The Spectrum of the Graph Laplacian as a Tool for Analyzing Structure and Evolution of Networks

Von der Fakultät für Mathematik und Informatik
der Universität Leipzig
angenommene

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM
(Dr.rer.nat.)

im Fachgebiet

Informatik

vorgelegt

von M.SC. Anirban Banerjee
geboren am 26.08.1978 in Bankura (Indien)

Die Annahme der Dissertation haben empfohlen:

1. Professor Dr. Jürgen Jost (Leipzig)
2. Professor Dr. Josef Leydold (Wien, Österreich)
3. Professor Dr. Ricard V. Solé (Barcelona, Spanien)

Die Verleihung des akademischen Grades erfolgt auf Beschluss des Rates der Fakultät für Mathematik und Informatik vom 21.01.2008 mit dem Gesamtprädikat magna cum laude.

Abstract

In recent years the study of complex systems with network theory, which is based on graph theory, has become quite popular. To analyze a system, different network models have been developed to capture the graph structure, constructed from local interactions of the components of the system, and many parameters have been introduced for analyzing the structure of the graph. These graph invariants can capture certain structural information, but they are not sufficient for capturing all qualitative aspects of a graph. One of the aims of graph theory is to identify the unique and special features of a network from a particular class on one hand and the universal qualities that are shared by other network structures on the other hand. It has been considered as a big challenge to find some sort of signature of networks from different sources and to measure how far one network is from another. It is difficult to distinguish the structural differences between networks from different sources, and thus it seems to be very hard to recognize the source of a real network by looking at its structure, though networks constructed from the same evolutionary process should share common structural properties. It is also difficult to say something about the evolutionary process from the internal connectivity pattern of a network.

The spectrum of the normalized graph Laplacian can reveal structural properties of a network and can be an important tool to help solve the structural identification problem. From the spectrum, we attempt to develop a tool that helps us to understand the network structure on a deep level and to identify the source of the network to a greater extent. The information about different topological properties of a graph carried by the complete spectrum of the normalized graph Laplacian is explored. We investigate how and why structural properties are reflected by the spectrum and how the spectrum changes when compairing different networks from different sources.

In order to understand the information contained in the spectrum of the normalized Laplacian, we systematically investigate the behavior of the eigenvalues.

The spectra of many biological and other networks, such as protein-protein interaction networks show a high multiplicity of the eigenvalue 1. Different internal structures of the graph that produce a high multiplicity of eigenvalue 1 are identified and their effect on the spectrum is proved. The nature of the changing of eigenvalue multiplicity under local and global operations like motif doubling, graph joining and splitting is also investigated. We also expand this to other eigenvalues. Thus we present the spectrum of the normalized graph Laplacian as a systematic tool for the investigation of network structure.

Given a class of empirical networks, reconstruction schemes for elucidating the evolutionary dynamics leading to these given data can be developed. This method is exemplified for protein-protein interaction networks. Traces of duplication and divergence processes in their evolutionary history are identified. In particular, we can identify typical specific features that robustly distinguish protein-protein interaction networks from other classes of networks, in spite of possible statistical fluctuations in the underlying data.

It is a basic question in biology and other fields to identify the characteristic properties that, on the one hand, are shared by structures from a particular realm, like protein-protein interaction, internet, neural or power-grid networks, and that, on the other hand, distinguish them from other structures. We apply our general method, based on the spectrum of the normalized graph Laplacian, to obtain representations, the spectral plots, that allow us to find and visualize such properties systematically. We present such visualizations for a wide range of different networks and compare them with those for networks derived from theoretical schemes. The differences that we find are quite striking and suggest that the search for universal properties of biological and other networks should be complemented by an understanding of the more specific features and organizational principles of the systems at different scales.

We introduce a tentative classification scheme for empirical networks based on qualitative global properties detected through the spectral plot of the Laplacian of the graph underlying the network. Our method identifies several distinct types of networks across different areas of application and indicates the hidden regularity properties of a given class of networks. Our study reflects that spectral distribution is an important characteristic of a network. So we infer that spectral distribution is an excellent diagnostic for categorizing different networks from different sources.

Computations that produce the whole spectrum of a large graph need a lot of

space and time. We introduce a scheme for coarsening a graph to reduce its size. The new graph produced by this method yields a spectral plot similar to that of the original graph. This substitute solution is proposed to reduce the computational complexity and space required for the computation while still giving an idea about the pattern of the spectral plot of a large graph.

Acknowledgements

It is needless to say that without the support, encouragement, and friendly care of a number of people this dissertation would not have been possible. First of all, it behooves me to thank my advisor, Prof. Dr. Jürgen Jost. I am substantially indebted to him for the constant support and supervision he had been providing through out by means of introducing me to the subject, rejuvenating my thoughts with bright insights and new ideas from seemingly nonintersecting branches of mathematics and thus guiding me through to the successful completion of the project we embarked upon.

I wish to thank Dr. Boris N. Khoromskij for his enthusiasm and advice on one of my research results (coarsening of graph). I would also like to thank Prof. Andreas Wagner for many valuable discussions concerning the evolutionary dynamics of protein-protein interaction networks in SFI summer school, Beijing, China 2006. Thanks to Dr. Nihat Ay and Dr. Fatihcan Atay for moral support and lot of interesting discussions.

I am also grateful to all my group members who made my research more enjoyable, and who could always be consulted with whenever in doubt. Sincere thanks to Tobias Elze. I have been benefitted greatly from his help and support with sorting out problems ranging from technical to personal grounds. It is my pleasure to thank Ulrich Steinmetz for the enthusiastic discussions we had from time to time. I am also thankful to Timm Lochmann for helping me with different aspects of LaTeX. I also spent a very nice time with my friends from Nihat's group. Many thanks to all of them. Wishing them good luck with their future research.

I would like to convey my heartfelt gratitude to all the members of MPI MIS, for their help and support that made my research move on more easily. A special thanks to Antje Vandenberg for helping me with professional and personal matters at each and every step of my research.

I am also grateful to my friends, Bhaswati Chatterjee, Chiranjib Mukherjee and Suman Gajula for their kind help during this write up. A very special thanks to Jessica Hill for the encouragement and motivation, I got from her life, which in many ways helped me achieve my cherished goal.

Last, but not the least, thanks to my family for unwavering support and love, and specially my Mom (Madhabi Banerjee) for her understanding and kindness. I am very grateful to Brian Clarke for his patient reading of my thesis and helping me with the English language.

Contents

List of Figures

Chapter 1

Introduction

Network theory is a useful tool for analyzing complex systems. Manifestation of complex behavior is very common in the physical, biological and social sciences. In a complex system, there are large numbers of components which interact with each other. They may act according to some rules that can change over time, and that might not be well understood (Amaral & Ottino, 2004). Since complex systems are self-organized, they are quite adaptive to changes of their environment. The interaction inside the system occurs between close neighbors, but since the system is coupled, each interaction has some effect on the properties and construction of the system. The interactions may follow some dynamics and according to that, the system adopts some inheritance structure. So the dynamics have a great influence on the structure of the system. Conversely the structure may reflect some properties of the dynamics. So there is an interplay between these two properties.

Because of this, it is useful to analyze the structure and its inheritance properties. Network theory, which is an emerging area of science; captures those properties to some extent. Complex systems in various disciplines like bioscience, neuroscience, meteorology, physics, computer science, artificial life, evolutionary computation, economics, earthquake prediction, heart cell synchronization, immune systems, reaction-diffusion systems have been extensively studied with the help of network theory. One can allow the network configuration to change with time, but nevertheless studying the structure at a fixed time reveals many underlying principles of the system. From biological, social, economical, ecological, and technical systems, one can construct networks in terms of the components and interactions between the components of the system. A few examples of the study of networks are in the areas of the world-wide web (WWW) (Albert et al., 1999; Barabási, Albert, & Jeong, 2000; Broder et al., 2000; Huberman & Adamic,

1999; Lawrence & Giles, 1998; Lawrence & Giles, 1999), the internet(Faloutsos et al., 1999; Chen et al., 2002), powergrids (Watts & Strogatz, 1998), worldwide air transportation (Guimera, Mossa, Turtschi, & Amaral, 2005), electronic circuits (Milo et al., 2002), movie actor collaborations (Barabási & Albert, 1999; Newman, Strogatz, & Watts, 2001), coauthorship (scientific collaborations) (Barabási et al., 2002; Castro & Grossman, 1999; Grossman & Ion, 1995; Newman, 2001c; Newman, 2001a; Newman, 2004), citation (Redner, 1998), linguistics (Cancho & Solé, 2001; Dorogovtsev & Mendes, 2001), human sexual contacts (Bearman, Moody, & Stovel, 2004; De, Singh, Wong, Yacoub, & Jolly, 2004; Kretzschmar, 2000; Liljeros, Edling, Amaral, Stanley, & Aberg, 2001), telephone calls (Aiello, Chung, & Lu, 2000; Aiello, Chung, & Lu, 2001), protein-protein interactions(Jeong et al., 2001), protein folding (Greene & Higman, 2003; Scala, Amaral, & Barthelemy, 2001), food webs (Huxham, Beaney, & Raffaelli, 1996; Martinez, 1991; Pimm, Lawton, & Cohen, 1991), nervous systems (Koch & Laurent, 1999; Lago-Fernandez, Huerta, Corbacho, & Siguenza, 2000; Watts & Strogatz, 1998; White et al., 1986), gene regulation (Alon, Surette, Barkai, & Leibler, 1999; Milo et al., 2002; Oltvai & Barabási, 2002), metabolic reactions (Jeong et al., 2000; Ravasz, Somera, Mongru, Oltvai, & Barbasi, 2002; Wagner & Fell, 2001).

For a long time graph theory has played a vital role in analyzing and understanding network structure. There are many approaches to using graph theory to analyze networks. The components of a network can be considered to be vertices and relations or interactions between these components can be considered to be edges of the graph. So whenever we wish to construct a network from a system, first we have to decide and be very clear about what we should consider to be vertices and which kinds of relations among these vertices should be considered to be edge. But it is better to think about whether the network construction from one system would be helpful to study that system or not, otherwise it wouldn't worth building a network from a given system. To analyze the network more intensely, new methodologies, tools and parameters have also been introduced. Furthermore, old methods of graph theory have been re-focused. Thus a new area of science, network science, has been emerging based on graph theory.

1.1 A very short history of network theory

For years, these systems (networks) were considered as sets of components with haphazard connections, mathematically framed in a random paradigm (Caldarelli, Erzan, & Vespignani, 2004). And a conventional way to model these networks was

to make random connections in-between the components (Solomonoff & Rapoport, 1951). Inspired by social structures, Erdős and Rényi proposed very simple model (Erdős & Réanyi, 1959) to construct a random graph[1] which can be represented as a random network where every pair of nodes has probability p of being connected by an edge. Many interesting properties of large random graphs emerge for different values of p (Erdős & Réanyi, 1959; Erdős & Réanyi, 1960). Different possible topologies of the random graph, with a given degree distribution, can be constructed by the process(Molloy & Reed, 1995). Various parameters have been constructed to analyze the inheritance structural properties of network, such as degree distribution, average path length, diameter, betweenness centrality, transitivity or clustering coefficient, etc. (see (Newman, 2003) for details). this suggests introducing different models to capture the structural properties of real networks. Erdős and Rényi's random graph has a typical kind ("Poisson") of degree distribution (Bollobás, 1985). To capture the properties of low average path length and high clustering coefficient of real networks, Watts and Strogatz proposed a model that exhibits a "small-world phenomenon" (Milgram, 1967) by randomizing a fraction Pr of links (edges) connecting nodes in a regular ring lattice (Watts & Strogatz, 1998). Networks generated by this model also have degree distributions similar to Erdős and Rényi's random graphs, but with a more pronounced peak, whereas most real networks have a power-law degree distribution (Albert et al., 1999; Barabási & Albert, 1999; Guimera et al., 2005; Jeong et al., 2001; Jeong et al., 2000; Redner, 1998). Degree distributions that follow a power law are invarient of scale, so they are also called *scale-free*[2] degree distributions. Thus, networks that have a power law degree distribution are sometimes called *scale-free networks* (Barabási & Albert, 1999), although only their degree distributions are scale-free[3]. Probably the first observation of a power law degree distribution in a real network was made by Price in a network of citations of scientific papers in 1965

[1]Probably first random graph model was proposed by Solomonoff and Rapoport in 1951 (Solomonoff & Rapoport, 1951).

[2]A function $f(x)$ which is invariant under a multiplicative factor (rescaling) of the independent variable x is called scale-free. Since the only solution of $f(ax) = bf(x)$ follows a power law, 'power-law' and 'scale-free' have been considered to be the same property. For details on the power law, see (Newman, 2005).

[3]In many papers, the term scale-free has been used for power-law degree distributions, though the function that has the characteristic 'power-law' is only invariant under multiplicative scaling. But in a more sensible way, two different scale-free properties could be considered. One is 'scale-free in space', related to the concept of *self similarity* of the subgraph or *local graph*. And another is 'scale-free in time', which is the property of a model for generating a network with a particular feature at all times. e. g. , a model is scale-free in time if it generates power-law graphs with the same exponent regardless of the choice of time scale. See more in (Chung & Lu, 2006).

(Price, 1965). Price explained the emergence of the property of a power law degree distribution with the help of a model (Price, 1976), which was based on the previous work by Simon on wealth distributions in 1950 (Simon, 1955). After looking at power law degree distributions in the WWW network (Albert et al., 1999; Kumar, Raghavan, Rajagopalan, & Tomkins, 1999), Barabási and Albert proposed a model (Barabási & Albert, 1999) to generate a network that exhibits the property of scale-free degree distribution. In this evolving network model, a new node is attached to the network not in random way, but by establishing connections towards higher-degree nodes. The preference towards getting a new connection with an existing node i with degree n_i depends on the probability $p_i = \frac{n_i}{\sum_j n_j}$. So growing a network by attaching a new node to already existing nodes by edges with this preference (sometimes referred to us *preferential attachment*), there emerges the network with a power law degree distribution. Since then, many models have been introduced to capture the structural properties of real networks.

1.2 Necessity of new method and outline of my work

Though there are many parameters and tool exist in graph theory, but they could not capture enough qualitative feature of networks. There is no good enough method to indicate the hidden regularity properties, to provide evidence for processes like node duplication behind the evolution or construction of a given class of networks, and to identify several distinct types of networks across different domains of applications.

Spectral analysis of the normalized graph Laplacian can reveal inheritance structural properties of a network and can be an important part of the path towards solving the above problem. From the spectrum of the normalized Laplacian matrix, I have attempted to develop a tool that helps understand deep properties of the network structure, so that we can recognize the source of the network. I have explored the information about different topological properties of a graph carried by the complete spectrum of the normalized graph Laplacian. I have investigated how and why structural properties are reflected by the spectrum and how the spectrum changes according to different networks from different sources. For a large network, not only the particular eigenvalues carry information about the structure, but also the density of the eigenvalues at particular points carries much information. Furthermore, I have categorized the different types of networks according to their spectrum. This study shows that spectral distribution is an

important characteristic of a network. So I have inferred that spectral distribution is an excellent diagnostic for categorizing different networks from different sources.

What could be the evolutionary process behind the formation of similar structures? There is an interplay between the dynamics of the network and inheritance structures. So evolutionary processes that are responsible for the construction of the network could be studied from the spectrum of the connectivity matrix. Different graph operations related to the evolution of a network produce specific eigenvalues. Constructions with these operations describe certain processes of graph formation that leave characteristic traces in the spectrum. So a useful plausible hypothesis about evolutionary processes could be made, and it would be easy to find the evolutionary assumption that is of highest relevance for the evolution of that system by investigating the spectrum of a graph constructed from actual data. Based on this idea, I have reconstructed a protein-protein interaction network which is structurally closer to real protein-protein interaction networks than previous models.

1.3 Overview of the dissertation

The presentation of the thesis is as follows.

Chapter 1 contains a brief introduction to the study of complex systems with network theory, which essentially developed from graph theory. The key milestones in the development of network modeling are briefly described. The necessity of spectral analysis for a qualitative structural study of a graph are explained.

Preliminaries needed for the following chapters are given in Chapter 2. The elementary notions of graph theory, network theory and spectral graph theory used in the text are defined. Three generic network models and the properties of the graphs constructed by them are briefly described.

In Chapter 3, constructions of different networks from real systems are demonstrated. Also, the difficulties and complications for qualitative structural analyses are explained. Questions related to common and varying features between the networks from different classes or domains are posed.

In the beginning of Chapter 4, a brief history and application of spectral analysis is given. An outline of previous studies on the spectral density of graphs is

drawn. Then we introduce the normalized graph Laplacian operator and described three important properties of this operator. Well-known results on the eigenvalues of this operator are discussed. The relation between the Cheeger constant and first nontrivial eigenvalue of this operator is described, At the end, we present examples of the spectra of different elementary graphs.

In Chapter 5, we turn to visualization of the spectrum of the normalized graph Laplacian operator. By presenting several spectral plots of the normalized graph Laplacian of different simulated artificial graphs and contrasting them with those of real networks from different domains, we show that the plots differ among graphs from different domains and classes. At the beginning, we discuss different methods for plotting the spectrum. In this chapter, we also show that the spectral plot is a useful and simple technique for visualizing the qualitative properties of a graph.

In Chapter 6, we investigate how and why structural properties are reflected by the spectrum of the normalized graph Laplacian. We systematically investigate the behavior of this spectrum under local and global operations like motif doubling, graph joining and splitting. The eigenvalue 1 plays a particular role, and we therefore emphasize those constructions that change its multiplicity in a controlled manner, like iterated duplication of nodes. Constructions with different graph operations related to the evolution of a network produce specific eigenvalues and describe certain processes of graph formation that leave characteristic traces in the spectrum. We show how useful and plausible hypotheses about evolutionary processes can be made by investigating the spectrum of a graph constructed from actual data.

From the spectral plot of the normalized graph Laplacian, the essential qualitative properties of a network can be simultaneously deduced. Given a class of empirical networks, reconstruction schemes for elucidating the evolutionary dynamics leading to the given data can be developed based on the evolutionary hypotheses made using the spectrum of the graph. This method is exemplified for protein-protein interaction networks in Chapter 7. We also identify traces of duplication and divergence processes in their evolutionary history, the typical specific features that robustly distinguish protein-protein interaction networks from other classes of networks, in spite of possible statistical fluctuations in the underlying data.

In Chapter 8, we introduce a tentative classification scheme for empirical networks based on qualitative global properties detected through the spectrum of the

(normalized) Laplacian of the graph underlying the network. It is shown that our method identifies several distinct types of networks across different domains of applications, indicating hidden regularity properties.

The computation of the whole spectrum of a large graph needs a lot of space and time. A scheme for coarsening a graph to reduce its size, is introduced in Chapter 9. The new graph produced by this method yields a spectral plot similar to that of the original graph. This substitute solution is proposed to reduce the computational complexity and space required for the computation while still giving an idea about the pattern of the spectral plot of a large graph.

The thesis concludes with Chapter 10, where the main conclusions of the analysis are described and possible directions for future research are proposed.

Chapter 2

Preliminaries

Before going on to explain the results of this thesis, we would like to look into the basics of graph theory (for more information and details on graph theory, see (Bollobás, 1998; Godsil & Royle, 2001; Merris, 2001)) and generic models of network theory. So in this chapter, basic notions and facts that are used to formulate and solve problems discussed in this thesis are explained. Some definitions are illustrated with short examples.

2.1 Basics of graph theory

The foundation of graph theory was influenced by the Königsberg bridge puzzle introduced by Euler in 1736 (Euler, 1736). Later, investigations of social problems by Erdős were a benchmark for the start of formal graph theory. Here we look into some basic notions of the subject.

Definition 2.1.1. A *graph* Γ (typically written as $\Gamma = (V, E)$) is an ordered pair of two sets, a non empty set $V = V(\Gamma)$, called vertex set, consisting of objects $i, j, \dots$, that are called *vertices* (sometimes also called *nodes*) and another set $E = E(\Gamma)$, called edge set, consists of edges. One *edge* connects two vertices. Two same vertices can be connected by multiple edges and one edge can connect one vertex to itself. The cardinality of the set V is called the *order* of the graph, and the cardinality of the set E is called the *size*[1] of the graph.

