{n_1,n_2})\sim H_{-1}(K_{n_1,n_2})$.
5. $H_{-1,1}(K_{1,n-1})\sim H_{-1}(K_{1,n-1})$.

Proof. Let us write the following limits of the ratios of both indices:

1)
$$\lim_{n\to\infty}\frac{H_{-1,1}(C_n)}{H_{-1}(C_n)}=\frac{\frac{n}{\pi}\int_0^{\pi}e^{-(2\cos\theta+1)^2}d\theta}{\frac{n}{\pi}\int_0^{\pi}e^{-(2\cos\theta+1)^2}d\theta}=1.$$

2)
$$\lim_{n\to\infty}\frac{H_{-1,1}(P_n)}{H_{-1}(P_n)}=\frac{\frac{n+1}{\pi}\int_0^{\pi}e^{-(2\cos\theta+1)^2}d\theta-e^{-9}}{\frac{n+1}{\pi}\int_0^{\pi}e^{-(2\cos\theta+1)^2}d\theta-\frac{1}{2}(e^{-1}+e^{-9})}=1$$

3)

$$\lim_{n\to\infty}\frac{H_{-1,1}\left(K_n\right)}{H_{-1}\left(K_n\right)}=\frac{n-1+e^{-n^2(n-2)^2}}{n-1+e^{-n^2}}=1.$$

4) when $n\to\infty$ we have also $n_1+n_2=n\to\infty$ and $n_1n_2\to\infty$, thus

$$\begin{aligned}\lim_{n\to\infty}\frac{H_{-1,1}\left(K_{n_1,n_2}\right)}{H_{-1}\left(K_{n_1,n_2}\right)} &= \lim_{n\to\infty}\frac{\frac{n_1+n_2-2}{e}+2e^{-(n_1n_2-1)^2}}{\frac{n_1+n_2-2}{e}+e^{-n_1n_2-1}\left(\frac{e^{2\sqrt{n_1n_2}}+e^{-2\sqrt{n_1n_2}}}{2}\right)}\\ &= \lim_{n\to\infty}\frac{\frac{n_1+n_2-2}{e}+2e^{-(n_1n_2-1)^2}}{\frac{n_1+n_2-2}{e}+\left(\frac{e^{2\sqrt{n_1n_2}-(n_1n_2+1)}+e^{-2\sqrt{n_1n_2}-(n_1n_2+1)}}{2}\right)}\\ &= 1\end{aligned}$$

5) We proved the general case in (4). □

4.1.3 Extremal Graphs for $H_{-1,1}$ Index

Let us start here by stating a result from Cioabă et al. [124]. Define $\mathcal{G}$ to be the set of connected graphs with eigenvalues $r>1$ and $s<-1$, and all other eigenvalues equal to ±1. Then, Cioabă et al. [124] proved the following result.

Lemma 4.1.5. No graph in $\mathcal{G}$ has one of the graphs presented in Figure 4.3 as an induced subgraph.

We calculated the $H_{-1,1}$ index for all 11,117 connected graphs with 8 nodes and determined those with the largest values of the index. These graphs are illustrated in Figure 4.4. The largest value of $H_{-1,1}$ is obtained for the complete graph K_8 (not illustrated in the Figure 4.4). We have verified that for graphs $G_{n\leq8}$, $H_{-1,1}\left(G_n\right)<H_{-1,1}\left(K_n\right)$. Therefore we have the following.

Conjecture 4.1.1. Let G be any connected graph of n nodes, then

$$H_{-1,1}\left(G\right)\leq H_{-1,1}\left(K_n\right). \tag{4.19}$$

In addition, none of the graphs in Figure 4.4 contains any of the graphs in Figure 4.3 as an induced subgraph. We then explore the graphs with the smallest values of $H_{-1,1}$ among all 11,117 connected graphs with 8 nodes. the 10 ones with the smallest

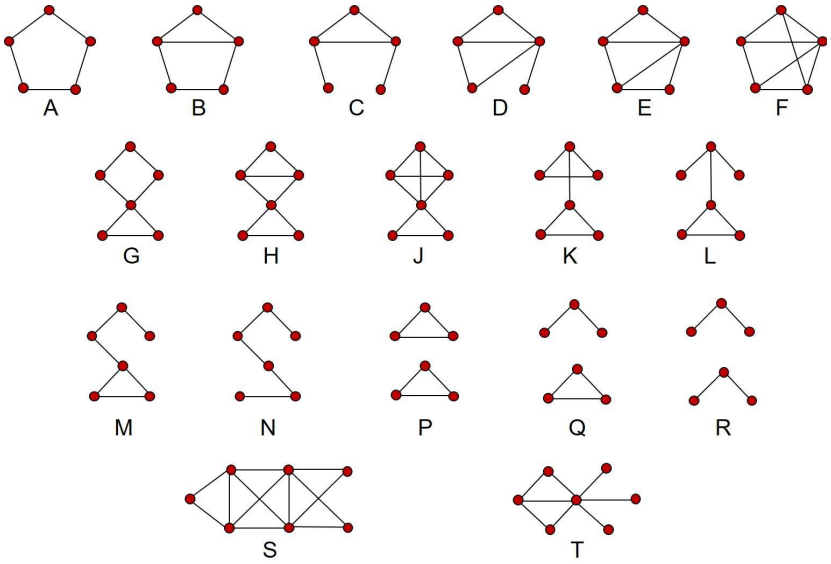


Figure 4.3: Illustration of the forbidden induced subgraphs found by Cioabă et al. We use the same labeling as in the paper of Cioabă et al.

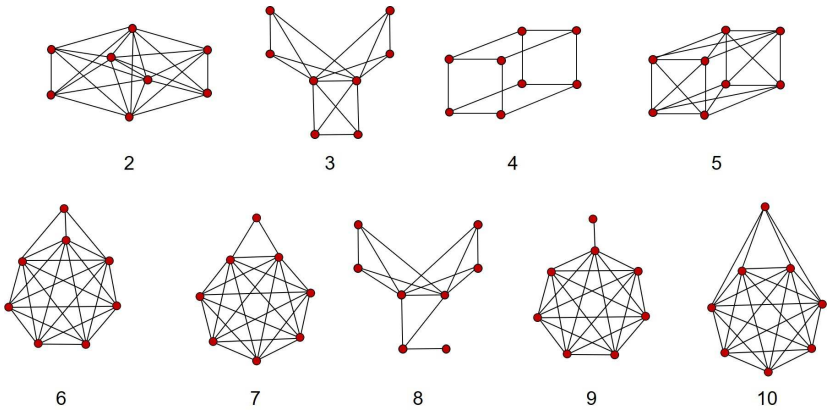


Figure 4.4: Illustration of the 10 graphs (K_8 is the number 1, which is omitted) with the largest values of $H_{-1,1}(G)$ among all connected graphs with 8 nodes.

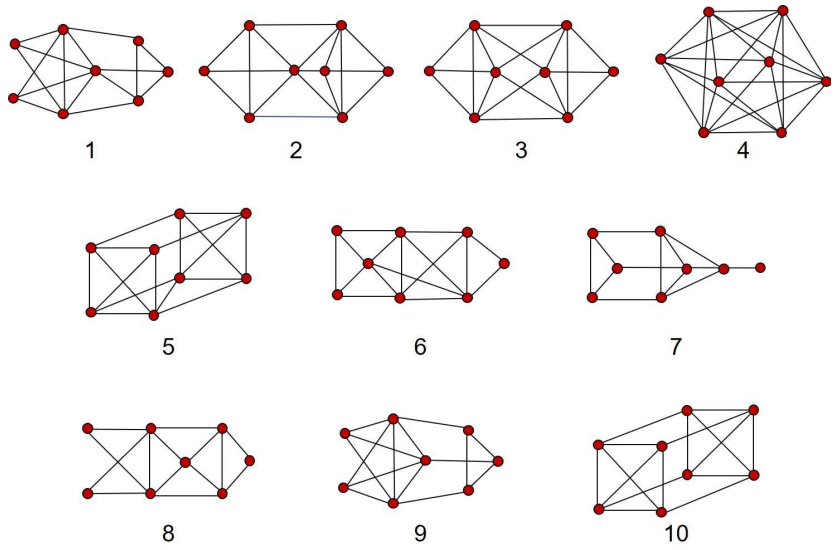


Figure 4.5: Illustration of the 10 graphs with the minimum values of $H_{-1,1}(G)$ among all connected graphs with 8 nodes.

values of this index are illustrated in Figure 4.5. We have calculated the number of each of the forbidden induced subgraphs in these 10 graphs displaying the minimum values of $H_{-1,1}(G)$. We have found that 11 out of the 18 forbidden induced subgraphs appear very frequently in these 10 graphs. The results are illustrated in Figure 4.6. for instance, the graph with the least value of $H_{-1,1}(G)$ has the forbidden induced sugraph B 8 times, D 6 times and L 2 times. Others, like graph 10 in Figure 4.6 contains only one forbidden subgraph, i.e., subgraph B 24 times.