Definition 2.1.2. If $e = \{i, j : i \text{ and } j \in V(\Gamma)\} \in E(\Gamma)$, we call the vertices i and j *adjacent* (to each other) or *connected* (to each other) or *neighbors* (of one another). We denote this by $i \sim j$. The edge e can be represented as a pair of vertices, (i, j), and denoted by ij.

[1] In literature, *size* of the graph has also been used to mean the cardinality of the vertex set.

Definition 2.1.3. An edge (i, i) is called a *self-loop* or simply a *loop*. By the definition, there could be more than one edge with the same vertices; edges are called *parallel edges* or *multi-edges*. Fig. 2.1(a) shows a graph where edges e_6 and e_9 are parallel edges and edge e_3 is self loop.

A graph with no self-loops or multi-edges is called a *simple graph*. Fig. 2.1(b) is an example of simple graph.

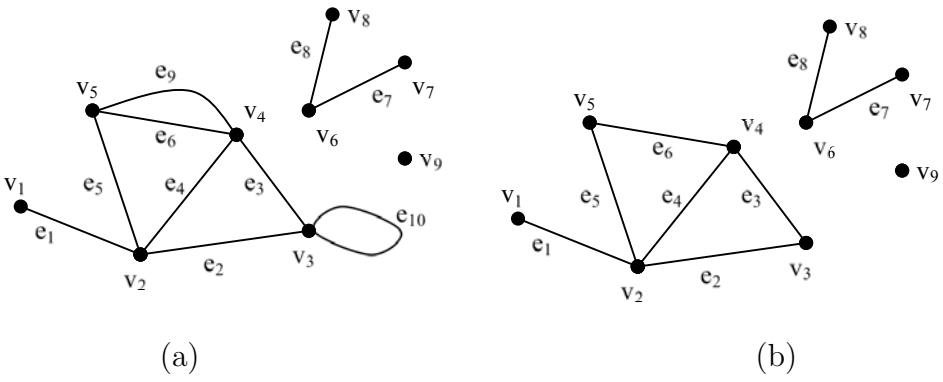


Figure 2.1: (a) A graph with 9 vertices and 10 edges. (b) A simple graph with 9 vertices and 8 edges.

Definition 2.1.4. For a simple graph, the *degree* n_i of a vertex i is the number of vertices wchich are adjacent to i. In Fig. 2.1(b), the vertices v_1, v_3, v_4 and v_5 are adjacent to vertex v_2. So the degree of the vertex v_2 is 4. The monotonic sequence of degrees of $V = V(\Gamma)$ is called *degree sequence* of the graph Γ. A vertex with degree zero is called an *isolated vertex*. A vertex with degree 1 is called a *pendant vertex*. In Fig. 2.1(b), vertex v_1 is a pendant vertex and vertex v_9 is an isolated vertex.

Definition 2.1.5. Let k_n be the number of vertices with degree n in a graph Γ. Then the distribution of k_n as a function of n is called the *degree distribution* of the graph Γ.

Definition 2.1.6. A graph $\Gamma' = (V', E')$ is called *subgraph* of a graph $\Gamma = (V, E)$, if $V' \subseteq V$ and $E' \subset E$. A subgraph of Γ *induced* by V' is $\Gamma[V'] = (V', E \cap V'^{(2)})$. For example, in Fig. 2.2, graph (b) is a subgraph of graph (a).

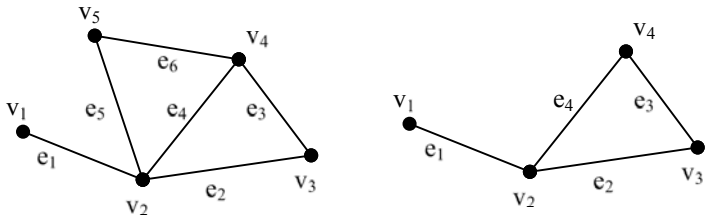


Figure 2.2: Graph (b) is a subgraph of graph (a).

Definition 2.1.7. Two graphs $\Gamma_1 = (V_1, E_1)$ and $\Gamma_2 = (V_2, E_2)$ are called *isomorphic* if there exists a bijection $\omega : V_1 \to V_2$ that posseses the adjacent vertices relation, i. e. , $i \sim j \iff \omega(i) \sim \omega(j)$, for all i, j. If there exists an isomorphism (other than the identity map) from a graph Γ to itself, then Γ is called *symmetric*. For instance, the two graphs in Fig. 2.3 are isomorphic. Here vertex p_l corresponds to vertex q_l, for $l = 1, \ldots, 5$. They are both also automorphic.

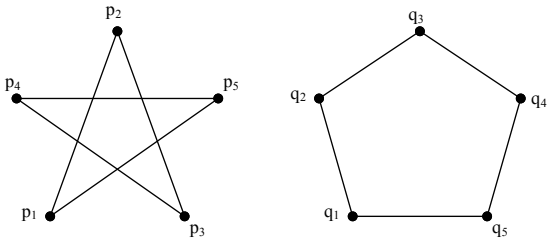


Figure 2.3: Two isomorphic graphs. They are symmetric also.

Definition 2.1.8. A graph is called *r-regular* or *regular of degree r* if each of its vertices have degree r. Fig. 2.4 is an example of a 3-regular graph.

Definition 2.1.9. A graph P is called *path* if it is of the form

$$V(P) = \{p_0, p_1, \ldots, p_l\}, \quad E(P) = \{p_0p_1, p_1p_2, \ldots, p_{l-1}p_l\}.$$

All vertices, $p_0, \ldots, p_l$ are not necessarily to be distinct. The vertices p_0 and p_l are called the *endvertices* of P. Specifically, vertex p_0 is called the *initial vertex*

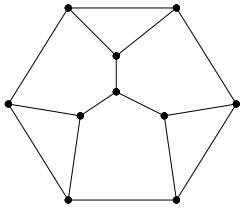


Figure 2.4: A regular graph of degree 3.

and vertex p_l is called the *terminal vertex*. The cardinality of the set $E(P)$, l, is called the *length* of P. Generally, a path of length N is denoted by P_N. Now, if all vertices, $p_1, \ldots, p_l$ are distinct and $p_0 = p_l$, the graph is called a *cycle*. The cardinality of the set E, l is called the *order* or *length* of the cycle C. Generally, a cycle of order N is denoted by C_N. Fig. 2.3 is an example of C_5.

Definition 2.1.10. A graph is called *connected* if, for every pair of vertices i and j, there exists a path where i and j are endvertices. Otherwise, the graph is called *disconnected*. A disconnected graph consists of more than one connected graphs, and these connected subgraphs are called the *components* of the disconnected graph. Fig. 2.2 shows examples of connected graphs, and Fig. 2.1 gives examples of disconnected graphs.

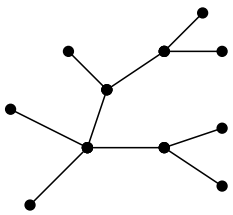


Figure 2.5: A tree.

Definition 2.1.11. A *tree* is a connected graph without any cycle. See Fig. 2.5 for an example.

Definition 2.1.12. A graph Γ is called a *bipartite graph* if the vertex set $V(\Gamma)$ can be decomposed into two disjoint subsets V_1 and V_2, such that each edge of Γ

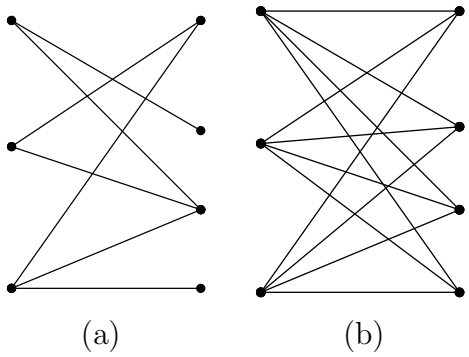


Figure 2.6: Bipartite graphs. (b) is a complete bipartite graph $K_{3,4}$.

connects a vertex in V_1 with a vertex in V_2. Hence there is no edge which joins two vertices in the same subset. Fig. 2.6 shows examples of bipartite graphs.

If every vetex of one subset is connected by edges with all vertices of other subset, then the bipartite graph is called a *complete bipartite graph*, and is usually denoted by $K_{m,n}$, where m and n are the cardinalities of the two subsets. Fig. 2.6(b) is an example of complete bipartite graph, $K_{3,4}$.

Definition 2.1.13. A *star* S_N is a complete bipartite graph where one subset of the vertex set contains only one vertex. Fig. 2.7 shows a star graph S_6.

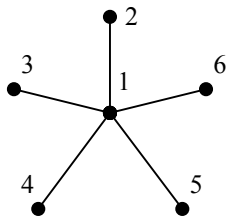


Figure 2.7: A star graph S_6.

Definition 2.1.14. The *path length* or *shortest distance* l_{ij} between two vertices i and j is the length of the shortest path (which is a subgraph of the original graph) where the vertices i and j are the endvertices. For both graphs in Fig. 2.2, the path length between vertices v_1 and v_3 is 2 (the length of the path v_1,v_2,v_3). The *average path length* of a connected graph is the average of all shortest distances over all vertex pairs.

Definition 2.1.15. The maximum over all shortest distances in a graph is called the *diameter*[2] of the graph. For example, the diameter of the graphs in Fig. 2.5 is 5, in Fig. 2.4 it is 3, in Fig. 2.7, Fig. 2.3, Fig. 2.2 it is 2, but in Fig. 2.8 it is 1.

Definition 2.1.16. A *cut-set* or *edge-cut* is a subset of edges whose removal disconnects the graph. Similarly, a *vertex-cut* is a subset of vertices whose removal disconnects the graph.

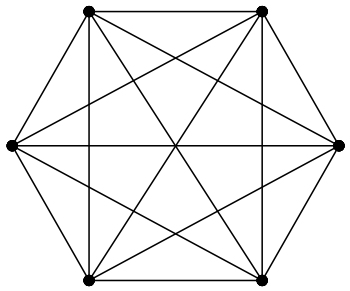


Figure 2.8: The completely connected graph of order 6, K_6.

Definition 2.1.17. A *totally* or *fully* or *completely connected graph* is a graph where every pair of vertices is connected by an edge. This graph is also called a *complete graph*. The complete graph of order N is denoted by K_N. Fig. 2.8 shows the graph K_6. A completely connected subgraph is called a *clique*.

Definition 2.1.18. A *directed graph* or *digraph* is consists of a set V of vertices $i, j, \ldots$ and a set of edges E which are ordered pairs (i, j) of vertices. We write the edge with ordered pair (i, j) as $i \to j$. Here, j is called the *head* or *terminal vertex*, and i is called the *tail* or *initial vertex* of the edge. The number edges with i as the initial vertex (resp. terminal vertex) is called the *outdegree* (resp. *indegree*) of the vertex i.

Transitivity or clustering coefficient: It has been observed in many graphs (networks) that neighbors of a vertex are connected to each other. In terms of the structural properties of a network, transitivity implies the presence of a large number of triangle–sets, or three vertices connected to one another. The quantitative definition of *transitivity* (sometimes called the *clustering coefficient*) is

$$C = \frac{3 \times \text{ number of tringles}}{\text{total number of connected triples of vertices}},$$

[2]In the literature, the term "diameter" sometimes also denotes the avarage path length.

i. e. , the fraction of triples that are triangles. So $0 \leq C \leq 1$. In simple terms, C is the mean probability that two vertices that have a common neighbor will themselves be neighbors. For the graph in Fig. 2.2(b) $C = \frac{3\times1}{5} = 3/5$.

Another definition of the clustering coefficient was proposed by Watts and Strogatz (Watts & Strogatz, 1998). Here $C = \frac{1}{N}\sum_i C_i$, where

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i} = \frac{2E_i}{n_i(n_i-1)}.$$

Where E_i is the number of edges between the neighbors of vertex i. Here, we calculate in reverse way – unlike the first definition, we have to find the mean of the ratio of triangle and triples, rather than the ratio of means. For the graph in Fig. 2.2(b), $C_1 = 0, C_2 = 1/3, C_3 = 1$, and $C_4 = 1$, so $C = 7/12$.

2.1.1 Connectivity matrices

A graph $\Gamma = (V, E)$ can be represented by different kinds of matrices. These matrices are called *connectivity matrices*. The eigenvalues of these connectivity matrices are important for analyzing the graph's structure. Here, we will have a look into a few such matrices, like the adjacency matrix, the Laplacian matrix and the normalized Laplacian matrix.
Let n_i and n_j be the degrees of the vertices i and j, respectively, of the graph Γ. Now we can define the aforementioned matrices as follows.

Adjacency matrix: The matrix $A = [a_{ij}]$ with the form

$$a_{ij} = \begin{cases} 1, & \text{if } ij \text{ is an edge} \\ 0, & \text{otherwise} \end{cases}$$

is called the adjacency matrix.

Laplacian matrix: The matrix $L = [a_{ij}]$ with the form

$$a_{ij} = \begin{cases} n_i, & \text{if } i = j \\ -1, & \text{if } ij \text{ is an edge} \\ 0, & \text{otherwise} \end{cases}$$

is called the Laplacian matrix.

Normalized Laplacian matrix: According to the normalization factor choosen, the normalized Laplacian matrix can have different forms. Here are examples of two different normalized Laplacian matrices.

1. The nomalized Laplacian matrix denoted by $\mathcal{L} = [a_{ij}]$ has the form

$$a_{ij} = \begin{cases} 1, & \text{if } i = j \text{ and } n_i \neq 0 \\ -\frac{1}{\sqrt{n_i n_j}}, & \text{if } ij \text{ is an edge} \\ 0, & \text{otherwise.} \end{cases}$$

2. The normalized Laplacian matrix $\Delta = [a_{ij}]$ has the form

$$a_{ij} = \begin{cases} 1, & \text{if } i = j \text{ and } n_i \neq 0 \\ -\frac{1}{n_j}, & \text{if } ij \text{ is an edge} \\ 0, & \text{otherwise.} \end{cases}$$

Relationship between these matrices

For a graph Γ, let D be the diagonal matrix with entries the degree of vertices, i. e. ,

$$D(i,j) = \begin{cases} n_i, & \text{if } i = j \\ 0, & \text{otherwise .} \end{cases}$$

The relation between the adjacency matrix A and the non-normalized Laplacian matrix L is

$$L = D - A.$$

The relationships with the normalized Laplacian matrix are

$$\begin{aligned} \mathcal{L} &= D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \\ &= I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}, \end{aligned}$$

where I is the identity matrix and $D^{-\frac{1}{2}}$ is a diagonal matrix with $D(i,i) = -\frac{1}{\sqrt{n_i}}$ and all other elements zero.

Relation between two normalized Laplacian matrices is

$$\Delta = D^{\frac{1}{2}} \mathcal{L} D^{-\frac{1}{2}}$$

These two matrices are similar and hence they have same spectrum.

2.1.2 Eigenvalues of a graph

Let A be the adjacency matrix of the graph Γ of order N. Let I be the identity matrix of order N, and let λ be a scalar. Then the determinant $|A - \lambda I|$ which is an ordinary polynomial in λ of N-th degree with scalar coefficients, is called the *characteristic polynomial* of Γ.

The roots of the equation

$$|A - \lambda I| = 0$$

are called the *eigenvalues* of the graph Γ (also of the matrix A). The set of eigenvalues is called the *spectrum* of the graph Γ. The multiplicity of an eigenvalue λ is called the *algebraic multiplicity* of λ.

The equation

$$Au = \lambda u \tag{2.1}$$

is called an eigenvalue equation. A nonzero solution u of the equation (2.1) is called an *eigenvector* or *eigenfunction* for the eigenvalue λ. The vector space constructed from the set of eigenvectors corresponding to a particular eigenvalue λ is called the *eigenspace* of λ. The dimension of the eigenspace of an eigenvalue λ is the *geometric multiplicity* of λ. For a symmetric matrix, the geometric and algebraic multiplicities of an eigenvalue are equal.

The *(normalized) Laplacian spectrum* of a graph is the set of all eigenvalues of its (normalized) Laplacian matrix.

Examples: For the graph $\Gamma = S_4$, the star graph with four vertices (Figure (2.9)), the connectivity matrices are as follows:

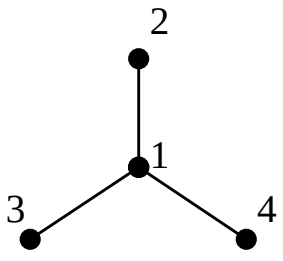


Figure 2.9: A star graph S_4 of order 4.

The adjacency matrix is

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

and the spectrum is $A(S) = \{-\sqrt{3}, 0, 0, \sqrt{3}\}$. The Laplacian matrix is

$$L = \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix},$$

and the spectrum is $L(S) = \{0, 1, 1, 4\}$. The normalized Laplacian matrices are

$$\mathcal{L} = \begin{pmatrix} 1 & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} & 1 & 0 & 0 \\ -\frac{1}{\sqrt{3}} & 0 & 1 & 0 \\ -\frac{1}{\sqrt{3}} & 0 & 0 & 1 \end{pmatrix}, \quad \Delta = \begin{pmatrix} 1 & -1 & -1 & -1 \\ -\frac{1}{3} & 1 & 0 & 0 \\ -\frac{1}{3} & 0 & 1 & 0 \\ -\frac{1}{3} & 0 & 0 & 1 \end{pmatrix},$$

and both have spectrum $\{0, 1, 1, 2\}$.

2.2 Three generic models

Very frequently, three different basic network models are reffered to in the literature. They are Erdős–Rényi's random graph model, Watts–Strogatz's small–world model, and Barabási–Albert's scale–free model. Due to their differing fundamental properties, these are often used for comparing with other models and considered to be basic models in network theory.

2.2.1 Erdős–Rényi's random graph

Erdős and Rényi proposed a very simple model (Erdős & Réanyi, 1959) for a random graph (network). The model is as follows:

For a graph constructed from N vertices, the existence of a connection between each vertex pair depends on a fixed probability p. That means that while constructing the network, one could make a decision to connect (resp. not connect) a pair of vertices by an edge with the probability p (or $1 - p$). So the expected

number of edges within the graph is $p\frac{N(N-1)}{2}$. The graph generated by this model with fixed N and p is usually denoted by $G_{N,p}$[3]. Most of the properties of this random graph have been studied for large graph size, i. e. , in the limit $N \to \infty$ (Erdős & Réanyi, 1959; Erdős & Réanyi, 1960; Erdős & Réanyi, 1961). One motivation behind taking the large limit is to keep the mean degree $\bar{n} = p(N-1)$ constant. The probability that a vertex has degree n is

$$p_n = \binom{N-1}{n} p^n (1-p)^{N-1-n} \simeq \frac{\bar{n}^n e^{-\bar{n}}}{n!}. \tag{2.2}$$

For fixed n, equality holds for $N \to \infty$. This is the reason why this graph is called a "Poisson random graph". Many internal structural properties emerge with different values of p. For instance, when p is small ($p < \frac{1}{N}$), there are few edges in the graph, so almost all vertices are disconnected. Most of the connected components are small and have size at most $O(\log N)$, which is independent of p. Now, if we increase the value of p, when it reaches $1/N$ (at $p = \frac{1}{N}$, $\bar{n} = 1$), suddenly a giant component appears in the graph. So there is a phase transition at $\bar{n} = 1$. For $p > \frac{1}{N}$, the graph tends to be connected. More precisely, when $p = \frac{\log N + b + o(1)}{N}$, for some positive real value b, the probability that the graph is connected tends to $exp(-e^{-b})$.