Lemma 4.1.6. Let G be connected bipartite graph of n nodes, then

$$H_{-1,1}(G) \leq H_{-1,1}(K_n). \tag{4.20}$$

Proof. For any graph G we have $\lambda_1 \geq d_{ave}$ where λ_1 is the principal eigenvalue and d_{ave} is the average degree of the graph G [39]. Then we have

$$\begin{aligned} \lambda_1 \geq d_{ave} &= \frac{2m}{n} \\ &\geq \frac{2(n-1)}{n}. \end{aligned} \tag{4.21}$$

Suppose that $n \geq 5$ (it is easy to check that the statement is true for all graphs of

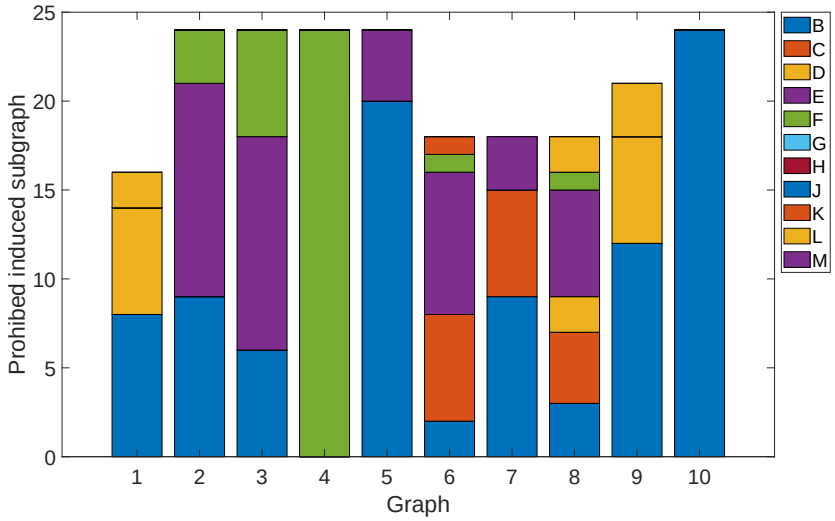


Figure 4.6: Frequency with which some of the forbidden induced subgraphs appear in the connected graphs with 8 nodes which display the minimum values of the index $H_{-1,1}(G)$. The induced subgraphs are given in Figure 4.3 and graphs are shown in Figure 4.5. The forbidden induced subgraphs not depicted in the figure do not appear in the graphs considered.

nodes less than 5) then we have $\lambda_1 \geq \frac{2(n-1)}{n} \geq \frac{2(4)}{5} = 1.6$. Thus, $(\lambda_1^2 - 1)^2 \geq 2.4336$ and that implies $\exp\left[-(\lambda_1^2-1)^2\right] \leq \exp(-2.4336)$. If G is bipartite, then we will have a symmetry in the spectra of G and we get $\lambda_n \leq -1.6$. Following the same steps we end with $\exp\left[-(\lambda_n^2 - 1)^2\right] \leq \exp(-2.4336)$. Now

$$\begin{aligned} H_{-1,1}(G) &= \sum_{j=1}^{n} e^{-\left(\lambda_j^2-1\right)^2} = \sum_{j=2}^{n-1} e^{-\left(\lambda_j^2-1\right)^2} + e^{-\left(\lambda_1^2-1\right)^2} + e^{-\left(\lambda_n^2-1\right)^2} \\ &\leq n - 2 + 2e^{-2.4336} \\ &< n - 2 + 1 = n - 1 \leq n - 1 + e^{-n^2(n-2)^2} = H_{-1,1}(K_n). \end{aligned}$$

□

4.1.4 $H_{-1,1}$ Index of Random Networks

In this section, we give a bound for the $H_{-1,1}$ index in two models of random graphs. We start by the following theorem that sets a bound for the index in ER random graphs.

Theorem 4.1.3. For an Erdős-Rényi random graph $G(n,p)$ with $\frac{\ln n}{n} \ll p$, the $H_{-1,1}(ER)$

is bounded as

$$H_{-1,1}\left(ER\right) \leq \frac{4n}{\pi r}. \tag{4.22}$$

almost surely, as $n \to \infty$, where $r = 2\sqrt{np(1-p)}$.

Proof. As $n \to \infty$ we have

$$\begin{aligned} H_{-1,1}\left(ER\right) &= n\int_{-r}^{r} \rho(\lambda) e^{-(\lambda^2-1)^2} d\lambda \\ &= \frac{4n}{\pi r^2}\int_{0}^{r} \sqrt{r^2-\lambda^2} e^{-(\lambda^2-1)^2} d\lambda \\ &\leq \frac{4n}{\pi r^2}\int_{0}^{r} \sqrt{r^2} e^{-(\lambda^2-1)^2} d\lambda \\ &\leq \frac{4n}{\pi r}\int_{0}^{\infty} e^{-(\lambda^2-1)^2} d\lambda \\ &\leq \frac{4n}{\pi r}(1) = \frac{4n}{\pi r}. \end{aligned} \tag{4.23}$$

□

Now we give the bound for BA random graphs.

Theorem 4.1.4. Let G be a BA random network. Then, when $n \to \infty$, the $H_{-1,1}(BA)$ is bounded as

$$H_{-1,1}\left(BA\right) \leq \frac{2n}{r}, \tag{4.24}$$

where $r = 2\sqrt{np(1-p)}$.

Proof. We know that the density of BA graphs follows a triangular distribution. Thus

$$\begin{aligned} H_{-1,1}\left(BA\right) &= \sum_{j=1}^{n} \rho(\lambda_j) e^{-(\lambda_j^2-1)^2} \\ &= n\left(\frac{1}{n}\sum_{j=1}^{n} \rho(\lambda_j) e^{-(\lambda_j^2-1)^2}\right). \end{aligned} \tag{4.25}$$

As $n \to \infty$ we have

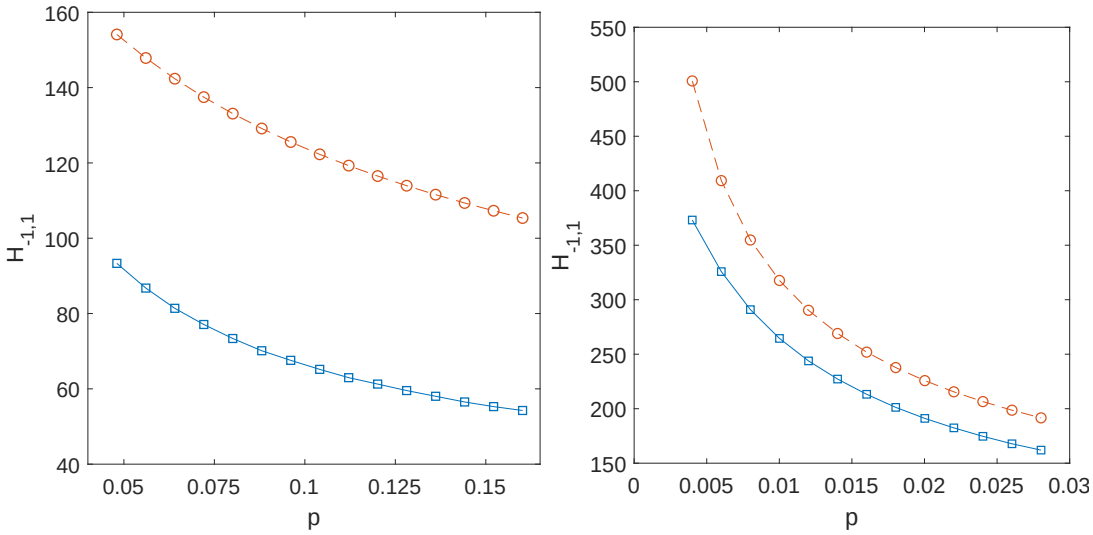
$$\begin{aligned} H_{-1,1}(BA) &= \int_{-r}^{r} \rho(\lambda) e^{-(\lambda^2-1)^2} d\lambda \\ &= n\left(\int_{-r}^{0} \frac{\lambda + r}{r^2} e^{-(\lambda^2-1)^2} d\lambda + \int_{0}^{r} \frac{r - \lambda}{r^2} e^{-(\lambda^2-1)^2} d\lambda \right) \\ &= \frac{n}{r^2}\left(2\int_{0}^{r} r e^{-(\lambda^2-1)^2} d\lambda - 2\int_{0}^{r} \lambda e^{-(\lambda^2-1)^2} d\lambda \right) \\ &= \frac{2n}{r^2}\left(r\int_{0}^{r} e^{-(\lambda^2-1)^2} d\lambda - \frac{1}{2}\int_{0}^{r^2-1} e^{-\lambda^2} d\lambda \right) \\ &\leq \frac{2n}{r^2}\left(r\int_{0}^{r} e^{-(\lambda^2-1)^2} d\lambda \right) \\ &\leq \frac{2n}{r}. \end{aligned} \tag{4.26}$$

□

In Figure 4.7 we compare the values of the $H_{-1,1}$ index calculated by using the function "expm" of Matlab with the bounds obtained in theorems 4.1.4 and 4.1.3 for the ER and BA random graphs, respectively. In this case we fixed the number of nodes to $n = 1000$ and change the values of p.

4.2 Carcinogenicity of Polycyclic Aromatic Hydrocarbons

Polycyclic aromatic hydrocarbns (PAHs) are compounds formed by carbon in fused hexagonal shapes and hydrogen, for which the eigenvalues ± 1 play an important role [114]. The excessive exposure to PAHs may result in cancer in humans. The general mechanism by which PAHs produce cancer is by their metabolic activation which leads to the formation of the active carcinogens like diol-epoxides, radical cations, and o-quinones (see first line in Figure 4.8) [125]. These metabolites then react with DNA forming DNA adducts which results in DNA mutations, alteration of gene expression profiles, and tumorigenesis (see second line in Figure 4.8). The metabolic activation of PAHs depends on the chemical reactivity of these compounds, and their electron donation/acceptance capacities, which are mainly determined by their HOMO and LUMO. Then, it is not strange to find reports on the use of these frontier molecular



(a) (b)

Figure 4.7: (a) Change of the H index with the increase of the probability p in ER random graphs $G_{ER}(1000, p)$ obtained using the function 'expm' in Matlab (squares and solid line) and the bound obtained in theorem 4.1.3 (circles and broken line). (b) Change of the H index with the increase of the probability p in ER random graphs $G_{BA}(1000, m_0)$ obtained using the function 'expm' in Matlab (squares and solid line) and the bound obtained in theorem 4.1.4 (circles and broken line). All the calculations are the average of 100 random realizations.

orbitals or electronic parameters like superdelocalizability in explaining the carcinogenic power of PAHs [126]. However, because chemical reactions occur at some specific atoms in a molecule, different atomic regions may have distinct contributions to the carcinogenicity of PAHs. This has been widely recognized in the literature where four main atomic regions have been identified with different contributions to the carcinogenic activity of PAHs. These regions are defined by the number of CH bonds in the ring and known as K, L, M and N. An L region is a ring with single CH bonds whereas K region is the ring with HC−CH bond. N regions are identified by the HC−CH−CH bond and M region is the ring with HC−CH−CH−CH bond. All regions are illustrated in Figure 4.8. The regions K and L where proposed by Pullman and Pullman [127] and have proved to be predictive for the carcinogenicity of a large number of PAHs. Due to more recent findings the other two regions, M and N, where proposed and studied in quantitative structure-carcinogenicity activity of PAHs for instance by Vijayalakshmi and Suresh [128]. Here we use the series of 28 PAHs for which the carcinogenic power has been reported and studied by Vijayalakshmi and Suresh [128]. The list of PAHs

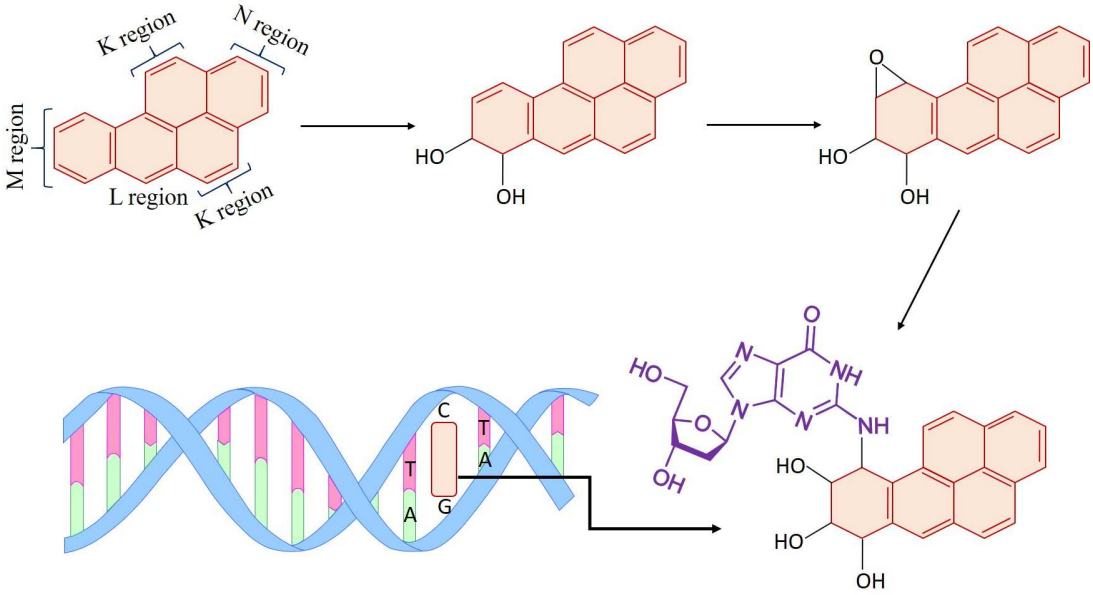


Figure 4.8: Schemetaic illustration of the metabolic activation of a PAH (first line) and the reaction of the reactive metabolite with a DNA base producing alteraions in DNA (second line).

and their carcinogenic activity (CA) is reported in Table 4.1. We consider here the $H_{-1,1}$ index split as follows:

$$H_{-1,1} = H_{-1,1}(K) + H_{-1,1}(L) + H_{-1,1}(M) + H_{-1,1}(N) + H_{-1,1}(F), \tag{4.27}$$

where $H_{-1,1}(K)$ is the sum of the contributions of the atoms in the region K to the global $H_{-1,1}$ index, and the term F is used for the atoms in the frame of the PAHs, i.e., those not in any of the four mentioned regions. We recall that the contribution of an atom p to the $H_{-1,1}$ index is:

$$H_{-1,1}(p) = \sum_{j=1}^{n} \psi_j^2(p) \exp\left(-\left(\lambda_j^2 - 1\right)^2\right). \tag{4.28}$$

We use here the average of the atomic contributions for each region $\bar{H}_{-1,1}$ whose values for the 28 PAHs analyzed are given in Table 4.1. We grouped the carcinogenic activity (CA) of these 28 PAHs into two categories, which correspond to the class I which groups inactive and the class A, which groups PAHs with CA ranging from + to ++++ (see Table 4.1). the main reason is that a classification based on the strength of the

carcinogenic activity is impossible as some of the classes contain only one member, e.g., CA +++. We now focus in classification techniques that allow to split the set of PAHs into the two groups devised here on the basis of the regional $\bar{H}_{-1,1}$ indices. For that purpose we explore the use of discriminant analysis, classification trees and K-nearest neighbors (KNN) techniques, all implemented in the "classification learner" toolbox of Matlab R2018b. In all cases we observe that the use of the $\bar{H}_{-1,1}$ indices for the K and L regions are enough for the classification of these compounds, with no improvement by adding information about M and N regions. Linear discriminant analysis (LDA) classifies correctly 85.7% of PAHs in the two classes. From the carcinogenic compounds it classifies correctly 87.5% of PAHs and 83.3% of inactive ones. In Figure 4.10 (a and b) we illustrate the results for the LDA. The classification tree improves a little bit the previous results and classifies correctly 93.75% of carcinogenic PAHs (see Figrue 4.10 (c and d). The best results are obtained by using KNN which classifies correctly 100% of compounds in the two classes as illustrated in Figure 4.10 (e and f). Our analysis has no exception like in the case of Vijayalakshmi et al. [128] for which 5 PAHs were excluded from the analysis as outliers (phenanthrene, chrysene, triphenylene, naphthalene and coronene). Figure 4.9 shows the confusion charts for the results obtained by classifying the PAHs using the $\bar{H}_{-1,1}$ indices for the K, L, M and N regions. These results coincide with those published long time ago by Pullman and Pullman [127] which shown that the K and L regions are enough to classify correctly PAHs into carcinogenic/inactive classes. As can be observed in Figure 4.10 (a, c, and e) carcinogenic compounds are those having large values of $\bar{H}_{-1,1}(K)$ as well as of $\bar{H}_{-1,1}(L)$ (red regions in the mentioned plots). This indicates that carcinogenic PAHs have large contributions of the HOMO/LUMO eigenvalues are those close to them, which parallel the idea of compounds of high reactivity. In fact, the three only PAHs having the strongest carcinogenicity, i.e., "++++", are the ones having the largest values of $\bar{H}_{-1,1}(K)$: dibenzo[a,i]pyrene (0.5474); dibenzo[a,h]pyrene (0.5410); benzo[a]pyrene (0.5299). However, in general having low values of either $\bar{H}_{-1,1}(K)$ or $\bar{H}_{-1,1}(L)$ result in inactive compounds althout the other index display large values. This result possibly indicates that a combined intervention of both regions, K and

N0.	compound	$H_{-1,1}$				CA	Class
		K	L	M	N		
1	dibenzo[a,i]pyrene	0.5474	0.3511	0.5243	0	++++	A
2	dibenzo[a,h]pyrene	0.541	0.3582	0.5229	0	++++	A
3	benzo[a]pyrene	0.5299	0.3531	0.5266	0.4902	++++	A
4	tribenzo[a,e,i]pyrene	0.5244	0.3806	0.5233	0	++	A
5	dibenzo[a,e]pyrene	0.5024	0.379	0.5222	0.4903	+++	A
6	naphtho[2,3,a]pyrene	0.4899	0.393	0.5088	0.5066	++	A
7	benzo[g,h,i]perylene	0.4799	0	0	0.488	++	A
8	dibenzo[a,h]anthracene	0.4234	0.4388	0.5088	0	++	A
9	dibenzo[a,j]anthracene	0.4192	0.4363	0.5087	0	++	A
10	dibenzo[a,c]anthracene	0	0.4229	0.5192	0	++	A
11	peropyrene	0.5372	0	0	0.4832	+	A
12	benzo[e]naphtho[3,4,a] pyrene	0.5007	0.4011	0.5138	0.4861	+	A
13	benzo[a]naphtho[2,1,8,h,i,j naphthacene	0.4914	0.417	0.5081	0.485	+	A
14	benzo[a]anthracene	0.4255	0.3947	0.5106	0	+	A
15	tribenzo[a,c,h]naphthacene	0.4225	0.4316	0.5207	0	+	A
16	dibenzo[a,c]naphthacene	0	0.4157	0.5214	0	+	A
17	pyrene	0.5183	0	0	0.4783	-	I
18	coronene	0.4934	0	0	0	-	I
19	anthanthrene	0.4833	0.4123	0	0.495	-	I
20	benzo[e]pyrene	0.48	0	0.5161	0.48	-	I
21	chrysene	0.4301	0	0.5091	0	-	I
22	phenanthrene	0.3975	0	0.5072	0	-	I
23	triphenylene	0	0	0.5258	0	-	I
24	dibenzo[e,l]pyrene	0	0	0.5224	0.4889	-	I
25	tetracene	0	0.3881	0.515	0	-	I
26	anthracene	0	0.3525	0.5114	0	-	I
27	naphthalene	0	0	0.4723	0	-	I
28	perylene	0	0	0	0.5008	-	I

Table 4.1: Names of the PAHs studied here, their carcinogenic action (CA), the values of $\bar{H}_{-1,1}$ for the four atomic regions of PAHs.