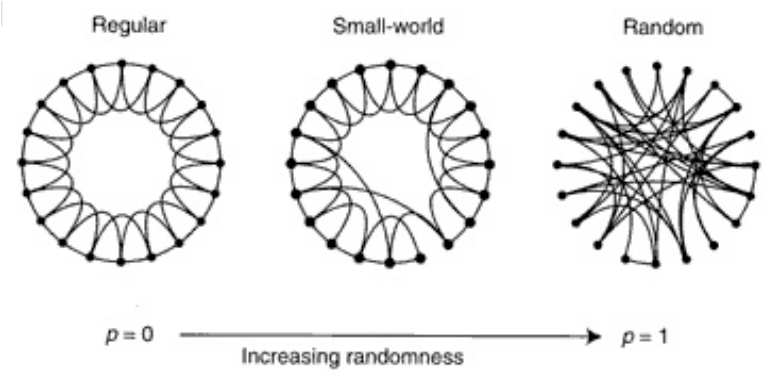


Figure 2.10: Transition from a lattice to a random network via small-world network. *Original figure is taken from (Watts & Strogatz, 1998).*

[3]Technically $G_{N,p}$ is the ensemble of all such graphs, where each graph with m edges appears with the probability $p^m(1-p)^{\frac{N(N-1)}{2}-m}$.

2.2.2 Watts–Strogatz's small–world network

To generate a graph with a high clustering coefficient which is independent of network size and a low average path length (the small world phenomenon (Milgram, 1967), popularly known as six degrees of separation (Guare, 1990)) Watts and Strogatz introduced a less sophisticated model (Watts & Strogatz, 1998). This model is based on the idea of rewiring a fraction, Pr, of edges within a regular lattice (Watts & Strogatz, 1998; Watts, 1999b; Watts, 1999a). This model could be constructed with a high-dimensional lattice, but most of the studies have been done on one-dimensional lattices. The construction scheme is as follows:

Take a regular ring lattice with N vertices where every node is connected to its first K neighbors ($K/2$ on either side). To get a sparse but connected network, it is better to consider $N \gg K \gg \ln(N) \gg 1$. Now rewire each edge of the lattice with probability $p = Pr$ (to exclude multiple edges and loops in the original model of Watts and Strogatz, only one end of an edge is rewired[4]). So $pNK/2$ edges will be rewired. A transition from order to randomness, i. e. , regular lattice to (an almost) random graph is observed when we vary p from 0 to 1 (see Fig. 2.10). Watts and Strogatz showed computationally that there exists a sizable region in between $p = 0$ and $p = 1$ where the model has the properties of low average path length and high clustering coefficient (see Fig. 2.11).

For the regular ring lattice (i. e. , when $p = 0$), the clustering coefficient is $\frac{3(K-2)}{4(K-1)}$, which tends to 3/4 for large K. The mean geodesic distance between vertices tends to $N/2K$ for large N. For $P = 1$, i. e. , where every edge is rewired, the graph is very similar to Erdős and Rényi's random graph, and has a very low clustering coefficient $\simeq K/N$ as well as typical geodesic distances of order $\ln(N)/\ln(K)$.

2.2.3 Barabási–Albert's scale–free network

Barabási and Albert proposed a growing model with preferential attachment of new vertex (Barabási & Albert, 1999). A graph constructed from this model follows a power law degree distribution (Barabási & Albert, 1999; Barabási, Albert, & Jeong, 1999). The model is as follows:

Start with a small number (m_0) of vertices. At each time step, add a new vertex to the network and connect that new vertex to $m(\leq m_0)$ already existing vertices in the network, but choosing any already existing vertex i with degree n_i

[4]In the model we have used in this work, both ends of an edge have been rewired, but self loops and multiple edges have always been avoided. This two-end rewiring does not change the statistical properties we are interested in.

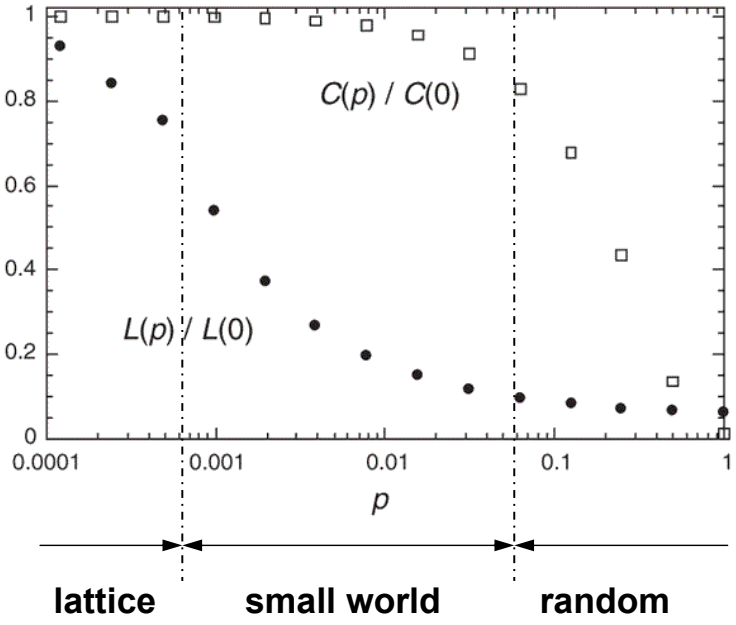


Figure 2.11: Characteristic path length $L(p)$ and clustering coefficient $C(p)$ for the family of randomly rewired graph described in Fig. 2.10. *Original figure is taken from the lecture notes of Réka Albert.*

depending on the probability $p = \frac{n_i}{\sum_j n_j}$.

Here we will see how the power-law property emerges in the degree distribution. For simplicity, we fix $m_0 = 1$, i. e., we start with a single isolated vertex. Let $n_{i,t}$ be the degree of a vertex i at time t. Let $p_{k,t}$ be the probability that a vertex, i has degree $(n_{i,t} =)$ k at time step t. Let the total number of vertices at time t be $N_t (= t)$ and $N_{k,t}$ be the number of vertices with degree k at time t. So, the probability that a new incoming vertex attaches to a node of degree k at time step $t + 1$ is:

$$p_{k,t} \frac{k}{\sum_i n_{i,t}} = \frac{k p_{k,t}}{2m N_{k,t}} \tag{2.3}$$

So, at each time step t, for a new arriving vertex with m edges,

$$\begin{aligned} N_{k,t+1} &= N_{k,t} + \frac{m(k-1)}{2mt} N_{k-1,t} - \frac{mk}{2mt} N_{k,t}, \text{ for } k > m, \\ N_{m,t+1} &= N_{k,t} + 1 - \frac{m^2}{2mt} N_{m,t}, \text{ for } k = m. \end{aligned} \tag{2.4}$$

Now $p_{k,t} = N_{k,t}/N_t = N_{k,t}/t \Rightarrow N_{k,t} = tp_{k,t}$. So

$$\begin{aligned}(t+1)p_{k,t+1} &= tp_{k,t} + \frac{(k-1)}{2}p_{k-1,t} - \frac{k}{2}p_{k,t}, \text{ for } k > m, \\ (t+1)p_{m,t+1} &= tp_{m,t} + 1 - \frac{m}{2}p_{m,t}, \text{ for } k = m.\end{aligned} \tag{2.5}$$

Solving p_k at the steady state, when $p_{k,t} = p_k$, we get

$$\begin{aligned}p_k &= \frac{(k-1)}{(k+2)}p_{k-1}, \text{ for } k > m \\ p_m &= \frac{2}{(m+2)}, \text{ for } k = m.\end{aligned} \tag{2.6}$$

With recursion, we get

$$p_k = \frac{(k-1)(k-2)\dots(m)}{(k+2)(k+1)\dots(m+3)}p_m = \frac{2m(m+1)}{(k+2)(k+1)k}. \tag{2.7}$$

So for $k \gg 1$, $p_k \sim k^{-3}$.

Chapter 3

Difficulties and Challenges in Analyzing Network Structure

3.1 Different real networks

Different networks have been constructed to study real systems. Here we shall give some idea about the construction principles of these networks. As we discussed before, to build a network from real system, one needs to decide what could be considered to be nodes (vertices) and what edges. According to the area and nature of the system, networks could be categorized roughly into four sections: social, information, technical and biological networks. Networks represented by movie actors, company directors, scientific co-authorship, telephone calls, email, friendships, and sexual contacts could be considered social networks; WWW, citation, and word co-occurance could be considered information networks; internet, power grids, train routes, electronic circuits, and software packages could be considered technological networks; metabolic and biochemical reactions, protein-protein interactions, gene regulation, food web, neural connectivity could be considered biological networks. Here are a few examples of these networks.

- *Movie actor collaboration network:*

  These networks have been constructed on the basis of information (movies and their casts) provided by different internet movie databases. Here, actors and actresses are considered as nodes, and two nodes are connected by an edge if the corresponding actors or actresses acted in the same movie. The type of the network is undirected.

- *Scientific collaboration network:*

Scientists are considered as nodes in this network, and the criteria for having an edge between two nodes is the joint publication of any article by the corresponding scientists. Another way to construct this network is by considering both scientists and articles as nodes and considering authorship between a scientist and an article as an edge. In this network, there is no edge that directly connects two scientists or two articles. This network (graph) is thus bipartite. The type of these networks is undirected.

- *Human sexual contacts network:*

  Here the nodes are men and women, and a sexual contact between two persons (nodes) is considered as an edge. These networks are used to study the transmission of infectious diseases. The nature of the network is undirected.

- *Telephone call network:*

  In this network, different telephone numbers are nodes, and any call from a caller number to receiver number is considered as a directed edge. So the networks here are directed.

- *World-Wide Web network:*

  The WWW has been represented as a large network. Here, the nodes are webpages and a hyperlink (URL) that links one webpage to another is considered as an edge. The type of the network is directed.

- *Citation network:*

  This network is usually constructed from a citation pattern of scientific publications. Here a published article is considered as a node and a directed edge stands for a reference from one article to another. So the networks are directed here.

- *Word adjacency network:*

  These networks are generally constructed from any particular piece of writing (like a book, journal, website etc.). Here, different words are considered as nodes and there is a connection between two words if they appear next to or one word apart from one another in a sentence. Another way to construct these networks is where edges are considered as being between two synonymous words. These networks can be either directed or undirected.

- *Internet network:*

Here, networks have been constructed based on the physical connections between computers and other telecommunication devices. Generally these networks have been considered with two different levels. At the router level, routers are considered as nodes and physical connections between them are considered as edges. At the inter-domain level (autonomous system), nodes are domains, composed of routers and computers, and connections between domains through routers are represented as edges. The type of these networks is undirected.

- *Metabolic network:*

  In a metabolic network, substrates or metabolites are considered as nodes and predominantly directed reactions in which these substrates are taking part are represented as edges. These networks are directed, but undirected versions of these networks have also been studied extensively.

- *Protein-protein interaction network:*

  Here proteins are considered as nodes and two proteins (nodes) are connected by an edge if there is a direct physical interaction (binding) between them. This network is undirected.

- *Protein folding network:*

  There are different consecutive conformational changes of a protein during its folding. In this network, the conformational states are represented by nodes, and two states are connected by an edge if they can be obtained from each other by an elementary change. There is also another type of protein folding network, constructed when one protein is in its secondary or tertiary state. Here, C_α carbons are considered as nodes, and two C_α are connected if their spatial Euclidean distance is lower than some threshold. These networks are undirected.

- *Gene regulatory network:*

  Here, nodes are different genes, and there is a connection between two nodes (genes) if one has a direct influence on the regulation of another. The network type is directed.

- *Food web network:*

  In these networks, nodes are different species and the predator-prey relationships are represented as edges. The networks are directed.

- *Neural network:*

  Here, neurons are considered as nodes, and direct synaptic connections between them are represented by directed edges. Thus, these networks are directed.

3.2 Complications in structural analysis

As we just discussed, many real data sets are, or can be, represented as networks, that is, in terms of the formal structure of a graph, where the vertices of a graph stand for the units in question, e. g. , in biological networks genes, proteins, cells, and neurons, and an edge between vertices expresses some correlation or interaction between the corresponding units. These edges can be directed to encode the direction of interaction, for example via a synaptic connection between neurons, and weighted to express the strength of interaction, like a synaptic weight. Here, for simplicity of presentation, we only consider the simplest type of a graph, the undirected and unweighted one, although our methods apply to and our considerations remain valid in the general situation. Thus, an edge expresses the presence of some interaction, connection or direct correlation between two vertices, regardless of its direction or strength. Clearly, this abstraction may neglect many important details, but we are concerned here with what it preserves.

When analyzing a network with graph theory, one has to remember that the number of non-isomorphic graphs is bewilderingly high. Thus, it becomes impractical, if not impossible, to list all different graphs with a given number of vertices, unless that number is rather small. Also, drawing a graph with a large number of vertices is not helpful for visual analysis, because the graph will just look too convoluted and complicated to make its structure transparent. So perceiving the structural topology on the basis of visualizing the graph is rather difficult.

In graph theory, many concepts have been developed that capture various quantitative or qualitative aspects of a graph. Various parameters (degree distribution, average path length, diameter, clustering coefficient, betweenness centrality, etc.) are considered in order to analyze the graph's structure. In recent years, many empirical network studies have based investigations on these parameters. Such studies managed to identify certain rather universal features valid for networks across a wide range of disciplines, like scale-free degree distributions. Conversely, on this basis, many algorithms have been developed that, perhaps after fitting certain free parameters, can construct networks with the same qualitative properties

and values for such variables as real networks. But graphs can be qualitatively quite different, and understanding this is obviously crucial for the analysis of the structure of the real networks to be represented. For example, the maximal distance (number of edges) between two vertices in a graph of size N can vary between 1 and $N-1$, depending on the particular graph. When the graph is complete, any two vertices have the distance 1, whereas for a chain, the first and the last vertex have distance $N-1$. For most graphs, of course, some intermediate value will be realized, and one knows from the theory of random graphs that for a typical graph this maximal distance is of the order $\log N$. So this maximal distance is one graph invariant, but still, rather different graphs can have the same value of this invariant. Adjoining a long sidechain to a complete subgraph can produce the same value as an everywhere loosely connected, but rather homogeneous graph. The question then emerges as to whether one should look for other, more comprehensive, invariants, or whether one should adopt an entirely different strategy for capturing the essential properties of some given graph. In fact, there are many graph invariants that each capture certain important qualitative aspects and that have been extensively studied in graph theory (see, e. g. , (Bollobás, 1998; Godsil & Royle, 2001)). These range from rather simple and obvious ones, like maximal or average degree of vertices or distance between them, to ones that reflect more global aspects, like how difficult it is to separate the graph into disjoint components (see, e. g. , (Chung, 1997)), commmunities (e. g. , (Newman, 2003)) or classes, or how difficult it is to synchronize coupled dynamics operating at the individual vertices (e. g. , (Jost & Joy, 2001)). For the sake of the subsequent discussion, we will call these properties cohesion and coherence, resp.

Recently, the power law behavior of the degrees has become quite popular, as it seems to be rather ubiquitous in biological and other data (see (Barabási & Albert, 1999)). Thus, scale-freeness seems a more or less universal feature among graphs coming from empirical data in a wide range of domains. Another powerful invariant of the graph is its first eigenvalue, which provides estimates for how difficult it is to cut the graph up into disjoint components (see (Chung, 1997), or for how easily dynamics at the vertices can be synchronized, (Atay, Jost, & Wende, 2004; Jost & Joy, 2001; Pecora & Carroll, 1990; A.Pikovsky, M.Rosenblum, & J.Kurths, 2001) and many other articles). These invariants are useful for analyzing particular structural (or dynamical) properties of a graph, but cannot capture all qualitative aspects of the graph. For example, graphs with the same degree distribution can have completely different inheritance and dynamical properties (like synchronizibility). Also, by their very nature, universal properties like a power law degree distribution capture what is common to large classes of

graphs, but fail to identify what is specific about graphs from a particular domain, and what distinguishes such graphs qualitatively from those of other fields. Also, depending on the details of the preferential attachment rule chosen, invariants like the average or maximal distance can vary widely.

3.2.1 Central questions

More recently the increasing availability of large data sets produced from real systems raises certain systematic questions, or, more precisely, brings new aspects to some old scientific issues. These, or at least the ones we wish to address in this chapter, are:

1. Given a particular structure, which features or qualities are universal, that is, shared by other structures within a certain class, and what is unique and specific to the structure at hand?

2. Given a large and complex structure, should we focus on particular aspects and quantities in detail, or should we try to obtain, at least at some coarse level, a simultaneous representation of all its qualitative features?

Some other questions also emerge:

- Do there exist systematic structural differences, e. g. , between protein-protein interaction, gene regulation and neural networks?

- Can one identify the domain of a given empirical graph on the basis of certain unique qualitative features?

In search of these answers, we are advocating a set of graph invariants that, on the one hand, give a complete qualitative characterization of a graph, and on the other hand, can be easily graphically represented and therefore visually analysed and compared.
This set is the spectrum of the graph Laplacian.

Chapter 4

Spectrum of the Graph Laplacian

Before going on to investigate the strength of the spectrum of the graph Laplacian for analyzing graph structure, we briefly look into the spectral analysis of a graph to gain some knowledge about different approaches in spectral graph theory and existing results.

4.1 Introduction to spectral analysis

The spectral analysis of graphs is not new. It has been one of the central attractions in the field of chemistry, mathematics and physics (Biggs, Lloyd, & Wilson, 1976; Cvetković, Doob, & Sachs, 1995). After the fundamental papers of Lihtenbaum (1956) and of Collatz and Sinogowitz (Collatz & Sinogowitz, 1957) (1957), spectra of graphs have been appearing in the mathematical literature very frequently. Even before, spectral analysis of graphs had been used in theoretical chemistry after the introduction of graph spectra in the thesis of Hückel (Hückel, 1931) (in 1931). One of the important applications of the spectrum of a graph in quantum chemistry is calculating the energy levels of electrons in hydrocarbons. Also, the stability of such molecules is studied with the graph spectrum and corresponding eigenvectors. Spectral graph theory plays an important role in theoretical physics and quantum mechanics. It is used in minimizing the energies of Hamiltonian systems. For a long time, the eigenvalues of a graph have been of deep interest for combinatorics and graph theory. In computer science, spectral techniques are used intensively for a wide range of problems. Graph spectra play a vital role in solving varies problems in communication networks. In general, physicists and chemists are interested in calculating the spectra from a graph for certain purpose, whereas graph theorists and combinatorialists are interested about the graph structure of a given spectra. Recently, dynamical properties like synchronization have been studied with the

help of the graph spectrum (Atay & Biyikoglu, 2005; Atay, Biyikoglu, & Jost, 2006; Atay et al., 2004; Atay, T.Bıyıkoğlu, & J.Jost, ; Jost & Joy, 2001).

4.1.1 Spectral analysis of graphs in mathematics

As I mentioned above, spectral graph theory has a long history. In the past, linear algebra and matrix theory have been used to analyze the adjacency matrix. (For details on algebraic aspects of spectral graph theory, see (Biggs, 1993; Cvetković, Doob, Gutman, & Torgašev, 1988; Cvetković, Doob, & Sachs, 1980; Cvetković et al., 1995; Seidel, 1989).). In the last twenty years, a new approach to analyzing spectral graph theory has arisen from a more geometric perspective. For instance, the isoperimetric property and eigenvalues of the graph play a major role in the explicit construction of expander graphs (Lubotzky, Phillips, & Sarnak, 1988; Margulis, 1984). Random walks and rapidly mixing Markov chains have been studied by rigorously applying the discrete analogue of the Cheeger inequality (Seidel, 1989). The interaction with differential geometry has been an important development in spectral graph theory. The analogy between spectral Riemannian geometry and spectral graph theory is very interesting for mathematicians. Spectral graph theory has benefitted from the powerful tools and methodology developed from the concepts and methods of spectral geometry. The importance of this is that it is one of the main interest in graph theory to deduce the inheritance structural properties of the graph from its spectra.