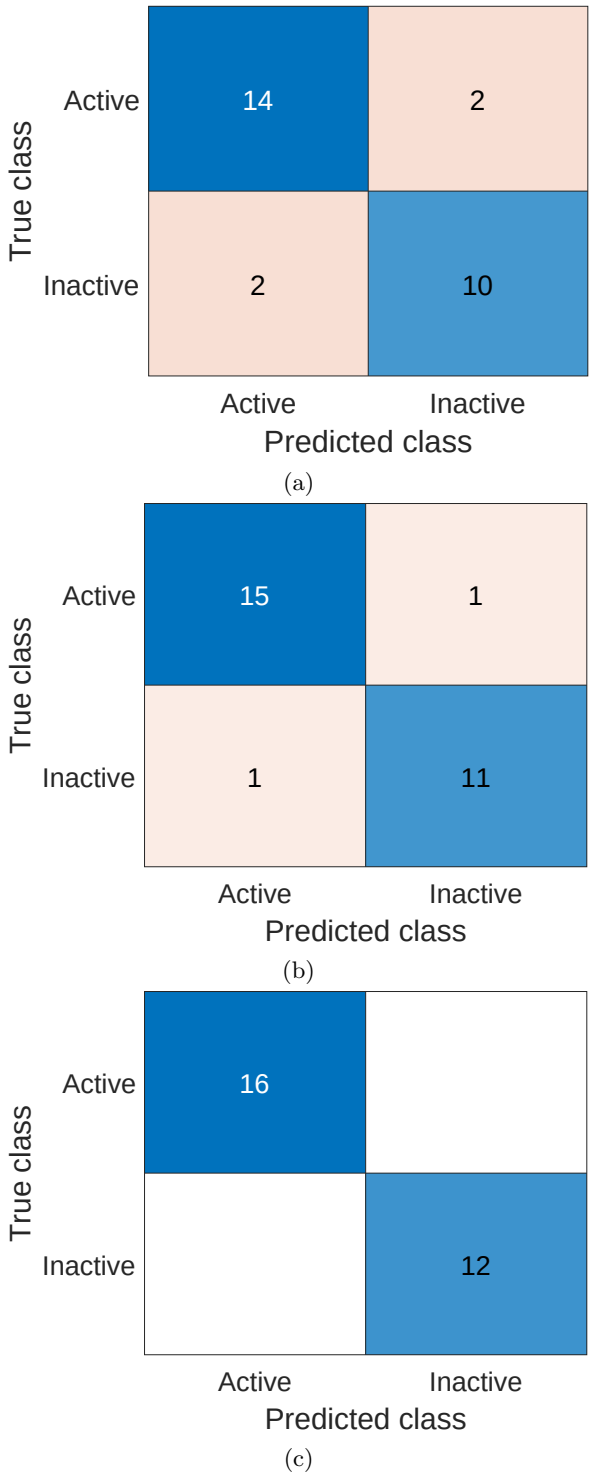


Figure 4.9: Confusion charts for the results obtained with the three classification methods: LDA (a), classification tree (b) and KNN (c) using the $\bar{H}_{-1,1}$ indices for the K, L, M and N reigons

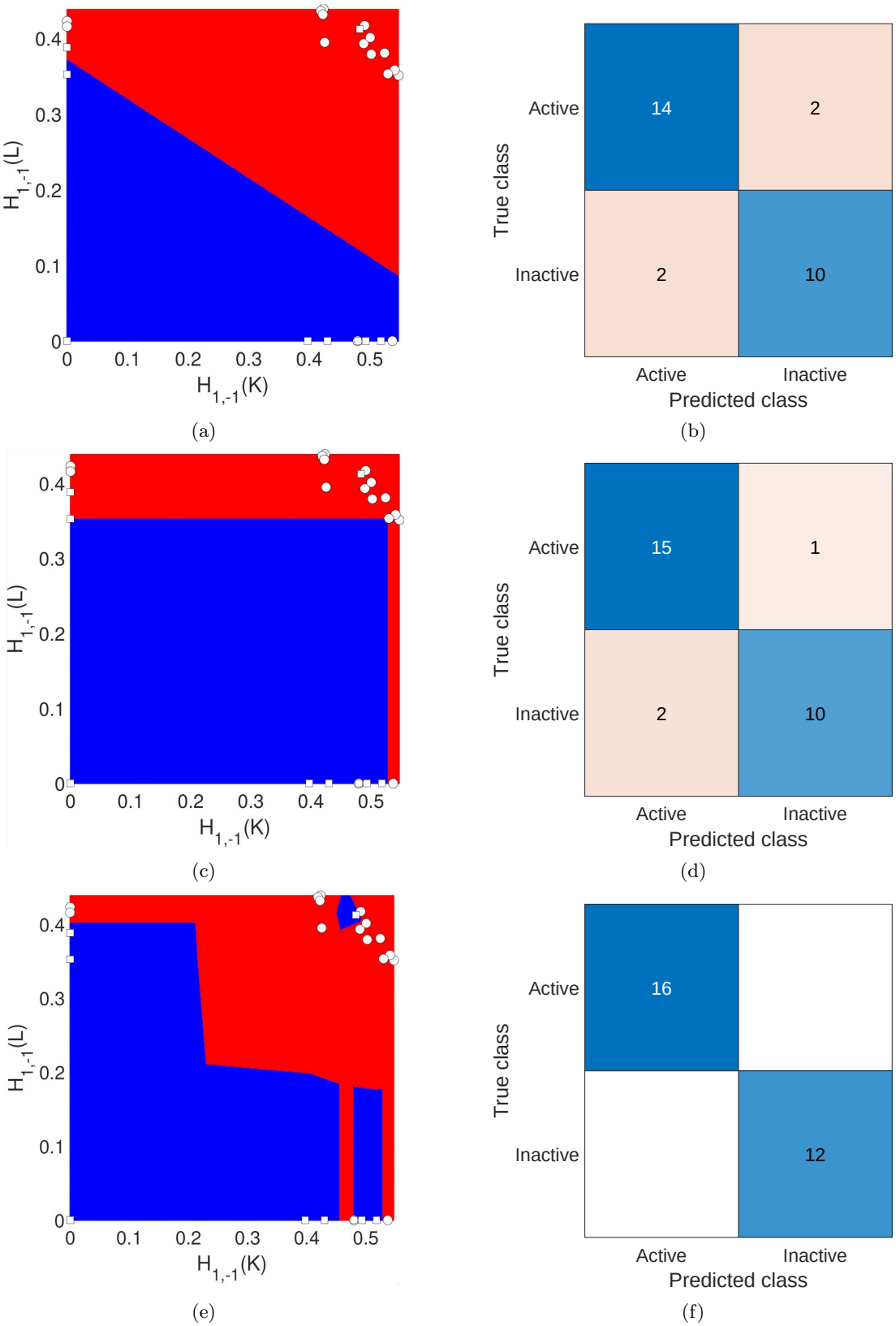


Figure 4.10: Illustration of the classification plots for carcinogenic PAHs (white circles) and inactive ones (white squares) using LDA (a), classification tree (c) and KNN (e). Confusion charts for the results obtained with the three classification methods used: LDA (b), classification tree (d) and KNN (f).

L, are important for the diverse processes giving rise to the carcinogenicity of these compounds.

Chapter 5

Conclusion

The study of spectral properties of adjacency matrices represents an entire area of research in algebraic graph theory. Most of the works using matrix functions for studying graphs are concentrated on the use of the exponential and the resolvent of the adjacency matrix of the graph. Other functions such as the hyperbolic sine and cosine, and ψ-matrix functions have also been reported. All these matrix functions give more weight to the largest eigenvalue and the corresponding eigenvector of the adjacency matrix than to the rest of eigenvalues/eigenvectors. In many real-world networks, where the spectral gap is relatively large, this situation gives rise to discarding important structural information contained in the other parts of the graph spectra. Thus, in this Thesis we have introduced a new adjacency matrix communicability function, the Gaussian communicability function which gives more weight to a selected reference eigenvalues λ_{ref}, which may be located in any region of the graph spectra which opens the door to study the unexplored parts of the spectra of graphs and their hidden structural properties.

We have studied the Gaussian matrix function $f(A) = \exp\left[-\left(\lambda_{\text{ref}}I - A\right)^2\right]$ which accounts for the information contained in the eigenvalues/eigenvectors close to λ_{ref} in the graph spectra. In particular, we have investigated the properties of the Gaussian Estrada index $H = tr\left(f(A)\right)$ for two reference eigenvalues $\lambda_{\text{ref}} = 0$ and $\lambda_{\text{ref}} = -1$ separately.