4.1.2 Spectral density of graphs

The study of the characteristics of the spectrum of a random matrix by Wigner in 1955 (Wigner, 1955) has increased interest in the behavior of the spectral distributions of different matrices and graphs. Wigner introduced a law, known as Wigner's semicircle law (Wigner, 1955; Wigner, 1957; Wigner, 1958). According to this, the distribution of eigenvalues of a large real symmetric matrix, with elements taken from a probability distribution, follow a semicircle distribution. (See bellow for the details of this law). In 1981, Füredi and Komlós (Füredi & Komlós, 1981) showed that the spectral density functions of Erdős and Rényi's random graphs follow Wigner's semicircle law. Afterwards Goh, Kahng and Kim (Goh, Kahng, & Kim, 2001) studied the spectra and eigenvectors of the adjacency matrix of Barabási and Albert's scale-free graph. They have found that the distribution of the spectra is quite far from a semicircle and the eigenvalues decay exponentially around the center and have power-law long tails at both spectrum edges. Farkas and others (Farkas, Derenyi, Barabási, & Vicsek, 2001) have also described that,

instead of following a semicircle law, the spectral density of Barabási and Albert's scale-free graph looks like a triangle-shaped curve when plotted in log-log scale. For Watts and Strogatz's small-world graph, they have observed that the spectral distribution consists of several sharp peaks (depending on the rewiring probability) in the centre and is also different from a semicircle. Later Chung, Lu and Vu (Chung, Lu, & Vu, 2003) showed that, depending on the matrix, one could get different spectral plots for power-law graphs. Under a mild condition (that the minimum expected degree is significantly larger than the square root of the expected average degree), they have proved that the eigenvalues of the normalized Laplacian of a random power-law graph with given expected degrees follow a semicircle law, whereas the spectrum of the adjacency matrix of the same graph obeys a power-law (see, (Chung et al., 2003) for details). They have mentioned that the spectrum of the normalized Laplacian matrix reflects global properties of the graph, and the spectrum of the adjacency matrix contains information about local properties of the graph. Also, it has been reported that the largest k eigenvalues of the adjacency matrix of random power-law graphs have a power-law distribution (provided that the largest k degrees are large in the terms of the second-order average degree) (Chung et al., 2003; Farkas et al., 2001; Goh et al., 2001; Mihail & Papadimitriou, 2002). So, for Erdős and Rényi's random graph, the spectral distribution of the three matrices (adjacency, Laplacian and normalized Laplacian) are very similar, but for any other (non-regular) graph they can be remarkably different.

Here, my study is based on the normalized Laplacian matrix. So from now onwards, all the terminology of spectral graph theory will refer to the normalized Laplacian matrix (e. g. ,"spectrum of a graph" will mean the spectrum of the normalized Laplacian matrix of the graph). Now, before explaining the development of new results and application of this tool, I would like to give a short review of the properties of the eigenvalues of the normalized Laplacian matrix. So in the next section, I shall discuss previous results on the normalized Laplacian matrix, considered as the graph Laplacian operator.

Wigner's Semicircle Law

Let V be a real symmetric matrix of large order N having random elements v_{ij} that for $i \leq j$ are independently distributed with equal densities, equal second moments m^2, and nth moments bounded by constants B_n independent of i, j, and N. Further, let $S = S_{\alpha,\beta}(v, N)$ be the number of eigenvalues of V that lie in the

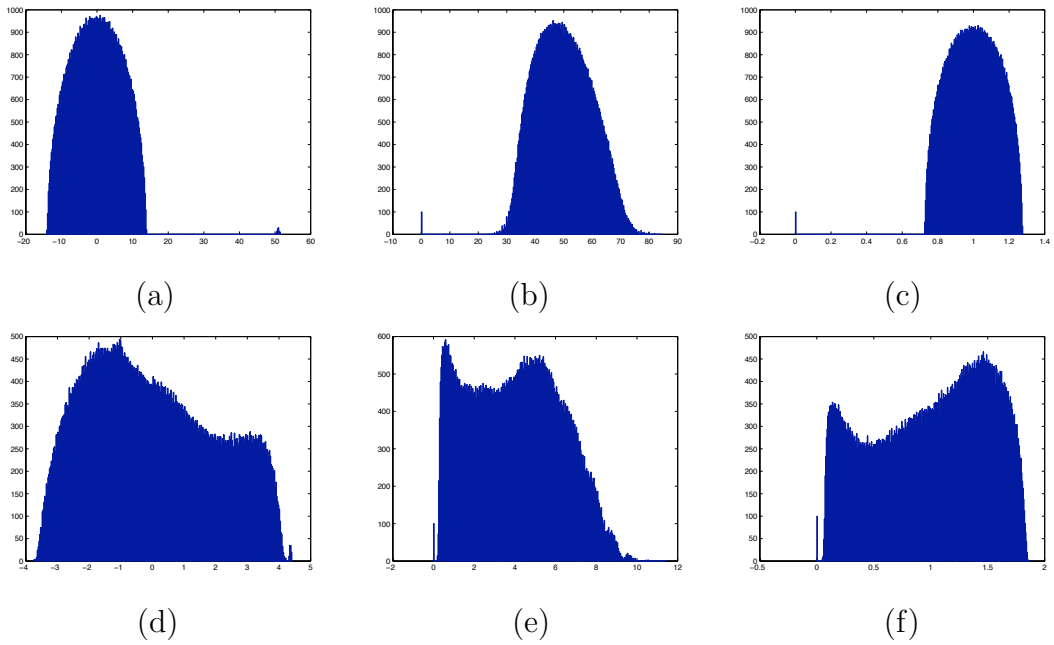


Figure 4.1: Plots of spectra–histograms (sum over 100 realizations) of generic networks. Random network by Erdős–Rényi model with $p = 0.05$ of (a) adjacency matrix, (b) Laplacian matrix, (c) normalized Laplacian matrix. Small-world network by Watts–Stogatz model (rewiring a regular ring lattice of average degree 4 with rewiring probability 0.3) of (d) adjacency matrix, (e) Laplacian matrix, (f) normalized Laplacian matrix. Number of bins within the interval $[\lambda_{min}, \lambda_{max}]$ is 317. Size of all networks is 1000.

interval $(\alpha N^{1/2}, \beta N^{1/2})$ for real $\alpha < \beta$. Let $E(S)$ be the expected value of S. Then

$$\lim_{N\to 0} \frac{E(S)}{N} = \frac{1}{2\pi m^2} \int_{\alpha}^{\beta} \sqrt{4m^2 - x^2}$$

This law is known as Wigner's semicircle law. This law was first observed by Wigner (1955) for certain special classes of random matrices arising in quantum mechanical investigations.

4.2 Eigenvalues of the normalized graph Laplacian

First, we will present some basic ideas about this operator (the normalized graph Laplacian), as well as some basic properties of its eigenvalues, known from the previous works (Chen et al., 2005; Chung, 1997; Jost & Joy, 2001).

Let Γ be a connected graph with vertex set $V = \{i : i = 1, \dots N\}$. Two vertices $i, j \in \Gamma$ are called neighbors, written $i \sim j$, if they are connected by an edge of Γ. For a vertex $i \in \Gamma$, let n_i be its degree, that is, the number of its neighbors.

Let u be a real-valued function on Γ, that is,

$$u : V \to \mathbb{R}$$

Now consider the space $L^2(\Gamma)$ of such functions, with the product

$$(u, v) := \sum_i n_i u(i) v(i) \tag{4.1}$$

So the corresponding norm in $L^2(\Gamma)$ is

$$\|u\| = (u, u)^{1/2} = \Big(\sum_i n_i u(i)^2 \Big)^{1/2} \tag{4.2}$$

Now we can get a Hilbert space $L^2(\Gamma)$ with this norm. In general, to find an orthogonal basis, consider an operator (graph Laplacian) (Banerjee & J.Jost, c; Jost, 2007; Jost, to appear; Jost & Joy, 2001)

$$\Delta : L^2(\Gamma) \to L^2(\Gamma)$$

$$\Delta u(i) := u(i) - \frac{1}{n_i} \sum_{j, j \sim i} u(j). \tag{4.3}$$

Remark. This operator has the same spectrum as the operator investigated in (Chung, 1997),

$$\mathcal{L} u(i) := u(i) - \frac{1}{n_i} \sum_{j, j \sim i} \frac{1}{\sqrt{n_i n_j}} u(j).$$

However, the spectrum is not the same as the operator usually studied in the graph theoretical literature, the (algebraic) graph Laplacian (see, e. g. , (Bollobás,

1998; Godsil & Royle, 2001; Merris, 1994; Mohar, 1997; T.Bıyıkoğlu, J.Leydold, & P.Stadler, 2007)):

$$Lu(i) := n_i u(i) - \sum_{j,j\sim i} u(j).$$

The normalized Laplacian is the operator underlying random walks on graphs, and it naturally incorporates a conversion law.

4.2.1 Important properties of this operator

1. Δ is self-adjoint with respect to (., .):

   $$(u, \Delta v) = (\Delta u, v)$$

   for all u, $v \in L^2(\Gamma)$.

2. Δ is nonnegative :

   $$(\Delta u, u) \geq 0$$

3. $$\Delta u = 0 \text{ when } u \text{ is constant}$$

4.2.2 Eigenvalues of this operator

- The property (1) of Section 4.2.1 implies all eigenvalues of Δ are real.
- The property (2) of Section 4.2.1 implies all eigenvalues of Δ are nonnegative. Now the eigenvalue equation becomes

  $$\Delta u - \lambda u = 0 \tag{4.4}$$

  A non-zero solution u is called an eigenfunction for the eigenvalue λ. Since Γ has N vertices, the function space on which Δ operates is N-dimensional. Therefore, it has N eigenvalues; some of them might occur with multiplicity > 1. The eigenfunctions corresponding to the eigenvalue λ constitute a vector space whose dimension is the multiplicity of the eigenvalue λ. So any suitable element of this vector space is an eigenfunction for the eigenvalue λ.

  Since all eigenfunctions are orthogonal to each other and the eigenfunction u_0 corrosponding to the eigenvalue zero, for any other eigenfunction u, $(u_0, u) = 0$. this implies

  $$\sum_i n_i u(i) = 0 \tag{4.5}$$

- Property (3) of Section 4.2.1 implies that the smallest eigenvalue is $\lambda_0 = 0$.

Since we assumed that Γ is connected, the other eigenvalues are greater than zero[1] i. e. ,

$$\lambda_k > 0 \text{ for } k > 0. \tag{4.6}$$

Let us order the eigenvalues in nondecreasing order like

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N-1}$$

The highest eigenvalue is bounded above.

$$\lambda_{N-1} \leq 2, \tag{4.7}$$

and equality holds *iff* the graph is bipartite. And the difference between 2 and the largest eigenvalue estimates how different the graph is from a bipartite graph. Another property of the spectrum of a bipartite graph is that if λ is an eigenvalue, then $2 - \lambda$ is also an eigenvalue of the graph.

Now, if the graph is complete, then

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{N-1} = \frac{N}{N-1}, \tag{4.8}$$

so

$$0 \leq \lambda_1 \leq \frac{N}{N-1} \leq \lambda_{N-1} \leq 2 \tag{4.9}$$

holds[2]. But if the graph is not complete, then λ_1 is always ≤ 1[3].

The precise value of

$$\lambda_1 = \min\left\{ \frac{\sum_{i,j;j\sim i} \big(u(i) - u(j)\big)^2}{\sum_i n_i u(i)^2} : \sum_i n_i u(i) = 0 \right\}, \text{ provided } u \text{ is not identically zero.} \tag{4.10}$$

and

$$\lambda_{N-1} = \max\left\{ \frac{\sum_{i,j;j\sim i} \big(u(i) - u(j)\big)^2}{\sum_i n_i u(i)^2} : \sum_i n_i u(i) = 0 \right\}, \text{ provided } u \text{ is not identically zero.} \tag{4.11}$$

[1] In general, the multiplicity of the eigenvalue 0 equals the number of connected components of Γ, with the corrosponding eigenfunctions being $\equiv 1$ on one and $\equiv 0$ on all other components.

[2] λ_1 and λ_{N-1} are important for studying the synchronization of a graph (Jost & Joy, 2001).

[3] This means $\lambda_1 \notin (1, \frac{N}{N-1})$ for any graph.

4.2.3 λ_1 and the Cheeger constant

λ_1 is a powerful invariant of a graph. It carries an estimate of how difficult it is to cut up the graph into two disjoint components.

One of the method for breaking a graph Γ into two components Γ_1 and Γ_2 is governed by the Cheeger constant. This constant was introduced by Cheeger in the context of Riemannian geometry, and that is given by

$$h(\Gamma) := \inf \left\{ \frac{|E_0|}{\min(\sum_{i\in\Gamma_1} n_i, \sum_{j\in\Gamma_2} n_j)} \right\} \tag{4.12}$$

where the infimum is taken over subsets E_0 of edges, such that removing E_0 disconnects Γ into components Γ_1 and Γ_2, and $|E_0|$ is the cardinality of the set E_0.

The relation between λ_1 and $h(\Gamma)$ for a connected graph is

$$\frac{1}{2}h(\Gamma)^2 \leq \lambda_1 \leq 2h(\Gamma). \tag{4.13}$$

4.2.4 Some more properties of the eigenvalues

1. $\sum_i \lambda_i \leq N$, and equality holds *iff* the graph is connected.

2. The spectrum of (disconnected) graph is the union of the spectra of its connected components.

3. For a connected graph,
$$\lambda_1 \geq \frac{1}{D \text{ vol } \Gamma},$$
where D is the diameter of the graph Γ and vol $\Gamma = \sum_i n_i$. But, if $D \geq 4$
$$\lambda_1 \leq 1 - 2\frac{n_{i_{max}} - 1}{n_{i_{max}}}(1 - \frac{2}{D}) + \frac{2}{D}$$
where $n_{i_{max}}$ is the maximum degree of the graph.

4. If $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{N-1}$ and $0 = \lambda'_0 \leq \lambda'_1 \leq \cdots \leq \lambda'_{N-1}$ are the eigenvalues of the graphs Γ and Γ' respectively, where $\Gamma' = \Gamma - e$ (e an edge of Γ), then
$$\lambda_{i-1} \leq \lambda'_i \leq \lambda_{i+1} \text{ for } i = 0, \ldots, N-1,$$
where $\lambda_{-1} = 0$ and $\lambda_N = 2$.

5. If $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{N-1}$ and $0 = \lambda'_0 \leq \lambda'_1 \leq \cdots \leq \lambda'_{N-1}$ are the eigenvalues of the graphs Γ and Γ', respectively, where Γ' is obtained from Γ by identifying two verties p_1 and p_2 which have no common neighbor, then

$$\lambda_{i-1} \leq \lambda'_i \leq \lambda_{i+1} \text{ for } i = 0, \ldots, N-1,$$

where $\lambda_{-1} = 0$ and $\lambda_N = 2$.

6. If a graph Γ of N vertices is t edges away from K_N, then $N/(N-1)$ will be an eigenvalue of Γ with multiplicity at least $N - 2t - 1$.[4]

7. If Γ is a graph of $m + n$ vertices and is obtained from $K_{m,n}$ by deleteing at most t edges, then 1 will be an eigenvalue of Γ with multiplicity at least $m + n - 2(t + 1)$.[5]

4.2.5 Eigenvalues of some elementary graphs

1. For the simple path P_N of N vertices, the eigenvalues are $1 - \cos \frac{\pi k}{N-1}$, where $k = 0, 1, \ldots, N-1$.

2. For the simple cycle C_N of N vertices, the eigenvalues are $1 - \cos \frac{2\pi k}{N}$, where $k = 0, 1, \ldots, N-1$.

3. For the complete bipartite graph $K_{m,n}$ of $m+n$ vertices, the eigenvalues are 0, 1 (with multiplicity $m + n - 2$) and 2. A special case of this graph is star S_N, which is a complete bipartite graph, $K_{1,N-1}$ and have eigenvalues 0, 1 (with multiplicity $N - 2$) and 2.

4. For the complete graph K_N of N vertices, the eigenvalues are 0 and $\frac{N}{N-1}$ (with multiplicity $N - 1$).

5. For the n-cube Q_n of 2^n vertices, the eigenvalues are $\frac{2k}{n}$ (with multiplicity $\binom{n}{k}$), where $k = 0, \ldots, n$.

[4]This result is only useful for graphs having more than $\frac{N^2}{2} - N + \frac{1}{2}$ edges.

[5]Some (real) networks are almost bipartite graphs, but far from being complete bipertite graphs. For these graphs, this result is not useful at all.

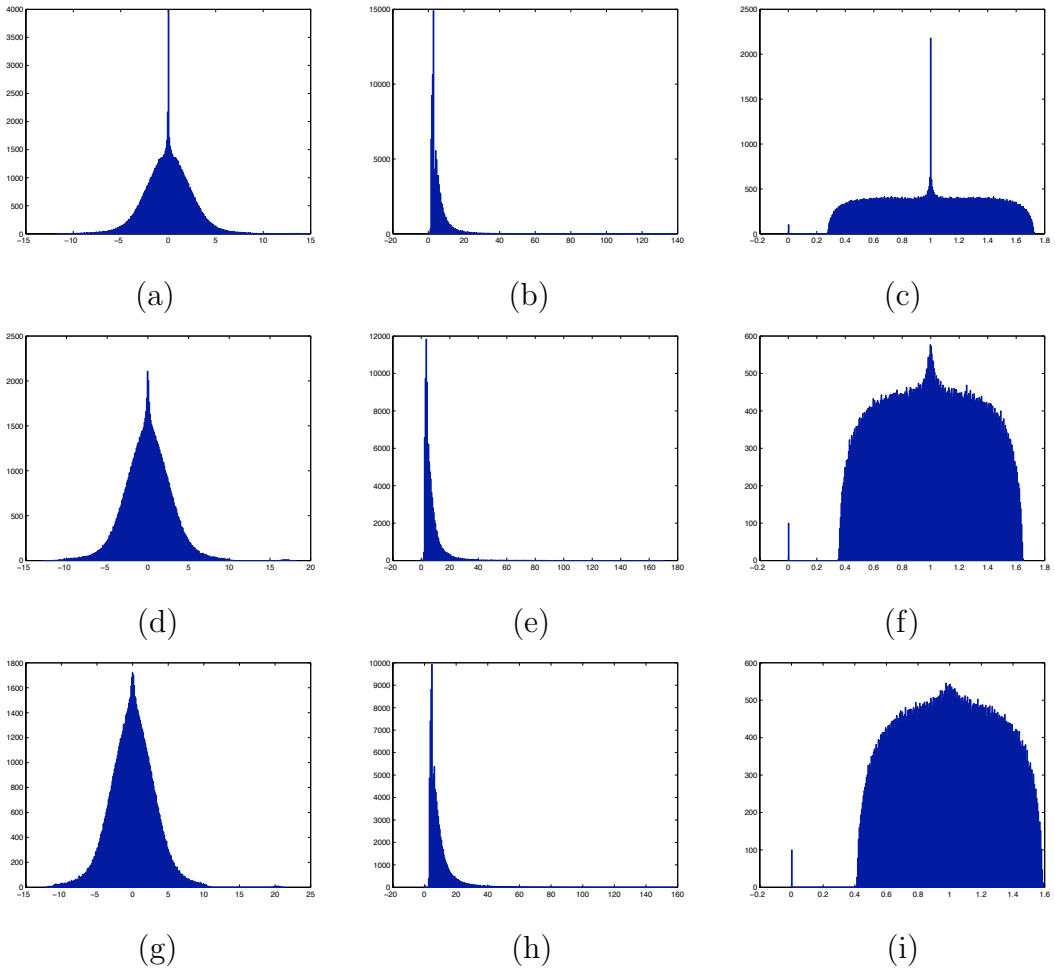


Figure 4.2: Plots of spectra–histograms (sum over 100 realizations) of generic networks. Scale-free network by Barabási–Albert model ($m_0 = 5$ and $m = 3$) of (a) adjacency matrix, (b) Laplacian matrix, (c) normalized Laplacian matrix. Scale-free network by Barabási–Albert model ($m_0 = 5$ and $m = 4$) of (d) adjacency matrix, (e) Laplacian matrix, (f) normalized Laplacian matrix. Scale-free network by Barabási–Albert model ($m_0 = 5$ and $m = 5$) of (g) adjacency matrix, (h) Laplacian matrix, (i) normalized Laplacian matrix. Number of bins with in the interval $[\lambda_{min}, \lambda_{max}]$ is 317. Sizes of all networks is 1000.

Chapter 5

Spectral Plots of Real Networks

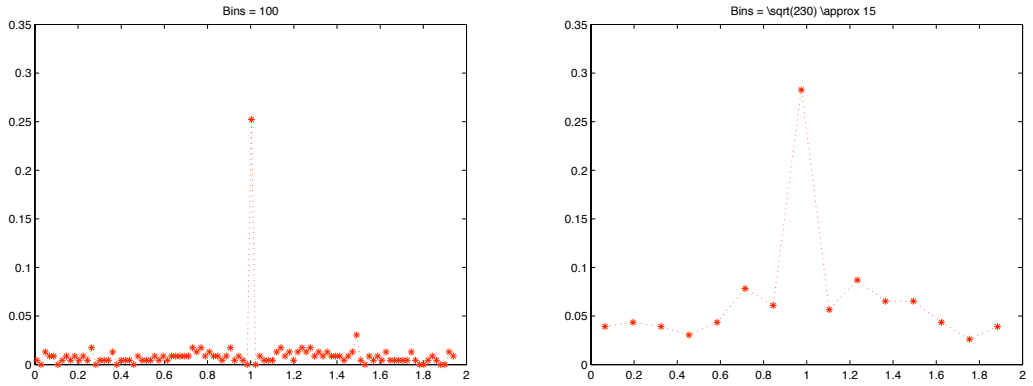
5.1 Discussion of spectral plotting

We have seen that spectrum differs with the different graphs. Now, we will explore the diversity of the spectral plots in this chapter with some examples. But before going on to investigate the relationships between graph structure and the spectral plot, it is important to find better plotting methods. Though the eigenvalues are bounded within $[0, 2]$, the number of eigenvalues varies among networks of different sizes. So one particular way cannot be perfect for all kind of networks. A simple way to plot is with a histogram or relative frequency plot with the desired number of bins. A very rough pattern of the spectrum could be realized by taking the square root of the size of the network as the number of bins within the interval [0, 2]. Sometimes a nice plot can be produced with 100 bins between the interval $[\lambda_{min}, \lambda_{max}]$ or $[0, 2]$ (see Fig. 5.1). Also, one could think about overlapping bins such as $[0, \frac{N_2}{N_1}], [\frac{1}{N_1}, \frac{N_2+1}{N_1}], \ldots, [\frac{2N_1-N_2}{N_1}, \frac{2N_1}{N_1}]$, where the number of bins $(= 1 + 2N_1 - N_2)$ and the width of each bin $(= \frac{N_2}{N_1})$ are controlled by the two parameters N_1 and N_2 (as an example, see Fig. 5.2).

Now, another possibility for looking into the pattern of the spectral density plot is to convolve the Dirac delta function $\sum_k \delta(\lambda, \lambda_k)$ (as a spectral density) with a smooth kernel $g(x, \lambda)$ and plot the density function

$$f(x) = \int g(x, \lambda) \sum_k \delta(\lambda, \lambda_k) \, \mathrm{d}\lambda = \sum_k g(x, \lambda_k).$$

We could choose many different kernels, like the Cauchy–Lorentz distribution $\frac{1}{\pi} \frac{\gamma}{(x-m)^2+\gamma^2}$ or the Gaussian distribution $\frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(x-m_x)^2}{2\sigma^2})$ (see Fig. 5.3, Fig. 5.4,



(a) (b)

Figure 5.1: Rrelative frequency spectrum plots of protein-protein interaction network of *E. coli*. Size: $N = 230$. (a) With 100 bins. (b) With $\sqrt(N) \approx 15$ bins.

Fig. 5.5, and Fig. 5.6.).

Tuning the parameter of the kernel, we might see sharp fractuation of the spectral density. But the problem for many networks is the high spectral density around 1. Thus, the relative height of the peaks differ greatly, and if a plot captures the high peak at 1, other peaks and patterns are often not prominent.

5.2 Visualization of a graph through its spectral plot

What is more, the spectral plot of a graph is much better amenable to visual inspection than a direct plot of the graph or any other method of representation that we know of. In other words, with a little experience in graph theory, one can quickly detect many important features of a graph through a simple look at its spectral plot. We now exhibit some examples.[1]

[1] All networks are taken to be undirected and unweighted. Thus, we suppress some potentially important aspects of the underlying data, but as our plots will show, we can still detect distinctive qualitative patterns. In fact, one can also compute the spectrum of directed and weighted networks, and doing so on our data will reveal further structures, but this is not explored in the present work.

First of all, the properties of the visualization will obviously depend on the display style, and this will be described first (see Fig. 5.7). That figure is based on the metabolic network of *C. elegans*. The first diagram displays the binned eigenvalues, that is, the range $[0, 2]$ is divided here into 35 disjoint bins, and the number of eigenvalues that fall within each such bin is displayed, normalized by the total number of eigenvalues (relative frequency plot). The next figure smoothes this out by using overlapping bins–see the figure legend for parameter values. The subsequent subfigures instead convolve the eigenvalues with a Gaussian kernel, that is, we plot the function

$$f(x) = \sum_{\lambda_j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{|x - \lambda_j|^2}{2\sigma^2})$$

where the λ_j are the eigenvalues. Smaller values of the variance σ^2 emphasize the finer details whereas larger values bring out the global pattern more conspicuously.

We present spectral plots of different artificial networks. First, we start with two different classes of regular networks: 1d regular ring lattice and 2d square grids. Spectral plots of circular regular ring, with different number of connectivity, have been shown in Fig. 5.4. Fig. 5.8 shows the plots of different dimensional square grid. Next we present plots for another 2d squar grid, but with one of the possible two diagonals (always the same) in each square (see, Fig. 5.9). We exhibit spectral plots of networks constructed by some formal schemes that have been suggested to capture important features of biological and other networks, namely an Erdős-Rényi random network, a Watts-Strogatz small-world network and a Barabási-Albert scale-free network (see Fig. 5.6). All graphs are with 1000 nodes. We first have an Erdős-Rényi random graph;[2] here, a single realization and the average of 100 such graphs will not exhibit substantially different spectral plots, that is, each realization already shows the typical spectral properties. This is an indication of the robustness of our scheme against random fluctuations – which, of course, are at the heart of the idea of a random graph. Next, we have a scale-free graph constructed by the algorithm of Barabási-Albert; here, averaging over 100 realizations smoothes the spectral plot out a bit. This is even more evident for a small-world graph á la Strogatz-Watts. We construct them by rewiring a regular graph, either of the square grid or the circle type, both with rewiring probability 0.3 (Fig. 5.10). The spectral plot becomes characteristically different from the regular one.

[2]Because we have normalized our Laplacian, we do not get Wigner's semicircle law for the spectrum of a random graph here.

It is obvious from a quick glance that these spectral plots are very different from those of the metabolic network. This suggests to us that such generic network constructions miss important features and properties of real biological networks. It is also true for networks from other domains. This will now be made more evident by considering further examples of different networks. In Fig. 5.11, we show some more metabolic networks. Fig. 5.12 and Fig. 5.13 display transcription and protein-protein interaction networks. Neurobiological networks and the food-webs network have been shown in Fig. 5.14 and Fig. 5.15 resp. Spectral plots of word-adjacency and Internet graph (see, Fig. 5.16 and Fig. 5.17) are similar with many biological networks. Fig. 5.18, Fig. 5.19 and Fig. 5.20 show network of hyperlinks between weblogs on US politics, protein folding network and e-mail interchange networks resp. Plots of power-grid network (Fig. 5.21), scientific collaboration networks (Fig. 5.22) and electronic circuite networks (Fig. 5.23) are very different from the plots of other networks. Fig. 5.24 and Fig. 5.25 shows the plots of co-purchasing of US political books networks and American football game networks resp. Spectral plots of networks from different domains as weel as from different classes are different.

We shall also see that biological networks from one given class typically have quite similarl-looking spectral plots, which, however, are easily distinguishable from those of networks from a different biological class. Transcription and protein-protein interaction networks (Fig. 5.12 and Fig. 5.13) look somewhat similar to the metabolic networks (Fig. 5.11), and this may reflect a common underlying principle. By way of contrast, the neurobiological networks of Fig. 5.14 and the food-webs in Fig. 5.15 are entirely different – which is not at all surprising, as they come from different biological scales.

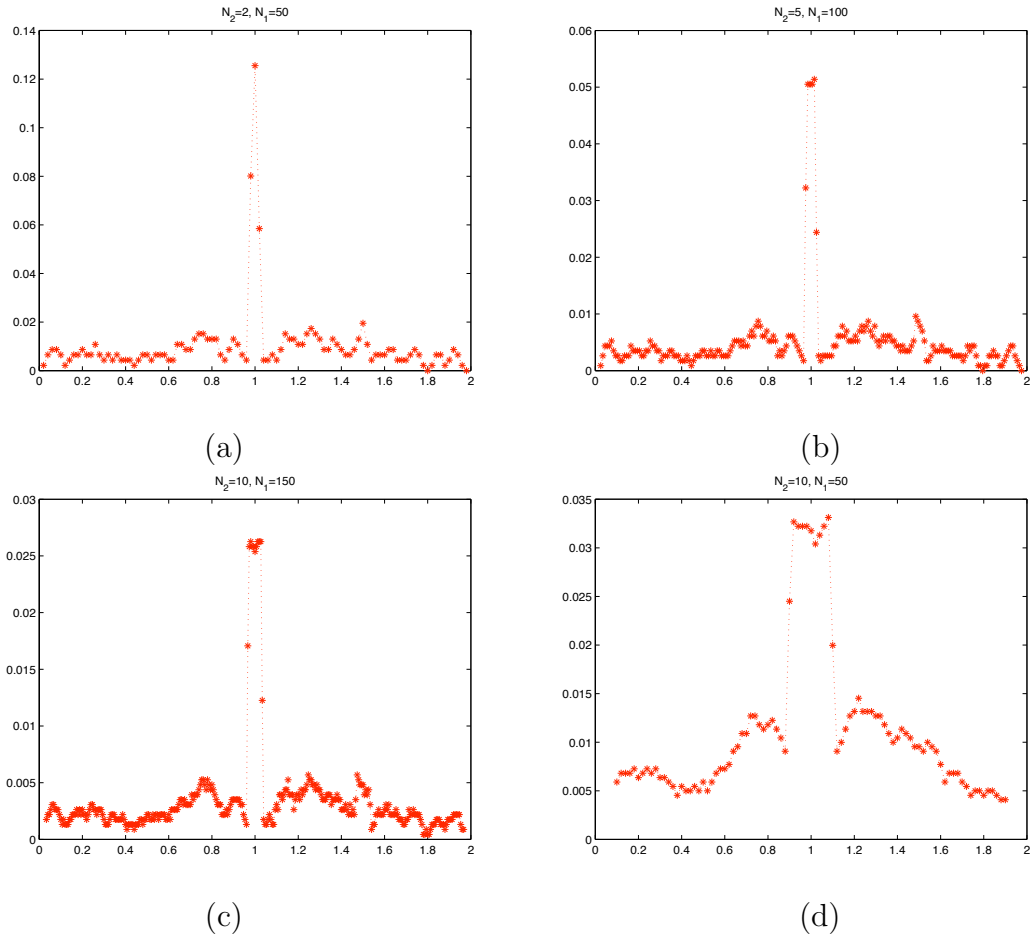


Figure 5.2: Relative frequency spectrum plots of protein-protein interaction network of *E. coli* with overlapping bins. Size of the network is 230. (a) $N_1 = 50$, $N_2 = 2$. (b) $N_1 = 100$, $N_2 = 5$. (c) $N_1 = 150$, $N_2 = 10$. (d) $N_1 = 50$, $N_2 = 10$.

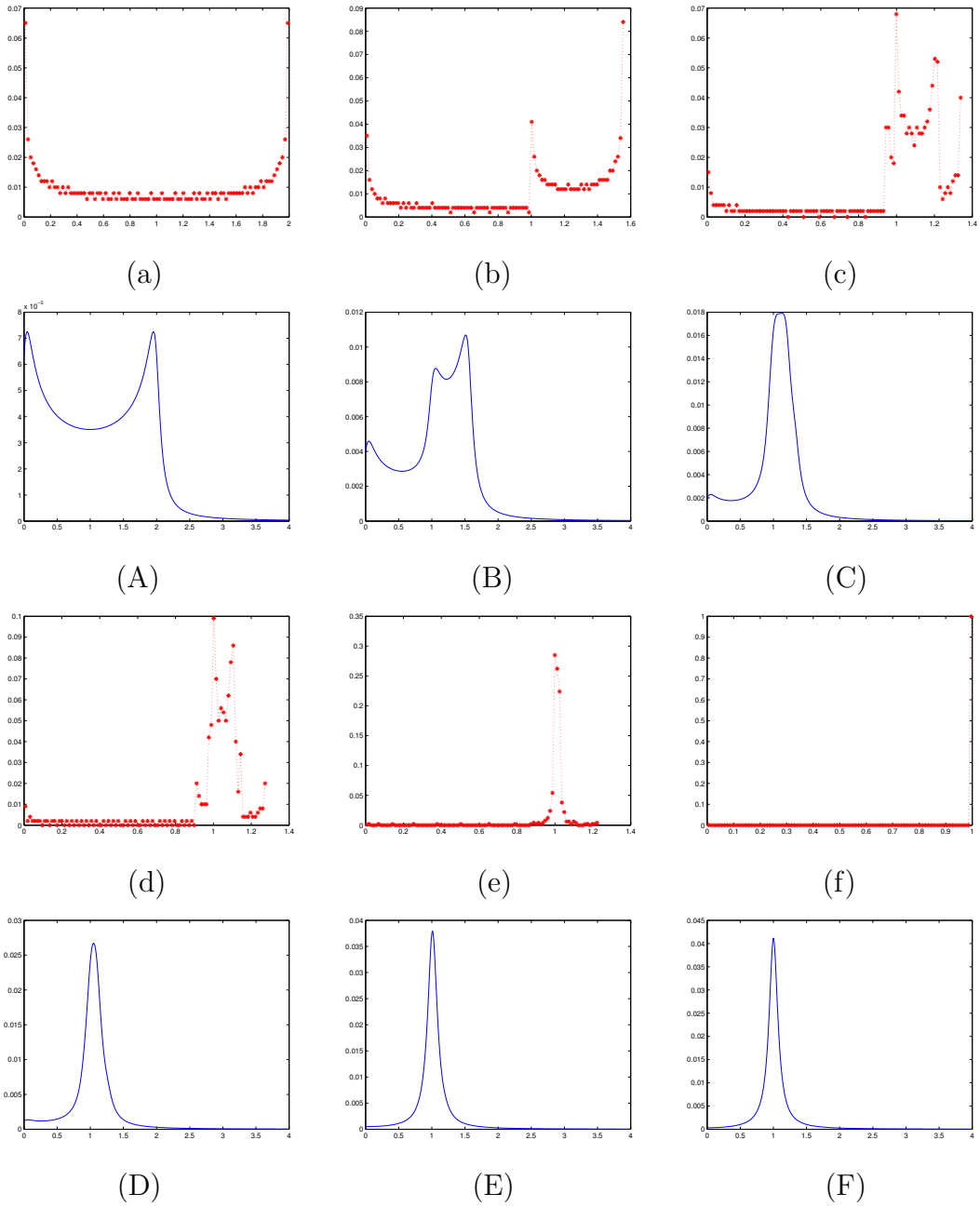


Figure 5.3: Rrelative frequency spectral plots of regular ring lattices with degree of each node (a) 2, (b) 4, (c) 10, (d) 20, (e) 100, (f) Plot of a complete graph. (A), (B), (C), (D), (E), (F) are respective spectral densities as a sum of Lorentz distributions, $\rho(\lambda) = \sum_{k=1}^{N-1} \frac{\gamma}{(\lambda_k - \lambda)^2 + \gamma^2}$, with width $\gamma = .08$. Size of all networks is 1000.

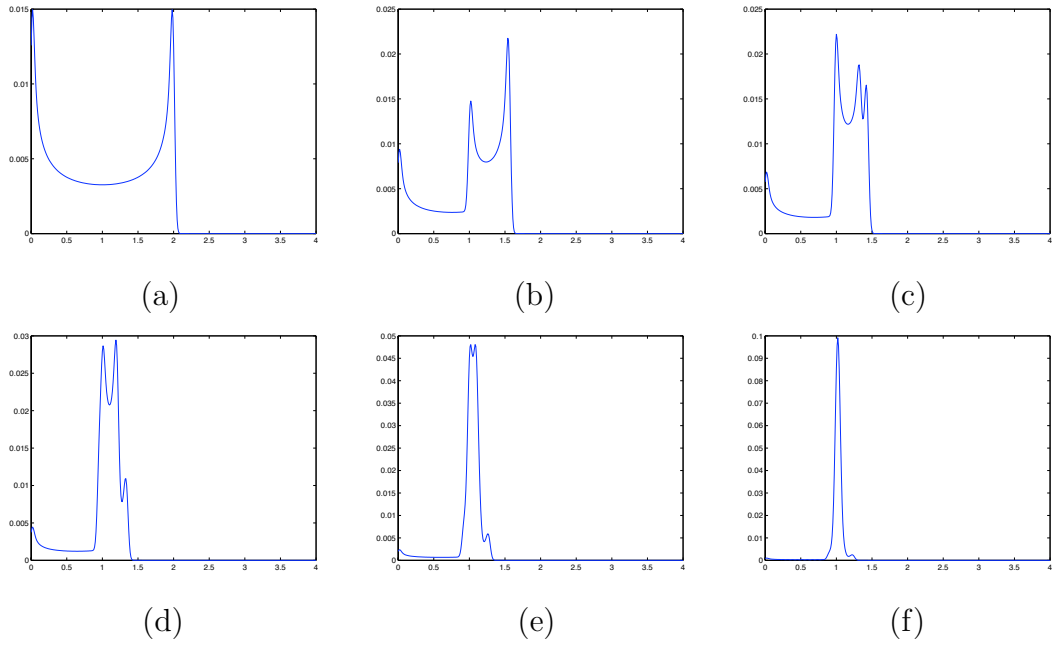


Figure 5.4: 1-dimensional regular ring lattice of size 1000 with degree of each vertex (a) 2 (b) 4 (c) 6 (d) 10 (e) 20 (f) 50. All plots are with Gaussian kernel, $\sigma = 0.025$.

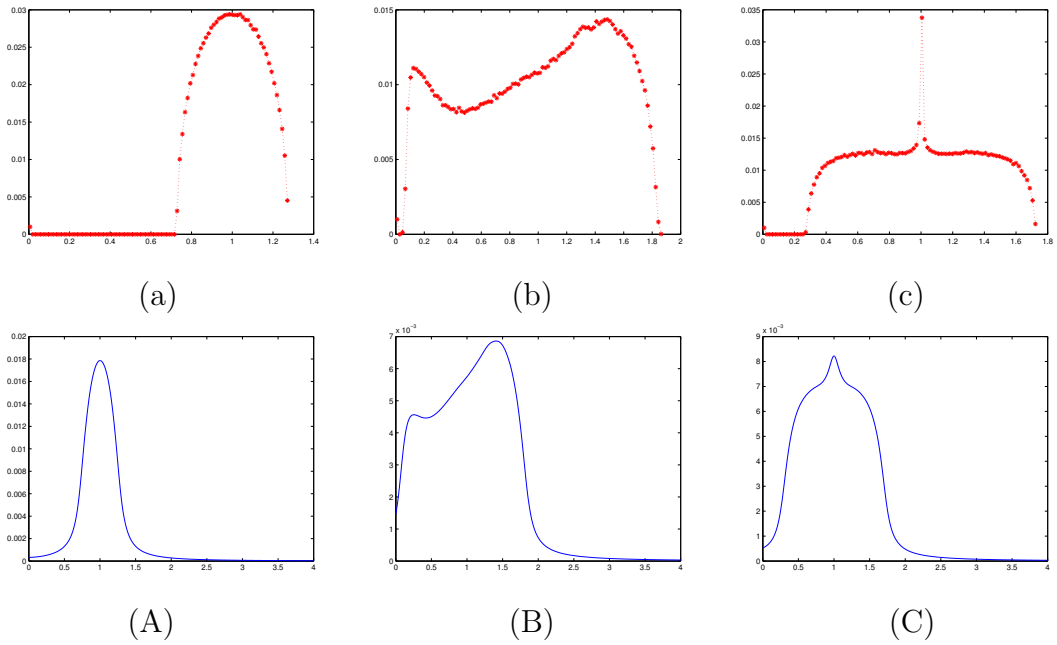


Figure 5.5: Spectral plots of (a) random network from Erdős and Rényi's model (Erdős & Réanyi, 1959) with $p = 0.05$, (b) small-world network from Watts and Stogatz's model (Watts & Strogatz, 1998) (rewiring a regular ring lattice of average degree 4 with rewiring probability 0.3), (c) scale-free network from Albert and Barabási's model (Barabási & Albert, 1999) ($m_0 = 5$ and $m = 3$). (A), (B), (C) are respective spectral densities as a sum of Lorentz distributions, $\rho(\lambda) = \sum_{k=1}^{N-1} \frac{\gamma}{(\lambda_k - \lambda)^2 + \gamma^2}$, with width $\gamma = .08$. Size of all networks is 1000. All figures are ploted with 100 realizations.