In the case when $\lambda_{\text{ref}} = 0$, we found analytic expressions for the Gaussian Estrada index

of some elementary graphs as well as for two known models of random networks, the Erdős-Rényi model and the Barabási-Albert model. We also proved that among the connected graphs with n nodes, the maximum of the H index is always obtained for the star graph followed by other complete bipartite graphs. We have shown that the Gaussian Estrada index is related to the existence of important structural patterns in graphs such as the existence of relatively large complete bipartite subgraphs (bicliques). Such bicliques appear naturally in many real-world networks as well as in the Barabási-Albert graphs and other networks with fat-tailed degree distributions. Also, we have considered λ_{ref}=-1 as a representative of the negative eigenvalues of the adjacency matrix. We found an analytic expressions for the index of some elementary graphs as well as for the Erdős-Rényi model and the Barabási-Albert model of random networks. We also proved that among the connected graphs with n nodes, the maximum of the H_{-1} index is always obtained for the complete graph.

Finally, we have generalized the Gaussian matrix function of the adjacency matrix of a graph to account for the spectral folding at two eigenvalues. We have concentrated on the pair $-1, 1$ motivated by the role played by these two eigenvalues in the spectra of molecular graphs. We found analytic expressions of a group of known graphs. useful upper bounds for the index in the two studied models of random graphs were obtained. also, we have shown that the indices derived from double Gaussianization of the graph spectra describe very well the carcinogenicity of PAHs.

Overall, we can say that the Gaussian matrix function allows us to explore the structural information encoded in certain unexplored parts of the graph spectrum that was not studied before which opens new research interests in the study of matrix functions for the structural characterization of graphs. One way to proceed this work is to consider the double Gaussian function resulted from the spectral folding at the pair 0,-1 which is defined by

$$f(A) = \exp\left[-A^2\left(I + A\right)^2\right]. \tag{5.1}$$

It is interesting to explore the structural properties that can arise in networks due to the multiplicity of this important pair of reference eigenvalues and the eigenvalues near it. A set of analytic expressions can be obtained for some elementary graphs and random

models of networks studied here together with proving the following conjectures:

Conjecture 5.0.1. The complete graph K_n has the maximum value of $H_{0,-1}$ among all connected graphs with the same number of vertices.

Conjecture 5.0.2. Let T_n be a tree of n nodes, then

$$H_{0,-1}(P_n) \leq H_{0,-1}(T_n) \leq H_{0,-1}(K_{1,n-1}). \tag{5.2}$$

Another area of research can be pursued is to consider the distance and the angle [53, 55] of the Gaussian communicability function. As proved before [54] that for any positive semidefinite matrix $f_{\alpha_k}(A)$, the quantity

$$\eta_{p,q} = \left[\left(f_{\alpha_k}(A)\right)_{pp} + \left(f_{\alpha_k}(A)\right)_{qq} - 2\left(f_{\alpha_k}(A)\right)_{pq} \right]^{\frac{1}{2}}, \tag{5.3}$$

defines a Euclidean distance between the nodes p and q of the graph. Also, we can define a communicability angle based on the Gaussian matrix function. For instance, the Gaussian communicability distance and angle for the Gaussian matrix function $f(A) = \exp\left(-A^2\right)$ are defined as follows:

$$\xi_{pq} = \sqrt{\left(e^{-A^2}\right)_{pp} + \left(e^{-A^2}\right)_{qq} - 2\left(e^{-A^2}\right)_{pq}}. \tag{5.4}$$

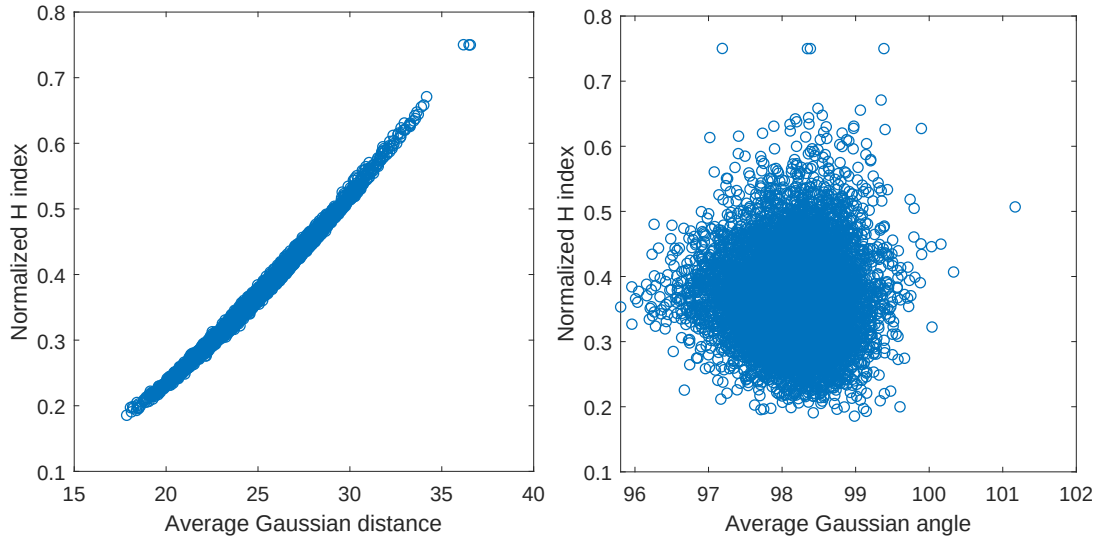
$$\cos\theta_{pq} = \frac{\left(e^{-A^2}\right)_{pq}}{\sqrt{\left(e^{-A^2}\right)_{pp}\left(e^{-A^2}\right)_{qq}}}. \tag{5.5}$$

The average Gaussian communicability distance and angle for graphs are defined by

$$\Upsilon = \sum_{p,q} \xi_{pq}. \tag{5.6}$$

$$\langle\theta\rangle = \frac{2}{n(n-1)} \sum_{p<q} \theta_{pq}. \tag{5.7}$$

Figure 5.1 shows a scatter plot between the average Gaussian communicability distance (the average Gaussian communicability angle) and the normalized H index for all graphs with 8 nodes. The scatter plot even if it is not mathematically sufficient to



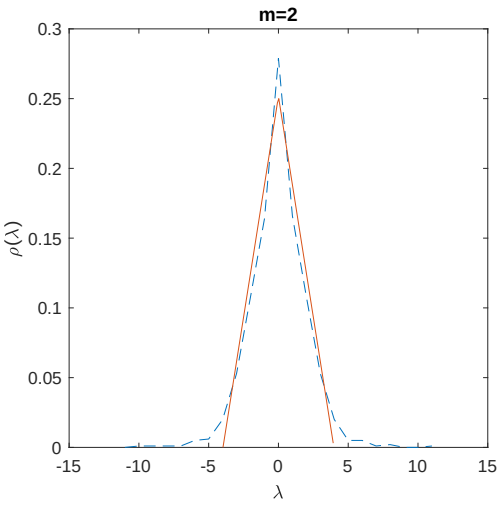
(a) (b)

Figure 5.1: Scatter plot of the normalized H index versus the average Gaussian communicability distance (a) and the average Gaussian communicability angle (b).

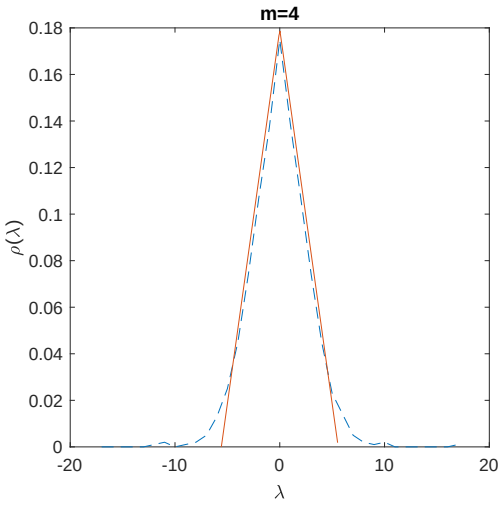
demonstrate the significance of these indices, but it suggests the existence of hidden potentials worth studying in detail. Also, a comparison of the average Gaussian distance and angle between the Gaussian communicability matrices defined in this Thesis may open the door to discover more about the structure of graphs and networks.

Appendix A

In the following, the eigenvalue distribution of some graphs generated using the Barabási-Albert model with $n = 1000$ and $m_0 = 4, 5, 10, 15$, where the dashed red line presents the eigenvalue distribution of the sample generated graph and the solid blue line presents the triangular distribution in equation 1.3. As can be seen, the BA networks display triangular eigenvalue distributions for relatively small edge densities and deformations of it occurs for larger densities.

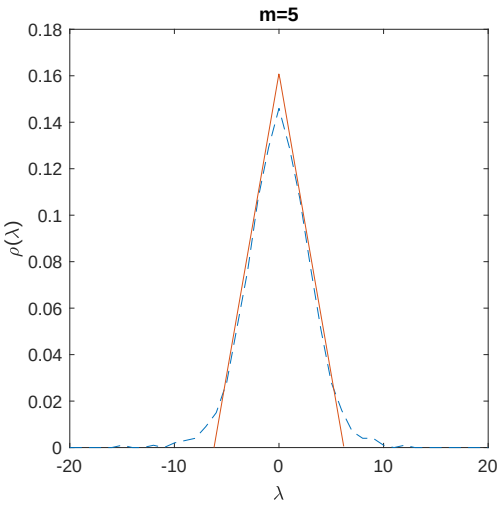


(a)

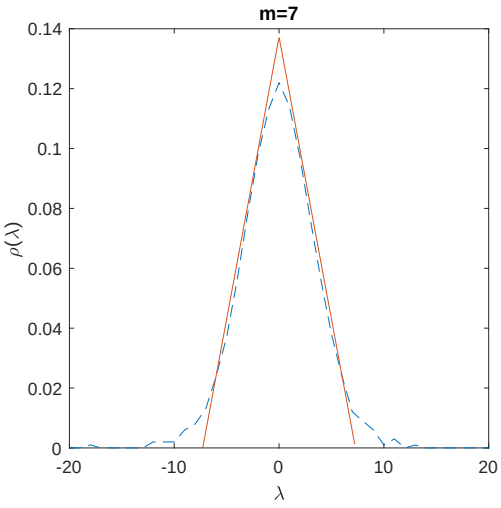


(b)

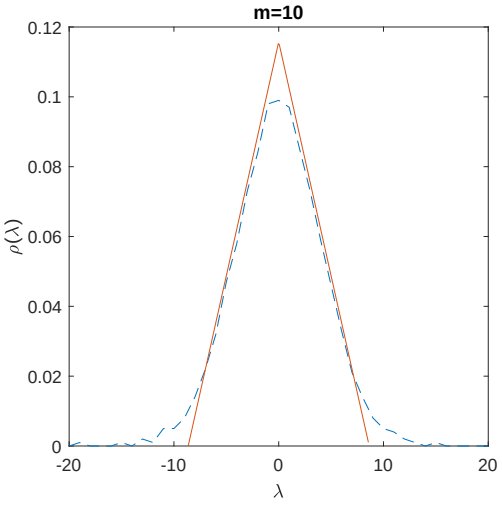
Appendix A.



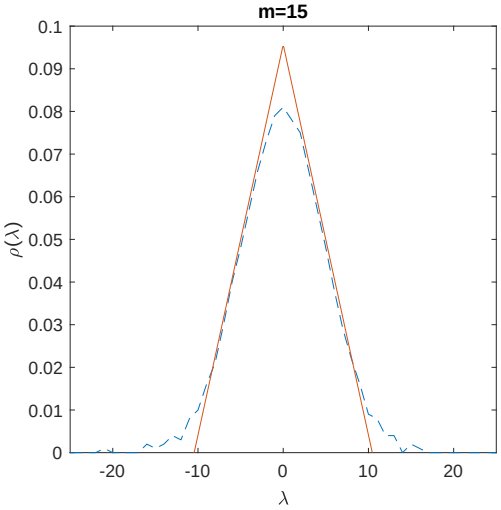
(a)



(b)



(c)



(d)

Appendix B

Datasets of Real-World Networks Studied in Chapter Two

Brain networks

- Neurons: Neuronal synaptic network of the nematode *C. elegans*. Included all data except muscle cells and using all synaptic connections [63]; Cat and macaque visual cortices: the brain networks of macaque visual cortex and cat cortex, after the modifications introduced by Sporn and Kötter [64].

Ecological networks

- Benguela: Marine ecosystem of Benguela off the southwest coast of South Africa [65]; Bridge Brook: Pelagic species from the largest of a set of 50 New York Adirondack lake food webs [66]; Canton Creek: Primarily invertebrates and algae in a tributary, surrounded by pasture, of the Taieri River in the South Island of New Zealand [67]; Chesapeake Bay: The pelagic portion of an eastern U.S. estuary, with an emphasis on larger fishes [68]; Coachella: Wide range of highly aggregated taxa from the Coachella Valley desert in southern California [69]; El Verde: Insects, spiders, birds, reptiles and amphibians in a rainforest in Puerto Rico [70]; Grassland: all vascular plants and all insects and trophic interactions found inside stems of plants collected from 24 sites distributed within England and Wales [71]; Little Rock: Pelagic and benthic species, particularly fishes,

zooplankton, macroinvertebrates, and algae of the Little Rock Lake, Wisconsin, U.S. [72]; Reef Small: Caribbean coral reef ecosystem from the Puerto Rico-Virgin Island shelf complex [73]; Scotch Broom: Trophic interactions between the herbivores, parasitoids, predators and pathogens associated with broom, Cytisus scoparius, collected in Silwood Park, Berkshire, England, UK [74]; Shelf: Marine ecosystem on the northeast US shelf [75]; Skipwith: Invertebrates in an English pond [76]; St. Marks: Mostly macroinvertebrates, fishes, and birds associated with an estuarine seagrass community, Halodule wrightii, at St. Marks Refuge in Florida [77]; St. Martin: Birds and predators and arthropod prey of Anolis lizards on the island of St. Martin, which is located in the northern Lesser Antilles [78]; Stony Stream: Primarily invertebrates and algae in a tributary, surrounded by pasture, of the Taieri River in the South Island of New Zealand in native tussock habitat [79]; Ythan_1: Mostly birds, fishes, invertebrates, and metazoan parasites in a Scottish Estuary [80] ;Ythan_2: Reduced version of Ythan1 with no parasites [81].

- Termite: The networks of three-dimensional galleries in termite nests [82]; Ant: The network of galleries created by ants [83]; Dolphins: social network of frequent association between 62 bottlenose dolphins living in the waters off New Zealand [84];

Informational networks

- Centrality: Citation network of papers published in the field of Network Centrality [85, 86]; GD: Citation network of papers published in the Proceedings of Graph Drawing during the period 1994-2000 [87]; ODLIS: Vocabulary network of words related by their definitions in the Online Dictionary of Library and Information Science. Two words are connected if one is used in the definition of the other [88]; Roget: Vocabulary network of words related by their definitions in Roget's Thesaurus of English. Two words are connected if one is used in the definition of the other [89]; Small World: Citation network of papers that cite S. Milgram's 1967 Psychology Today paper or use Small World in title [90].

Appendix B. Datasets of Real-World Networks Studied in Chapter Two

Biological networks

- Protein-protein interaction networks in: *Kaposi sarcoma herpes virus* (KSHV) [91]; *P. falciparum* (malaria parasite) [92]; *S. cerevisiae* (yeast) [93,94]; *A. fulgidus* [95]; *H. pylori* [96]; *E. coli* [97] and *B. subtilis* [98].

- Trans_E.coli: Direct transcriptional regulation between operons in *Escherichia coli* [99, 100]; Trans_sea_urchin: Developmental transcription network for sea urchin endomesoderm development. [99]; Trans_yeast: Direct transcriptional regulation between genes in *Saccaromyces cerevisae*. [63, 99].

Social and economic networks

- Corporate: American corporate elite formed by the directors of the 625 largest corporations that reported the compositions of their boards selected from the Fortune 1000 in 1999 [101]; Geom: Collaboration network of scientists in the field of Computational Geometry [90]; Prison: Social network of inmates in prison who chose "What fellows on the tier are you closest friends with?" [102]; Drugs: Social network of injecting drug users (IDUs) that have shared a needle in the last six months [103]; Zachary: Social network of friendship between members of the Zachary karate club [104]; College: Social network among college students in a course about leadership. The students choose which three members they wanted to have in a committee [105]; ColoSpring: The risk network of persons with HIV infection during its early epidemic phase in Colorado Spring, USA, using analysis of community wide HIV/AIDS contact tracing records (sexual and injecting drugs partners) from 1985-1999 [106]; Galesburg: Friendship ties among 31 physicians [86]; High_Tech: Friendship ties among the employees in a small high-tech computer firm which sells, installs, and maintain computer systems [86, 107]; Saw Mills: Social communication network within a sawmill, where employees were asked to indicate the frequency with which they discussed work matters with each of theircolleagues [86, 108];

Technological and infrastructural networks

- Electronic: Three electronic sequential logic circuits parsed from the ISCAS89 benchmark set, where nodes represent logic gates and flip-flop [63]; USAir97: Airport transportation network between airports in US in 1997 [90]; Internet: The internet at the Autonomous System (AS) level as of September 1997 and of April 1998 [109]; Power Grid: The power grid network of the Western USA [110].

Software networks

- Collaboration networks associated with six different open-source software systems, which include collaboration graphs for three Object Oriented systems written in C++, and call graphs for three procedural systems written in C. The class collaboration graphs are from version 4.0 of the VTK visualization library; the CVS snapshot dated 4/3/2002 of Digital Material (DM), a library for atomistic simulation of materials; and version 1.0.2 of the AbiWord word processing program. The call graphs are from version 3.23.32 of the MySQL relational database system, and version 1.2.7 of the XMMS multimedia system. Details of the construction and/or origin of these networks are provided in Myers [111].

Bibliography

[1] L. Collatz, U. Sinogowitz, Spektren endlicher Grafen. In Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg. 21 (1957) 63-77.

[2] Hückel, E. Zur Quantentheorie der Doppelbindung. Zeitschrift für Physik. 60 (1930) 423-456.

[3] Hückel E. Quantentheoretische Beiträge zum Benzolproblem. Zeitschrift für Physik. 70 (1931) 204-86.

[4] K. Yates, Hückel Molecular Orbital Theory. Elsevier, 2012.

[5] B.J. Powell, An introduction to effective low-energy Hamiltonians in condensed matter physics and chemistry. arXiv preprint arXiv:0906.1640. 2009 Jun 9.

[6] D. Cvetković, Applications of Graph Spectra: An Introduction to the Literature, in: D. Cvetković, I. Gutman (Eds.), Applications of Graph Spectra, Math. Inst, Belgrade, 2009, pp. 123–140.

[7] N. Biggs, Algebraic Graph Theory, 2nd. ed., Cambridge University Press, 1993.

[8] P. Bonacich, Factoring and weighting approaches to status scores and clique identification. J. Math. Sociol. 2, 113–120.

[9] P. Bonacich, Power and centrality: A family of measures. Am. J. Soc. 92, 1170–82.

[10] S. Brin, L. Page, The anatomy of a large-scale hypertextual Web search engine. Comput. Net. ISDN Syst. (1998) 33, 107–117.