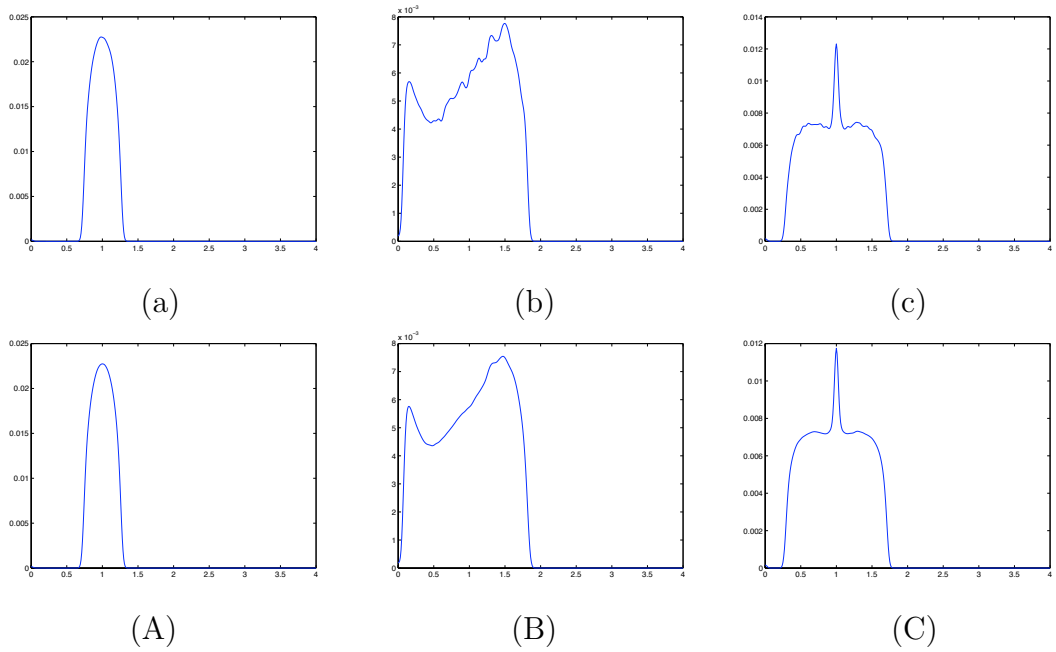


Figure 5.6: Specral plots of generic networks. (a) Random network from the Erdős-Rényi's model (Erdős & Réanyi, 1959) with $p = 0.05$. (b) Small-world network from the Watts-Strogatz's model (Watts & Strogatz, 1998) (rewiring a regular ring lattice of average degree 4 with rewiring probability 0.3). (d) Scale-free network from the Albert-Barabási's model (Barabási & Albert, 1999) ($m_0 = 5$ and $m = 3$). *Figures (a-c) obtained from a single realization, (A-C) represent the averages of* 100 *realizations. Size of all networks is* 1000. . All plots are with Gaussian kernel, $\sigma = 0.025$.

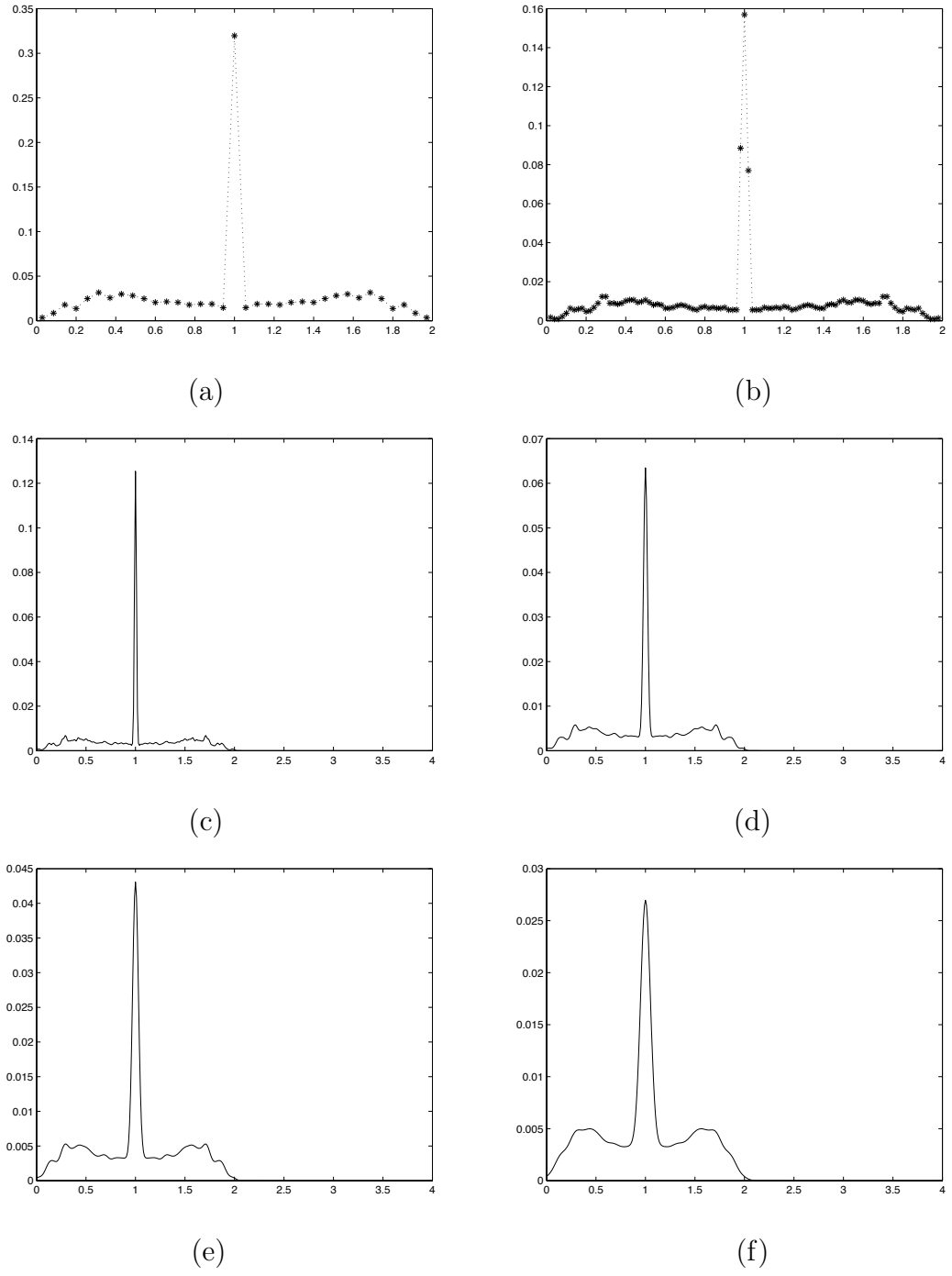


Figure 5.7: Spectral plots of the metabolic network of *Caenorhabditis elegans*. Size of the network is 1173. Nodes are substrates, enzymes and intermediate complexes. Data obtained from (Jeong et al., 2000). Data Source: http://www.nd.edu/∼networks/resources.htm. [Download date: 22 Nov. 2004]. (a) Relative frequency plot with 35 bins. (b) Relative frequency polygon with overlapping bins, bin width 0.04, and 99 bins; bins used are $[0, .04], [.02, .06], [.04, .08], \ldots, [1.96, 2]$. (c) with Gaussian kernel, $\sigma = 0.01$. (d) with Gaussian kernel, $\sigma = 0.02$. (e) with Gaussian kernel, $\sigma = 0.03$. (f) with Gaussian kernel, $\sigma = 0.05$.

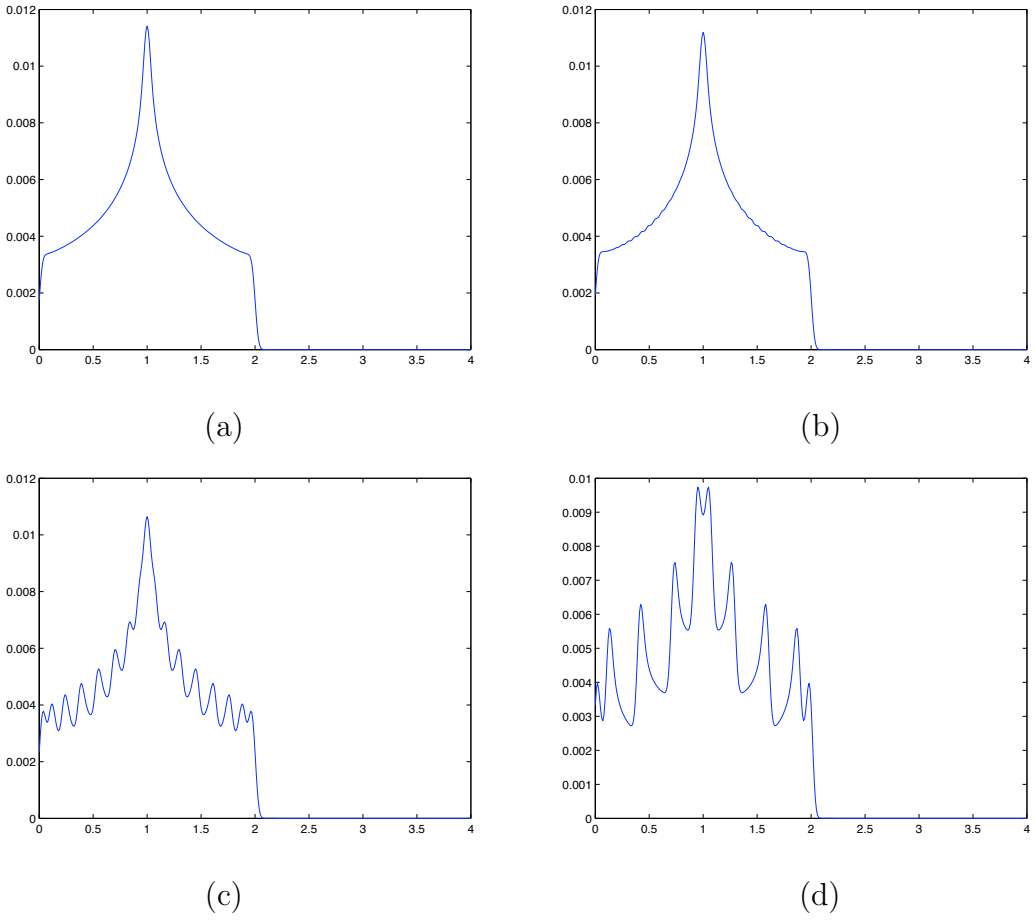


Figure 5.8: 2-dimensional grid with dimension M by N. (a) M= 100, N= 100. (b) M= 25, N= 400. (c) M= 10, N= 1000. (d) M= 5, N= 2000. All plots are with Gaussian kernel with $\sigma = 0.025$

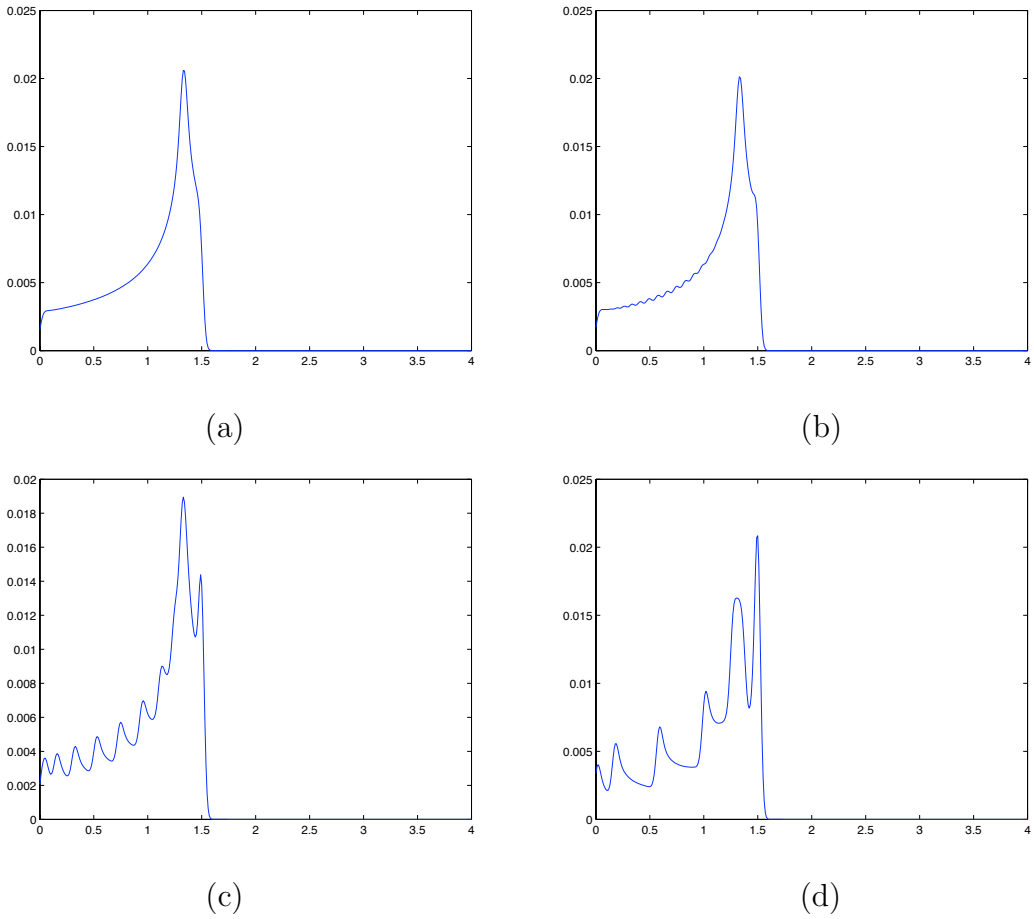


Figure 5.9: 2-dimensional grid with one of the two possible diagonal (always the same) in each square with dimension M by N. (a) M= 100, N= 100. (b) M= 25, N= 400. (c) M= 10, N= 1000. (d) M= 5, N= 2000. All plots are with Gaussian kernel, $\sigma = 0.025$.

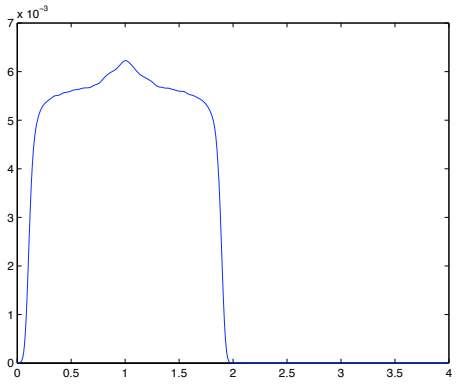


Figure 5.10: Specral plot (with Gaussian kernel, $\sigma = 0.025$) of a small-world network created by rewiring a 2-dimensional grid of dimension 100 by 100 with rewiring probability 0.3. *Plot with single realization.*

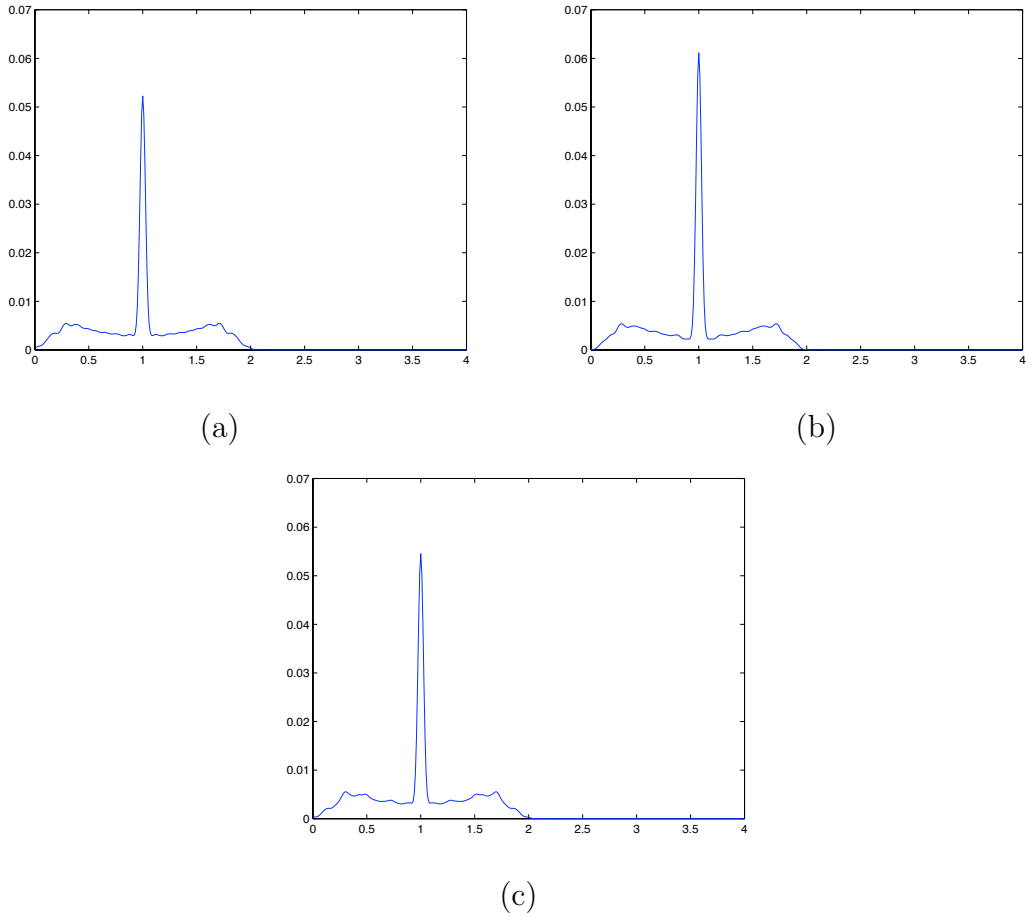
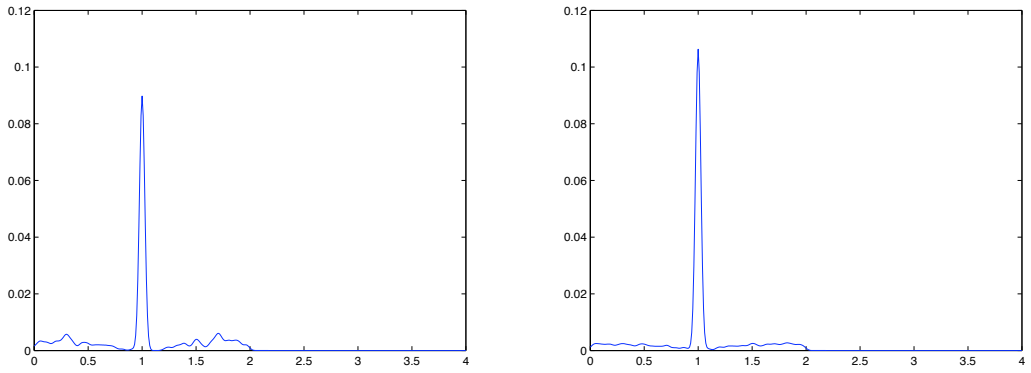


Figure 5.11: Metabolic networks; nodes represent substrates, enzymes and intermediate complexes. Data obtained from (Jeong et al., 2000). Data Source: http://www.nd.edu/∼networks/resources.htm/. [Download date: 22 Nov. 2004] (a) *Archaeoglobus fulgidus*. Network size: 1268. (b) *Escherichia coli*. Network size: 2268. (c) *Saccharomyces cerevisiae*. Network size: 1511. All plots are with Gaussian kernel, $\sigma = 0.025$.