[11] G. Chen, X. Wang, X. Li, Fundamentals of Complex Networks: Models, Structures and Dynamics. John Wiley & Sons; 2014 .

[12] X. Li, I. Gutman, Y. Shi, Graph Energy (1. Aufl. ed.). New York, NY: Springer-Verlag (2012).

[13] M. Benzi, P. Boito, Matrix functions in network analysis, GAMM Mitteilungen (2020).

[14] E. Estrada, P. Knight, A First Course in Network Theory. Oxford University Press, 2015.

[15] N. J. Higham, Functions of Matrices: Theory and Computation. (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2008).

[16] L. Katz, A new index derived from sociometric data analysis, Psychometrika 18 (1953) 39-43.

[17] M. Benzi, C. Klymko, On the limiting behavior of parameter-dependent network centrality measures, SIAM J. Matrix Anal. Appl. 36 (2015) 686-706.

[18] M. Benzi, E. Estrada, C. Klymko, Ranking hubs and authorities using matrix functions, Linear Algebra Appl. 438 (2013) 2447-2474.

[19] F. Arrigo, M. Benzi, Updating and downdating techniques for optimizing network communicability, SIAM J. Sci. Comp. 38 (2016) B25-B49.

[20] V. Ejov, J.A. Filar, S.K. Lucas, P. Zograf, Clustering of spectra and fractals of regular graphs, J. Math. Anal. Appl. 333 (2007) 236-246.

[21] V. Ejov, S. Friedlan, G.T. Nguyen, A note on the graph's resolvent and the multifilar structure, Linear Algebra Appl. 431 (2009) 1367-1379.

[22] H. Deng, S. Radenković, I. Gutman, The Estrada index, in: D. Cvetković, I. Gutman (Eds.), Applications of Graph Spectra, Math. Inst, Belgrade, 2009, pp. 123-140.

[23] J.A. de la Peña, I. Gutman, J. Rada, Estimating the Estrada index, Linear Algebra Appl. 427 (2007) 70–76.

[24] I. Gutman, H. Deng, and S. Radenković, The Estrada index: an updated survey, in: D. Cvetković, I. Gutman (Eds.), Selected Topics on Applications of Graph Spectra, Math. Inst., Beograd, 2011, pp. 155–174.

[25] E. Estrada, Characterization of the folding degree of proteins, Bioinformatics 18 (2002) 697–704.

[26] E. Estrada, J.A. Rodríguez-Velázquez, Subgraph centrality in complex networks, Phys. Rev. E 71 (2005) 056103.

[27] E. Estrada, N. Hatano, Communicability in complex networks, Phys. Rev. E 77 (2008) 036111.

[28] E. Estrada, Generalized walks-based centrality measures for complex biological networks, J. Theor. Biol. 263 (2010) 556–565.

[29] E. Estrada, J. Gómez-Gardeñes, Network bipartivity and the transportation efficiency of European passenger airlines, Physica D 323-324 (2016) 57–63.

[30] E. Estrada, Characterization of 3D molecular structure, Chemical Physics Letters, Vol 319, Issues 5–6, 2000, 713-718.

[31] E. Estrada, G. Silver, Accounting for the role of long walks on networks via a new matrix function. J. Math. Anal. Appl. 449 (2017) 1581-600.

[32] E. Estrada, M. Benzi, What is the meaning of the graph energy after all? Discr. Appl. Math. 230 (2017) 71-77.

[33] E. Estrada, The electron density function of the Hückel (tight-binding) model. Proc. R. Soc. A. 474 (2018) 20170721.

[34] E. Estrada, N. Hatano, M. Benzi, The physics of communicability in complex networks, Phys. Rep. 514 (2012) 89–119.

[35] I.J Farkas, I. Derényi, A.-L. Barabási, T Vicsek, "Spectra of real-world graphs: Beyond the semicircle law." Phys. Rev. E 64, 026704 (2001).

[36] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," Science 286, 509–512 (1999).

[37] E. Estrada, Spectral scaling and good expansion properties in complex networks, Europhys. Lett. 73 (2006) 649.

[38] E. Estrada, Topological structural classes of complex networks, Phys. Rev. E 75 (2007) 016103.

[39] A. E. Brouwer, W. H. Haemers, Spectra of Graphs / [internet Resource]. 1st Ed. 2012.. ed. Universitext. 2012.

[40] R. Balakrishnan, K. Ranganathan. Textbook of Graph Theory / [internet Resource]. 2nd ed. Heidelberg: Springer, 2012.

[41] J. E. Hirsch, and J. R. Schrieffer,"Dynamic correlation functions in quantum systems: A Monte Carlo algorithm," Phys. Rev. B 28, 5353–5356 (1983).

[42] J. E. Barrios-Vargas and G. G. Naumis, "Doped graphene: the interplay between localization and frustration due to the underlying triangular symmetry," J. Phys.: Condens. Matter 23, 375501 (2011).

[43] E. Estrada, and M. Benzi, "Atomic displacements due to spin-spin repulsion in conjugated alternant hydrocarbons," Chem. Phys. Lett. 568-569, 184–189 (2013).

[44] E. Estrada, J.A. Rodríguez-Velázquez, Spectral measures of bipartivity in complex networks. Phys. Rev. E 72 (2005) 046105.

[45] E. Estrada, and D. J. Higham, "Network properties revealed through matrix functions," SIAM Rev. 52, 696–714 (2010).

[46] J. J. Crofts, D. J. Higham, A weighted communicability measure applied to complex brain networks, J. Royal Soc. Interface, 6 (2009), pp. 411–414.

[47] J. Serra-Musach, F. Mateo, E. Capdevila-Busquets, et al. Cancer network activity associated with therapeutic response and synergism. Genome Med 8, 88 (2016).

[48] I. M. Campbell, R. A. James, E. S. Chen, C. A. Shaw, NetComm: a network analysis tool based on communicability. Bioinformatics. 2014;30(23):3387-3389.

[49] S. Rönnqvist, P. Sarlin, From text to bank interrelation maps, 2014 IEEE Conference on Computational Intelligence for Financial Engineering.

[50] D. M. Walker, A. Tordesillas, Topological evolution in dense granular materials: A complex networks perspective, International Journal of Solids and Structures, Vol 47, Issue 5, 2010, 624-639.

[51] J. A. Khan, Y. Ghamri-Doudane, SAVING: socially aware vehicular information-centric networking, IEEE Communications Magazine 54 (8), 100-107.

[52] E. Estrada and N. Hatano, Communicability graph and community structures in complex networks, Appl. Math. Comput., 214 (2009), 500–511.

[53] E. Estrada, The communicability distance in graphs, Linear Algebra and its Applications, Vol 436, Issue 11, 2012, 4317-4328.

[54] E. Estrada, M. G. Sanchez-Lirola, and J. A. de la Pẽna, Hyperspherical Embedding of Graphsand Networks in Communicability Spaces, Discr. Appl. Math., 176 (2014). 53–77.

[55] E. Estrada, N. Hatano, Communicability angle and the spatial efficiency of networks. SIAM Review, 58(4) 2016, 692-715.

[56] E. Estrada, E. Vargas-Estrada, and H. Ando, Communicability angles reveal the critical links in network consensus dynamics, Phys. Rev. E, 92 (2015), 052809.

[57] L. W. Wang and A. Zunger, "Solving Schroedinger's equation around a desired energy: Application to silicon quantum dots," J. Chem. Phys. 100, 2394 (1994).

[58] A. Canning, L. W. Wang, A. Williamson, and A. Zunger, "Parallel empirical pseudopotential electronic structure calculations for million atom systems," J. Comput. Phys. 160, 29–41 (2000).

[59] K.B. Oldham, J.C. Myland, J. Spanier, An atlas of functions : with Equator, the atlas function calculator, Second ed 2009.,

[60] M. Benzi, and G. H. Golub, "Bounds for the entries of matrix functions with application to preconditioning," BIT 39, 417–438 (1999).

[61] M. Benzi, and P. Boito, "Quadrature rule-based bounds for functions of adjacency matrices," Lin. Algebra Appl. 433, 637–652 (2010).

[62] M. Krivelevich, and B. Sudakov, "The largest eigenvalue of sparse random graphs," Combin. Probab. Comput. 12, 61–72 (2003).

[63] R. Milo, S. Shen–Orr, S. Itzkovitz, et al., "Network motifs: simple building blocks of complex networks," Science 298, 824–827 (2002).

[64] O. Sporns, and R. Kötter. "Motifs in brain networks," PLoS Biology 2, e369 (2004).

[65] P. Yodzis, "Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem," J. Anim. Ecol. 67, 635-658 (1998).

[66] G. A. Polis, "Complex trophic interactions in deserts: an empirical critique of food-web theory," Am. Nat. 138, 123-155 (1991).

[67] C. Townsend, R. M. Thompson, and A. R. McIntosh, et al. "Disturbance, resource supply, and food-web architecture in streams," Ecol. Lett. 1, 200 (1998).

[68] R. R. Christian, and J. J. Luczkovich, "Organizing and understanding a winter's seagrass foodweb network through effective trophic levels," Ecol. Model. 117, 99-124 (1999).

[69] P. H. Warren,"Spatial and temporal variation in the structure of a fresh-water food web," Oikos 55, 299-311 (1989).

[70] R. B. Waide, and W. B. Reagan, (Eds.) *The Food Web of a Tropical Rainforest.* (University Chicago Press, Chicago, 1996).