(a) (b)

Figure 5.12: Transcription networks. Data source: Data published by Uri Alon (http://www.weizmann.ac.il/mcb/UriAlon). [Download date: 13 Oct. 2004]. Data used in (Milo et al., 2002; Shen-Orr et al., 2002). (a) *Escherichia coli*. Network size: 328. (b) *Saccharomyces cerevisiae*. Network size: 662. All plots are with Gaussian kernel, $\sigma = 0.025$.

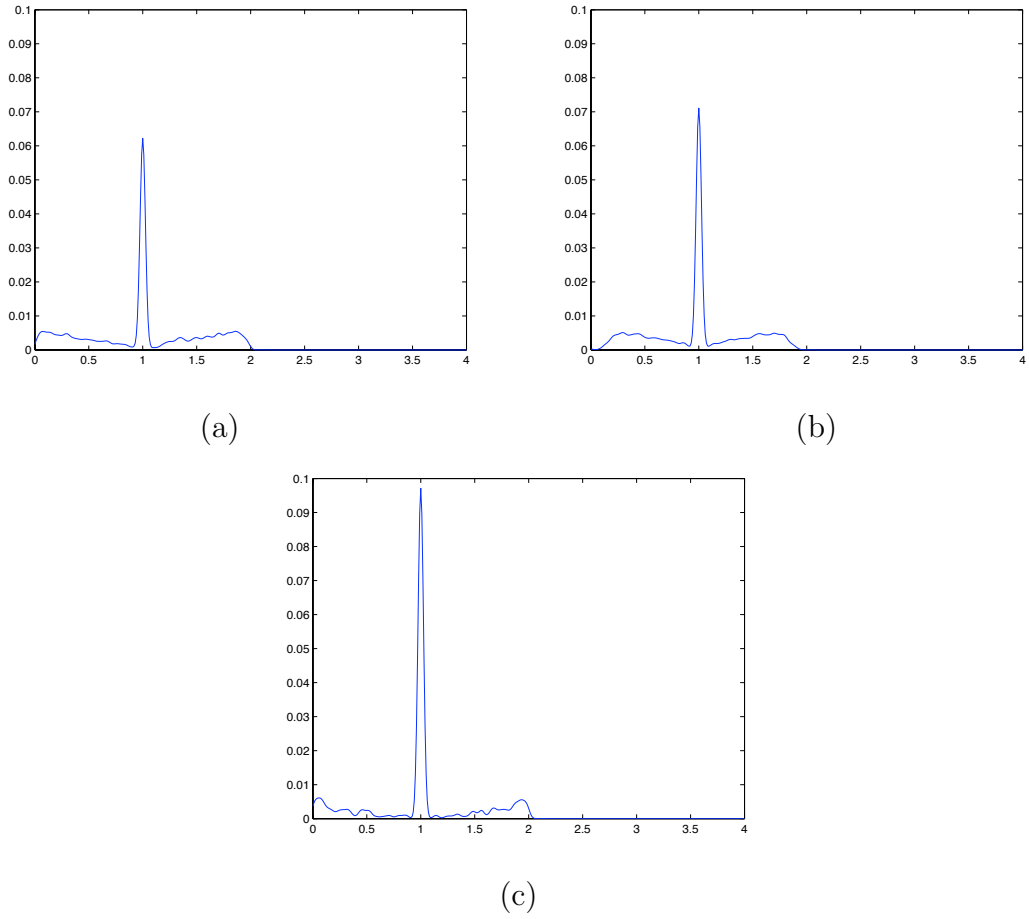
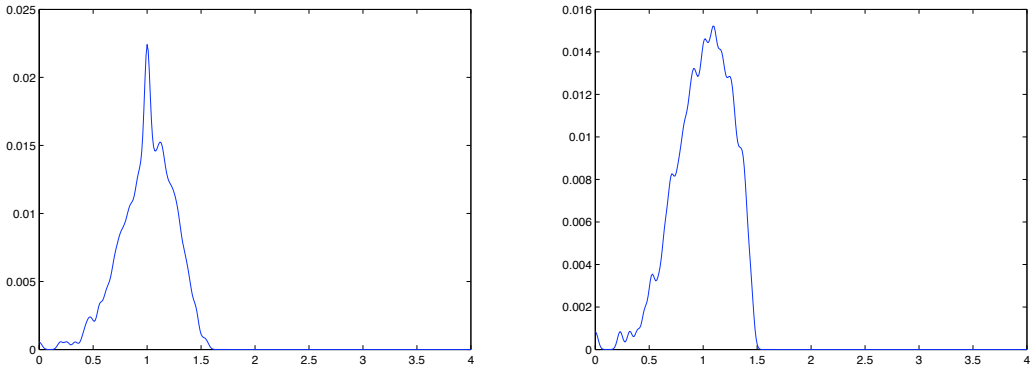
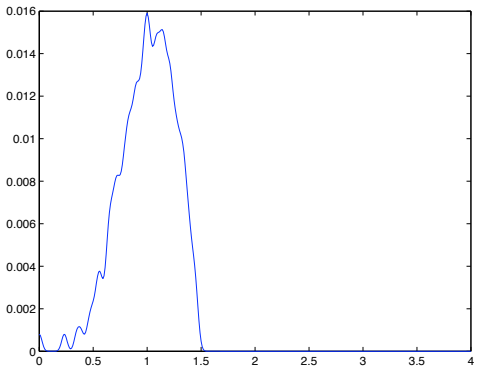


Figure 5.13: Protein-protein interaction networks. (a) *Saccharomyces cerevisiae*. Network size 1458. Data downloaded from http://www.nd.edu/∼networks and data obtained from (Jeong et al., 2001) [download date: 17 September, 2004]. (b) *Helicobacter pylori*. Network size: 710. (c) *Caenorhabditis elegans*.Network size: 314. For (b) and (c), data collected from http://www.cosin.org [download date: 25 September, 2005]. All plots are with Gaussian kernel, $\sigma = 0.025$.

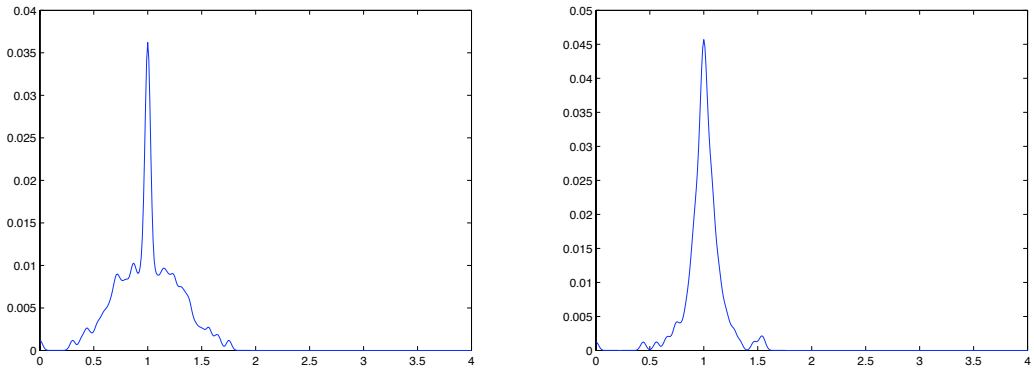


(a) (b)

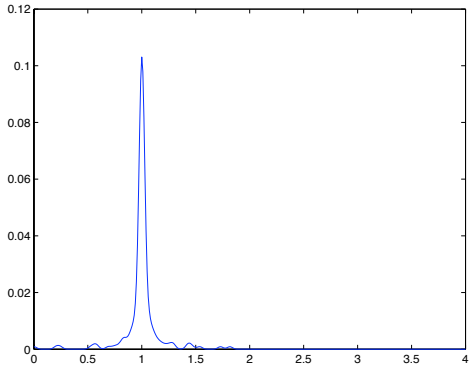


(c)

Figure 5.14: Neuronal connectivity. (a) *Caenorhabditis elegans*. Network size: 297. Data obtained from (Watts & Strogatz, 1998; White et al., 1986). Data Source: http://cdg.columbia.edu/cdg/datasets [Download date: 18 Dec. 2006]. (b) *Caenorhabditis elegans* (animal JSH, L4 male) in the nerve ring and RVG regions. Network size: 190. Data source: Data assembled by J. G. White, E. Southgate, J. N. Thomson, S. Brenner (White et al., 1986) and revisited by R. M. Durbin (Ref. http://elegans.swmed.edu/parts). [Download date: 27 Sep. 2005]. (c) *Caenorhabditis elegans* (animal N2U, adult hermaphrodite) in the nerve ring and RVG regions. Network size: 199. Data source: Data assembled by J. G. White, E. Southgate, J. N. Thomson, S. Brenner (White et al., 1986) and revisited by R. M. Durbin (Ref. http://elegans.swmed.edu/parts). [Download date: 27 Sep. 2005]. All plots are with Gaussian kernel, $\sigma = 0.025$.



(a) (b)



(c)

Figure 5.15: Food-web. (a) From "Ythan estuary". Data downloaded from http://www.cosin.org. [Download Date 21st December, 2006]. Network size: 135. (b) From "Florida bay in wet season". Data downloaded from http://vlado.fmf.uni-lj.si/pub/networks/data (main data resource: Chesapeake Biological Laboratory. Web link: http://www.cbl.umces.edu). [Download Date 21 December, 2006]. Network size: 128. (c) From "Little rock lake". Data downloaded from http://www.cosin.org. [Download Date 21 December, 2006]. Size of the network is 183. All plots are with Gaussian kernel, $\sigma = 0.025$.

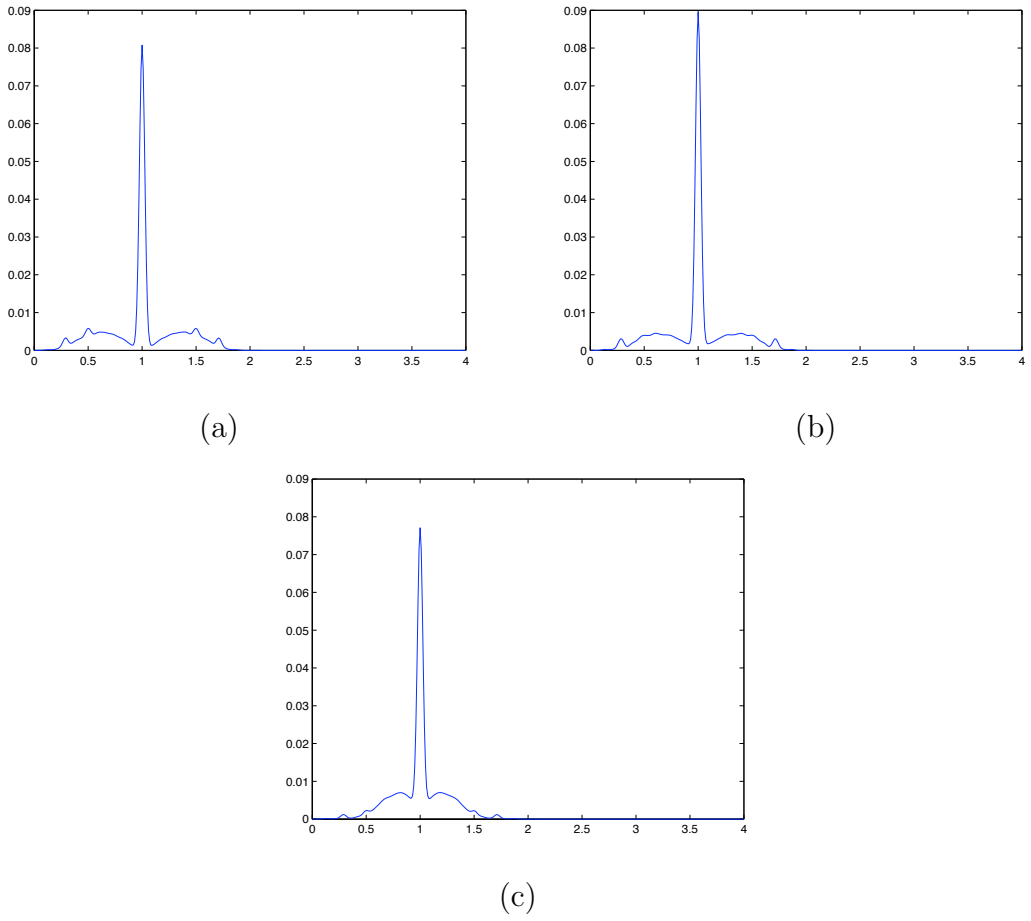
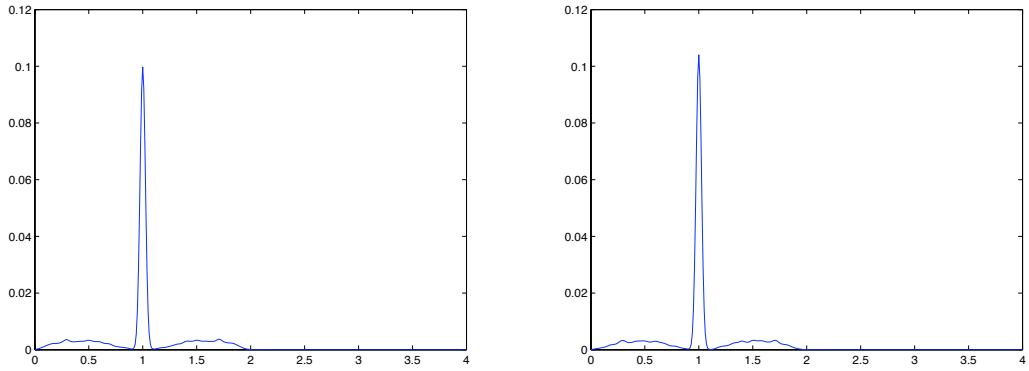
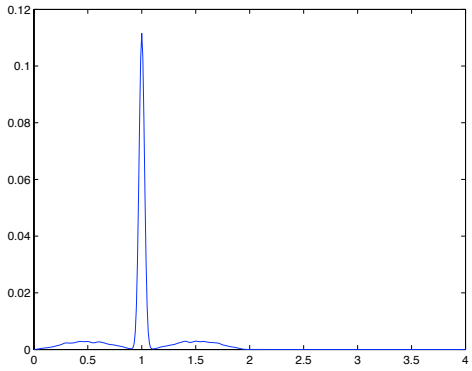


Figure 5.16: Word-adjacency networks of a text in (a) French. Network size: 8308. (b) Japanese. Network size: 2698. (c) English. SNetwork size: 7377. Data downloaded from http://www.weizmann.ac.il/mcb/UriAlon [Download date 3 Feb. 2005]. Data obtained (Milo et al., 2004). All plots are with Gaussian kernel, $\sigma = 0.025$.



(a)

(b)



(c)

Figure 5.17: Autonomous Systems topology of the Internet. Every vertex represents an autonomous system, and two vertices are connected, if there is at least one physical link between the two corresponding Autonomous Systems. (a) AS graph of 8 Nov. 1997. Network size: 3015. (b) AS graph of 2 July 1999. Network size: 5357. (c) AS graph of 16 March 2001. Network size: 10515. Data collected from http://151.100.123.37/extra/data/internet/nlanr.html and data obtained from (Faloutsos et al., 1999) [download date: 23 September, 2005]. Main source: BGP routing data collected by University of Oregon Route Views Project, then processed and made available in various formats at the Global ISP interconnectivity by AS number page of NLANR (National Laboratory of Applied Network Research). All plots are with Gaussian kernel, $\sigma = 0.025$.

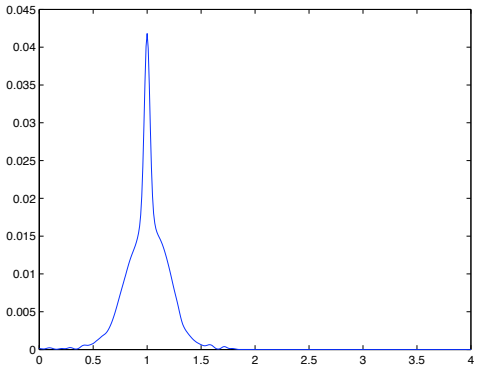


Figure 5.18: Network of hyperlinks between weblogs on US politics, recorded in 2005 by Adamic and Glance (Adamic & Glance, 2005). Network size: 1222. Data downloaded from http://www-personal.umich.edu/∼mejn/netdata [Download date: 23 April 2007]. Plot is with Gaussian kernel, $\sigma = 0.025$.

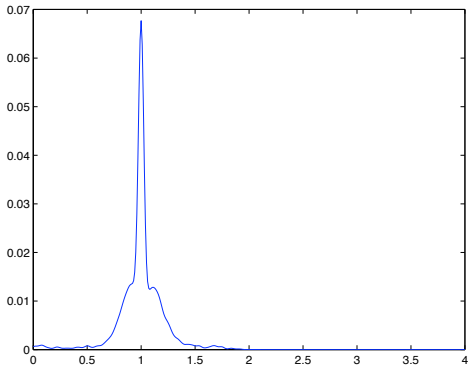


Figure 5.19: Network of conformation space (only conformations that are visited at least 20 times during the simulation are considered in the building of the network) of a 20 residue antiparallel beta-sheet peptide sampled by molecular dynamics simulations (Rao & Caflsich, 2004). Snapshots saved along the trajectory are grouped according to secondary structure into nodes of the network and the transitions between them are links. Network size: 1199. Downloaded from Caflisch group, University of Zurich, http://www.biochem-caflisch.unizh.ch [Download date: 18th Dec. 2006]. Plot is with Gaussian kernel, $\sigma = 0.025$.

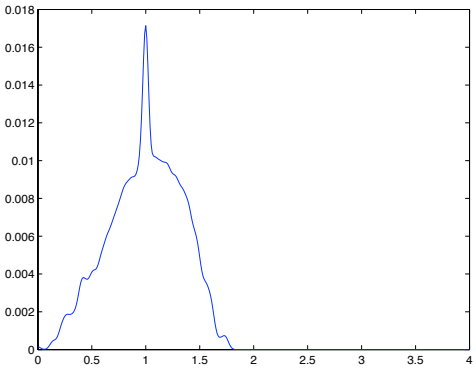


Figure 5.20: E-mail interchanges between members of the Univeristy Rovira i Virgili (Tarragona) (Guimera et al., 2003). Network size: 1133. Data downloaded from http://www.etse.urv.es/∼aarenas/data/welcome.htm [Download date: 21 March, 2007]. Plot is with Gaussian kernel, $\sigma = 0.025$.

5.3 Conclusion

Here we have presented a simple technique for visualizing the important qualitative aspects of biological and other networks and for distinguishing networks of different origins. This technique can be used for qualitative classification of different networks. Now, we will explore how and why inheritance structure of a network is reflected by its spectrum and so, reflected by the spectral plot.

Remark. Plots are usefull for discerning patterns. Different spectral plots from various networks have different patterns. Since the spectral plots of many biological networks have a high peak at 1 (because of the high multiplicity of the eigenvalue 1), one must always tune the parameter of the kernel to get the unique pattern from the plot. One must, however, remember that the plot does not give precise information or even an estimate about the multiplicity of an eigenvalue. E.g., in Fig. 5.7, the eigenvalue 0.3 seems to have a higher multiplicity than the eigenvalue 0.5. However, the precise data shows that the multiplicity of eigenvalue 0.5 is 3 and there is no eigenvalue 0.3, but there are many eigenvalues close to 0.3. So for the multiplicity, it is better to directly check the data on the eigenvalues.