[71] N. D. Martinez, B. A. Hawkins, H. A. Dawah, and B. P. Feifarek, "Effects of sampling efforts on characterization of food web structure," Ecology 80, 1044–1055 (1999).

[72] K. Havens, "Scale and structure in natural food webs," Science 257, 1107-1109 (1992).

[73] S. Opitz, "Trophic Interactions in Caribbean coral reefs," ICLARM Tech. Rep. 43, Manila, Philippines, (1996).

[74] J. Memmott, N. D. Martinez, and J. E. Cohen, "Predators, parasites and pathogens: species richness, trophic generality, and body sizes in a natural food web," J. Animal Ecol. 69,1-15 (2000).

[75] J. Link, "Does food web theory work for marine ecosystems?," Mar. Ecol. Prog. Ser. 230, 1-9 (2002).

[76] P. Yodzis, "Diffuse effects in food webs," Ecology 81, 261-266 (2000).

[77] L. Goldwasser, and J. A. Roughgarden, "Construction and analysis of a large Caribbean food web," Ecology 74, 1216-1233 (1993).

[78] N. D. Martinez, "Artifacts or attributes? Effects of resolution on the Little Rock Lake food web," Ecol. Monogr. 61, 367-392 (1991).

[79] D. Baird, and R. E. Ulanowicz, "The seasonal dynamics of the. Chesapeake Bay ecosystem," Ecol. Mon. 59, 329-364 (1989).

[80] M. Huxman, S. Beany, and D. Raffaelli, "Do parasites reduce the chances of triangulation in a real food web?," Oikos 76, 284-300 (1996).

[81] S. J. Hall, and D. Rafaelli, "Food-web patterns- lessons from a species-rich web," J. Anim. Ecol. 60, 823–842 (1991).

[82] A. Perna, S. Valverde, and J. Gautrais, et al. “Topological efficiency in the three-dimensional gallery networks of termite nests,” Physica A 387, 6235-6244 (2008).

[83] J. Buhl, J. Gautrais, and R. V. Solé, et al. “Efficiency and robustness in ant networks of galleries,” Eur. Phys. J. B 42, 123-129 (2004).

[84] D., Lusseau, “The emergent properties of a dolphin social network,” Proc. R. Soc. Lond. B (Suppl.) 270, 186-188 (2003).

[85] N. P. Hummon, P. Doreian, and L. C. Freeman, “Analyzing the structure of the centrality-productivity literature created between 1948 and 1979,” Know.-Creat. Diffus. Util. 11, 459-480 (1990).

[86] W. de Nooy, A. Mrvar, and V. Batagelj, Exploratory Social Network Analysis with Pajek, Cambridge University Press, Cambridge (2005).

[87] V. Batagelj, and A. Mrvar, “Graph Drawing Contest 2001,” http://vlado.fmf.uni-lj.si/pub/GD/GD01.htm. (2001).

[88] ”ODLIS: Online Dictionary of Library and Information Science,” http://vax.wcsu.edu/library/odlis.html. (2002).

[89] Roget’s Thesaurus of English Words and Phrases, Project Gutenberg. http://gutenberg.net/etext/22 (2002).

[90] V. Batagelj, and A. Mrvar, Pajek datasets. Available at: http://vlado.fmf.uni-lj.si/pub/networks/data/ (2006).

[91] P. Uetz, Y.-A. Dong, and Ch. Zeretzke, et al., “Herpesviral protein networks and their interaction with the human proteome,” Science 311, 239-242 (2006).

[92] D. LaCount, M. Vignali, and R. Chettier, et al. “A protein interaction network of the malaria parasite Plasmodium falciparum,” Nature 438, 103–107 (2005).

[93] D. Bu, Y. Zhao, and L. Cai, et al. “Topological structure analysis of the protein-protein interaction network in budding yeast,” Nucleic Acids Res. 31, 2443-2450 (2003).

[94] C. von Mering, R. Krause, and B. Snel, et al. "Comparative assessment of large-scale data sets of protein-protein interactions," Nature 417, 399-403 (2002).

[95] M. Motz, I. Kober, and C. Girardot, et al. "Elucidation of an Archaeal Replication Protein Network to Generate Enhanced PCR Enzymes," J. Biol. Chem. 277, 16179–16188 (2002).

[96] C. Y. Lin, C. L. Chen, C. S. Cho, et al. "hp-DPI: Helicobacter pylori database of protein interactomes. A combined experimental and inferring interactions," Bioinformatics 21, 1288–1290 (2005).

[97] G. Bultland, J. M. Peregrín-Alvarez, and J. Li, et al. "Interaction network containing conserved and essential protein complexes in Escherichia coli," Nature 433, 531-537 (2005).

[98] P. Noirot, and N. F. Noirot-Gross, "Protein interaction networks in bacteria," Curr. Op. Microb. 7, 505–512 (2004).

[99] R. Milo, S. Itzkovitz, and N. Kashtan, et al. "Superfamilies of evolved and designed networks," Science 303, 1538-1542 (2004).

[100] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of Escherichia coli," Nature Gen. 31, 64-68 (2002).

[101] G. F. Davis, M. Yoo, and W. E. Baker, "The Small World of the American Corporate Elite, 1982-2001," Strategic Organization 1, 301-326 (2003).

[102] D. MacRae, "Direct factor analysis of sociometric data," Sociometry 23, 360-371 (1960).

[103] J. Moody, Data for this project was provided in part by NIH grants DA12831 and HD41877, those interested in obtaining a copy of these data should contact James Moody (moody.77@sociology.osu.edu), (2001).

[104] W. Zachary, "An information flow model for conflict and fission in small groups," J. Anthropol. Res. 33, 452-473 (1977).

[105] L. D. Zeleny, "Adaptation of research findings in social leadership to college classroom procedures," Sociometry 13, 314-328 (1950).

[106] J. J. Potterat, L. Philips-Plummer, and S. Q. Muth, et al. "Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs," Sex. Transm. Infect. 78, i159-i163 (2002).

[107] D. Krackhardt, "The ties that torture: Simmelian tie analysis in organizations" Res. Sociol. Org. 16, 183-210 (1999).

[108] J. H. Michael, and J. G. Massey, "Modeling the communication network in a sawmill" Forest Prod. J. 47, 25-30 (1997).

[109] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology" Comp. Comm. Rev. 29, 251-262 (1999).

[110] D. J. Watts, and S. H. Strogatz, "Collective dynamics of small-world networks" Nature 393, 440-442 (1998).

[111] C. R. Myers, "Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs," Phys. Rev. E 68, 046116 (2003).

[112] M. E. J. Newman, "Assortative mixing in networks," Phys. Rev. Lett. 89, 208701 (2002).

[113] Y. Morita, S. Suzuki, K. Sato, T. Takui, Synthetic organic spin chemistry for structurally well-defined open-shell graphene fragments. Nature Chem. 3 (2011) 197.

[114] Y.S. Jiang, G.Y Chen, On subspectral problem—benzenoid hydrocarbons with common eigenvalues ±1. Theor. Chim. Acta 76 (1990) 437-450.

[115] P. Ball, Elusive triangulene created by moving atoms one at a time. Nature 542 (2017) 284-285.

[116] M. Cámara, W. H. Haemers, Spectral characterizations of almost complete graphs. Discr. Appl. Math. 176 (2014) 19-23.

[117] P. Blackwell, M. Edmondson-Jones, J. Jordan, Spectra of adjacency matrices of random geometric graphs. University of Sheffield. Department of Probability and Statistics; 2007.

[118] I. Gutman, A. Graovac, Estrada index of cycles and paths, Chem. Phys. Lett. 436 (1–3) (2007) 294–296

[119] I. S. Gradshteyn, I. M. Ryzhik, A. Jeffrey, and D. Zwillinger. Table of Integrals, Series and Products. 7th ed. Oxford: Academic, 2007. Print.

[120] J.H. Smith, Some properties of the spectrum of a graph, in Combinatorial Structures and their Application, R. Guy, H. Hanani, N. Sauer, and J. Schönheim, eds., Gordon and Breach, Science Publ., Inc., New York, London, Paris, 1970, pp. 403–406.

[121] F. Esser, F. Harary, On the spectrum of a complete multipartite graph. Europ. J. Combinat. 1 (1980) 211-21.

[122] I. Fleming, Molecular Orbital and Organic Chemical Reactions: Student Edition, John Wiley and Sons Ltd, 2009.

[123] P. W. Fowler, T. Pisanski, Homo-lumo maps for chemical graphs, MATCH Commun. Math. Comput. Chem 64 (2) (2010) 373–390.

[124] S. M. Cioabă , W. H. Haemers, J. R. Vermette, W. Wong, The graphs with all but two eigenvalues equal to ± 1, Journal of Algebraic Combinatorics41 (3) (2015) 887–897.

[125] B. Moorthy, C.Chu, D. J. Carlin, Polycyclic aromatic hydrocarbons: from metabolism to lung cancer, Toxicological Sciences 145 (1) (2015), 5-15

[126] I. A. Smith, G. D. Berger, P. G. Seybold, M Serve, Relationships between carcinogenicity and theoritical reactivity indices in polycyclic aromatic hydrocarbons, Cncer research 38 (9) (1978) 2968–2977

[127] B. Pullman, Recent developments on the mechanism of chemical carcinogenesis by aromatic hydrocarbons, International Journal of Quantum Chemistry 16 (3) (1979) 669–689.

[128] K. P. Vijayalakshmi, C. H. Suresh, Theoretical studies on the carcinogenicity of polycyclic aromatic hydrocarbons, Journal of computational chemistry 29 (11) (2008) 1808–1817.