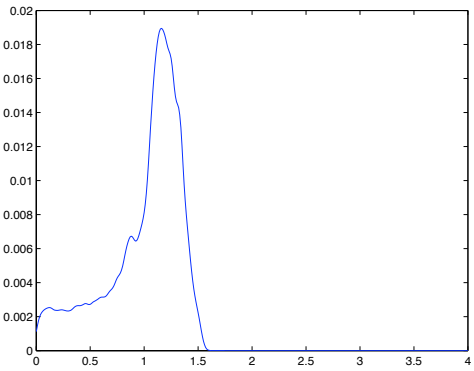
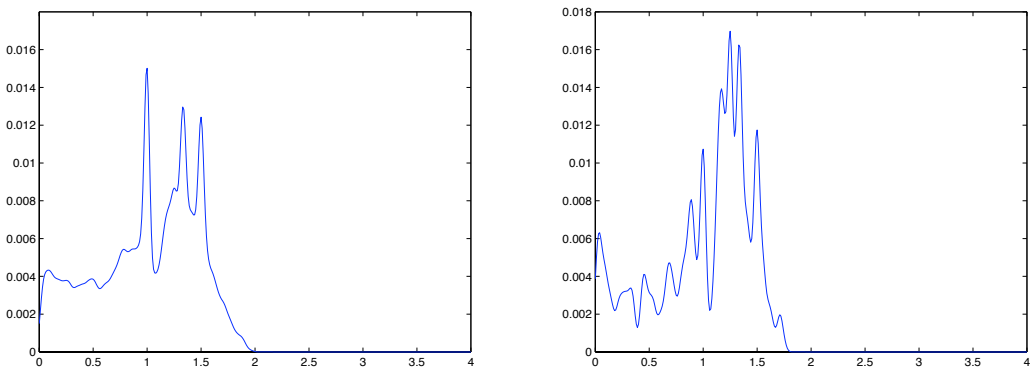
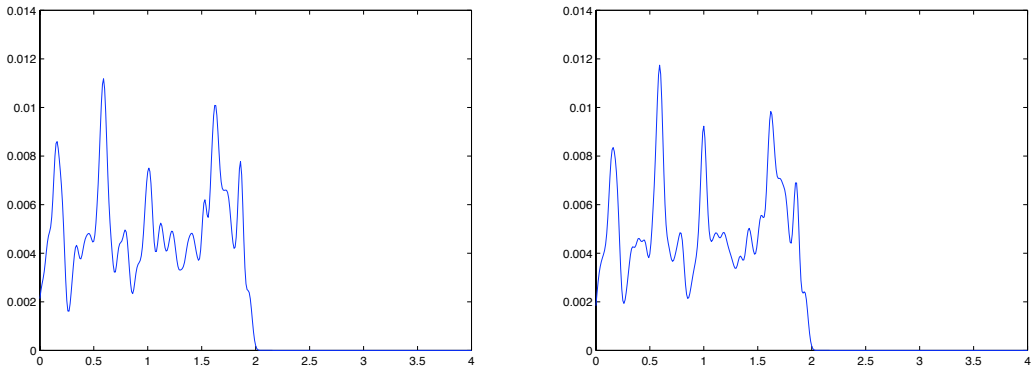


Figure 5.21: Topology of the Western States Power Grid of the United States (Watts & Strogatz, 1998). Network size: 4941. Data downloaded from http://cdg.columbia.edu/uploads/datasets [Download date: 1 March, 2007.]. Plot is with Gaussian kernel, $\sigma = 0.025$.



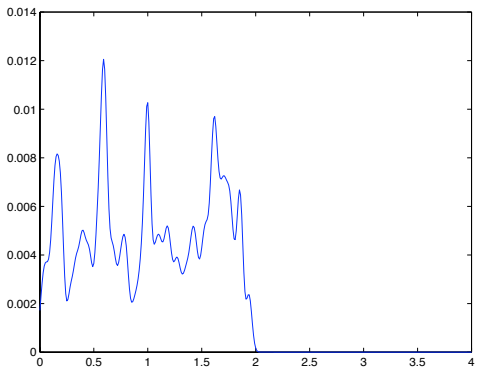
(a) (b)

Figure 5.22: Network of (a) coauthorships between scientists posting preprints on the High-Energy Theory E-Print Archive, http://arxiv.org/archive/hep-th between 1 Jan, 1995 and 31 December 1999 (Newman, 2001c). Network size: 5835. (b) coauthorships of scientists working on network theory and experiment (Newman, 2001b). Network size: 379. Data downloaded from http://www-personal.umich.edu/∼mejn/netdata [Download date: 23 April, 2007]. All plots are with Gaussian kernel, $\sigma = 0.025$.



(a)

(b)



(c)

Figure 5.23: Electronic circuits. (a) With size 122. (b) With size 252. (c) With size 512. Data downloaded from http://www.weizmann.ac.il/mcb/UriAlon [Download date: 15 March, 2005]. Data obtained from (Milo et al., 2002). All plots are with Gaussian kernel, $\sigma = 0.025$.

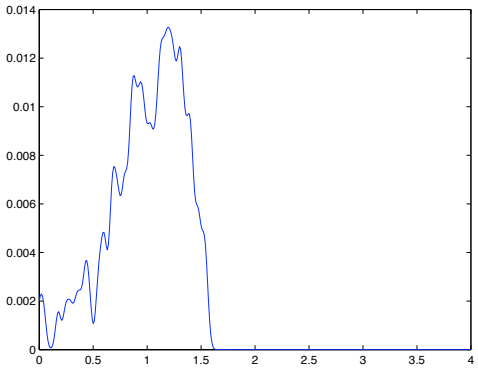


Figure 5.24: Network of copurchasing of books about recent US politics sold by the online bookseller Amazon.com. Edges between books represent frequent copurchasing of books by the same buyers. Network compiled by V. Krebs (unpublished). Network size: 105. Data downloaded from http://www-personal.umich.edu/∼mejn/netdata [original source http://www.orgnet.com. Download date: 23 April, 2007]. Plot is with Gaussian kernel, $\sigma = 0.025$.

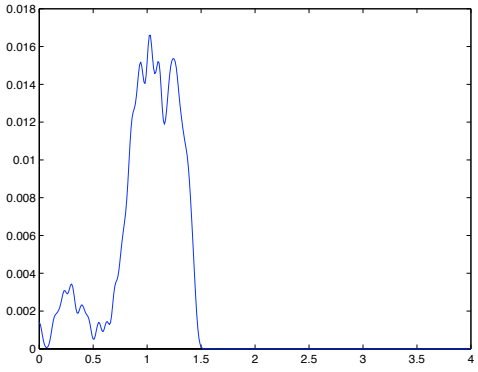


Figure 5.25: Networks of American football games between division IA colleges during regular season fall 2000, as compiled by M. Girvan and M. Newman (Girvan & Newman, 2002). Network size: 115. Data downloaded from http://www-personal.umich.edu/∼mejn/netdata [Download date: 23 April, 2007]. Plot is with Gaussian kernel, $\sigma = 0.025$.

Chapter 6

Eigenfunctions and Graph Structure

We are interested in the spectrum of the normalized graph Laplacian as yielding important invariants of the underlying graph Γ and incorporating its qualitative properties. As in the case of the algebraic Laplacian, one can essentially recover the graph from its spectrum, up to isospectral graphs. The latter are known to exist, but are relatively rare and qualitatively quite uniform in most respects (see e.g. (Zhu & Wilson, 2005) for a systematic discussion). For a heuristic algorithm for the algebraic Laplacian, which can be easily modified for the normalized Laplacian, see (Ipsen & Mikhailov, 2002).

Now we shall see how eigenfunctions vary according to the structure and substructure, how this leads to specific eigenvalues (see (Banerjee & J.Jost, c)). On the spectral plots of real networks, we observerd that a peak at 1 is very common. Also, in some cases it is very sharp. So here we shall give special attention to the graph evolutionary processes that produce high multiplicity of the eigenvalue 1 or eigenvalues that are very close to 1. Other frequently observed eigenvalues are 1/2 and 3/2. Here we explore some situations that create these eigenvalues.

Let us think of a graph Γ representing real data as a structure that has evolved from some simpler precursors, for example by joining smaller graphs into a larger one, or by duplicating certain sets of vertices in a precursor graph. It is important to find some indications of this process in the spectrum of Γ.

Two notions are important for understanding how the structure or substructures of a graph influence the occurrence of an eigenfunction that produces a

specific eigenvalue.

In some cases, a solution u_k of the eigenvalue equation

$$\Delta u_k - \lambda_k u_k = 0$$

can be *localized*, that is, be 0 outside a small set of vertices. In other cases, it has to be *global*, that is, be 0 only at relatively few vertices. These notions provide some insight into the behavior of graphs under certain operations as we shall now explore.

We recall some elementary properties of the eigenvalues of the normalized Laplacian operator given in the previous section (see also (Chung, 1997; Jost & Joy, 2002)).
The eigenvalue equation of the normalized Laplacian, (henceforth simply called the Laplacian)

$$\Delta u - \lambda u = 0.$$

becomes

$$\frac{1}{n_i} \sum_{j \sim i} u(j) = (1 - \lambda) u(i) \text{ for all } i. \tag{6.1}$$

In particular, when the eigenfunction u vanishes at i, then also

$$\sum_{j \sim i} u(j) = 0 \tag{6.2}$$

and conversely (except when $\lambda = 1$). This observation will be useful for us below.

6.1 The eigenvalue 1

For the eigenvalue $\lambda = 1$, (6.1) becomes simply

$$\sum_{j \sim i} u(j) = 0 \text{ for all } i, \tag{6.3}$$

that is, the average of the neighboring values vanishes for each i. We call a solution u of (6.3) *balanced*. The multiplicity m_1 of the eigenvalue 1 then equals the number of linearly independent balanced functions on Γ.

There is an equivalent algebraic formulation: Let $A = (a_{ij})$ be the adjacency matrix of Γ. Then (6.3) simply means

$$Au = \sum_j a_{ij} u(j) = 0, \tag{6.4}$$

that is, the vector $u(j)_{j\in\Gamma}$ lies in the kernel of the adjacency matrix. Thus,

$$m_1 = \dim \ker A. \tag{6.5}$$

We are interested in the question of estimating the multiplicity of the eigenvalue 1 of a graph. An obvious method for this is to determine restrictions on the corresponding eigenfunctions f_1. We shall do this by graph theoretical considerations, and in this sense, this constitutes a geometric approach to the algebraic question of determining or estimating the kernel of a symmetric 0-1 matrix with vanishing diagonal. In (Bevis, Blount, Davis, Domke, & Miller, 1997) the effect of the addition of a single vertex to m_1 was systematically investigated. Here, we are also interested in the effect of more global graph operations. We start with the following simple observation

Lemma 6.1.1. *Let q be a vertex of degree 1 in Γ (such a q is called a* pending *vertex). Then any eigenfunction f_1 for the eigenvalue 1 vanishes at the unique neighbor of q.*

Definition 6.1.1. A *motif* Σ is a connected small subgraph of Γ (whereas the graph Γ is supposed to be large), containing all edges of Γ between vertices of Σ.

6.2 Motif doubling, graph splitting and joining

6.2.1 Motif doubling

Let Σ be a motif of a graph Γ with vertices $p_1, \ldots, p_m$. The situation we have in mind is where N, the number of vertices of Γ, is large, while m, the number of vertices of Σ, is small. Let 1 be an eigenvalue of Σ with eigenfunction f_1^Σ. Now, if we extend f_1^Σ, to all of Γ by setting it equal to 0 outside Σ, (i. e. we construct a new function f such that $f(p_\alpha) = f_1^\Sigma(p_\alpha)$ for $\alpha = 1, \ldots, m$ and $f = 0$ for all other vertices of Γ) then f need not be an eigenfunction of Γ, and 1 need not even be an eigenvalue of Γ. But we can, however, enlarge Γ by doubling the motif Σ so that the enlarged graph also possesses the eigenvalue 1, with a localized eigenfunction:

Theorem 6.2.1. *Let Γ^Σ be obtained from Γ by adding a copy of the motif Σ consisting of the vertices $q_1, \ldots, q_m$ and the corresponding connections between them, and connecting each q_α with all $p \notin \Sigma$ that are neighbors of p_α. Then Γ^Σ possesses the eigenvalue 1, with a localized eigenfunction that is nonzero only at the p_α and the q_α.*

Proof. We define a function

$$f_1^{\Gamma^\Sigma}(p) = \begin{cases} f_1^\Sigma(p_\alpha) \text{ if } p = p_\alpha \in \Sigma \\ -f_1^\Sigma(p_\alpha) \text{ if } p = q_\alpha \\ 0 \text{ else.} \end{cases} \tag{6.6}$$

Now, our claim is that $f_1^{\Gamma^\Sigma}$ is an eigenfunction of the graph Γ^Σ corresponding to the eigenvalue 1.

Let Σ' be the copy of Σ and $p_\beta \notin \Sigma$ be a neighbor of p_α. So p_β will also be a neighbor of q_α. By the construction (6.6) of the the function $f_1^{\Gamma^\Sigma}$, $f_1^{\Gamma^\Sigma}(p_\beta) = 0$. Now,

$$\begin{aligned} \sum_{p\sim p_\alpha} f_1^{\Gamma^\Sigma}(p) &= \sum_{p_\beta \sim p_\alpha} f_1^{\Gamma^\Sigma}(p_\beta) + \sum_{p\in\Sigma; p\sim p_\alpha} f_1^{\Gamma^\Sigma}(p) \\ &= 0 + \sum_{p\in\Sigma; p\sim p_\alpha} f_1^\Sigma(p), \text{ by (6.6)} \\ &= 0. \end{aligned} \tag{6.7}$$

since, being an eigenfunction corresponding to the eigenvalue 1 of Σ, f_1^Σ stisfies the equation (6.3) on Σ, i. e. , f_1^Σ is balanced on Σ.

Equation (6.7) is true for all p_α, and similarly for all q_α. Now, for all p_β

$$\begin{aligned} \sum_{p\sim p_\beta} f_1^{\Gamma^\Sigma}(p) &= \sum_{p\notin\Sigma; p\notin\Sigma'; p\sim p_\beta} f_1^{\Gamma^\Sigma}(p) + \sum_{p_\alpha\in\Sigma; p_\alpha\sim p_\beta} f_1^{\Gamma^\Sigma}(p_\alpha) + \sum_{q_\alpha\in\Sigma'; q_\alpha\sim p_\beta} f_1^{\Gamma^\Sigma}(q_\alpha) \\ &= 0 + \sum_{p_\alpha\in\Sigma; p_\alpha\sim p_\beta} f_1^\Sigma(p_\alpha) + \sum_{q_\alpha\in\Sigma'; q_\alpha\sim p_\beta} f_1^\Sigma(q_\alpha) \\ &= 0 + \sum_{p_\alpha\in\Sigma; p_\alpha\sim p_\beta} f_1^\Sigma(p_\alpha) + \sum_{p_\alpha\in\Sigma; q_\alpha\sim p_\beta} -f_1^\Sigma(p_\alpha), \text{ using (6.6)} \\ &= 0. \end{aligned} \tag{6.8}$$

For any other vertices $p \in \Gamma$, the function $f_1^{\Gamma^\Sigma}(p)$ will satisfy equation (6.3), since $f_1^{\Gamma^\Sigma}(p) = 0$ for all $p \notin \Sigma \cup \Sigma'$. Thus $f_1^{\Gamma^\Sigma}$ is balanced in Γ^Σ and is an eigenfunction of Γ^Σ. □

The theorem also holds for the case where Σ is a single vertex p_1 (even though such a motif does not possess the eigenvalue 1 itself). Thus, we can always produce the eigenvalue by **vertex doubling**. This is a reformulation of a result of (Ellingham, 1993).

Corollary 6.2.1. *Let Γ^Σ be obtained from Γ by adding Σ', a copy of the motif Σ consisting of the vertices $q_1, \ldots, q_m$ and the corresponding connections between them, and connecting each q_α with all p that are neighbors of p_α. Then Γ^Σ possesses m more eigenvalues 1 than Γ, with localized eigenfunctions f_1^α $(\alpha = 1, \ldots, m)$ that are 1 at p_α, -1 at q_α and zero elsewhere.*

Proof. We can obtain Γ^Σ from Γ by m sequential duplications of vertices p_α $(\alpha = 1, \ldots, m)$. So there will be an increment of the multiplicity of the eigenvalue 1 by m in Γ^Σ. □

Thus, if we wish to produce a high multiplicity for the eigenvalue 1, we can perform many vertex doublings. We could either duplicate different vertices, or we could duplicate one vertex repeatedly. In fact, the repeated doubling of one vertex leaves a characteristic trace in the number of certain small motifs in the graph. Let p_1 be a vertex and q_1 its double. We consider any motif Σ consisting of a certain collection $p, p', p'', \ldots$ of neighbors of p_1 together with their connections to both p_1 and q_1 and possibly some connections among them.

Theorem 6.2.2. *Let the graph $\bar{\Gamma}$ be obtained from Γ by n successive doublings of the vertex p_1, and let Σ be any motif of the type just described. Then $\bar{\Gamma}$ contains at least $\binom{n}{2}$ instances of the motif Σ.*

Proof. An instance of the motif Σ is obtained by taking any two copies of p_1 and the vertices $p, p', p'', \ldots$ together with the connections defining Σ. There exist $\binom{n}{2}$ such pairs of copies of p_1 in $\bar{\Gamma}$. □

Theorem 6.2.1, however, does not apply to eigenvalues other than 1 because for $\lambda \neq 1$, the vertex degrees n_i in (6.1) are important, and this is affected by embedding the motif Σ into another graph Γ. However, we have the following variant in the general case.

Theorem 6.2.3. *Let Σ be a motif in Γ. Suppose f satisfies*

$$\frac{1}{n_i} \sum_{j \in \Sigma, j \sim i} f(j) = (1 - \lambda) f(i) \text{ for all } i \in \Sigma \text{ and some } \lambda. \tag{6.9}$$

Then the motif doubling of Theorem 6.2.1 produces a graph Γ^Σ with eigenvalue λ and an eigenfunction f^{Γ^Σ} agreeing with f on Σ, with $-f$ on the double of Σ, and which is identically 0 on the rest of Γ^Σ.

Proof. Let Σ' be the copy of Σ; let $p_1, \ldots, p_m$ be the vertices of Σ and $q_\alpha \in \Sigma'$ be the corresponding copy of $p_\alpha \in \Sigma$.

Let $p_\beta (\notin \Sigma, \Sigma')$ be the neighbors of p_α and q_α.
According to the construction of f^{Γ^Σ},

$$f^{\Gamma^\Sigma}(p) = \begin{cases} f(p_\alpha) \text{ if } p = p_\alpha \in \Sigma \\ -f(p_\alpha) \text{ if } p = q_\alpha \in \Sigma' \\ 0 \text{ else.} \end{cases} \tag{6.10}$$

So for any $p_\alpha \in \Sigma$

$$\begin{aligned} \frac{1}{n_{p_\alpha}} \sum_{s \sim p_\alpha} f^{\Gamma^\Sigma}(s) &= \frac{1}{n_{p_\alpha}} \sum_{p_\beta \sim p_\alpha} f^{\Gamma^\Sigma}(p_\beta) + \frac{1}{n_{p_\alpha}} \sum_{p_\gamma \in \Sigma; p_\gamma \sim p_\alpha} f^{\Gamma^\Sigma}(p_\gamma) \\ &= 0 + \frac{1}{n_{p_\alpha}} \sum_{p_\gamma \in \Sigma; p_\gamma \sim p_\alpha} f^{\Gamma^\Sigma}(p_\gamma), \text{ by (6.10)} \\ &= (1-\lambda) f(p_\alpha), \text{ using (6.9) and (6.10)} \\ &= (1-\lambda) f^{\Gamma^\Sigma}(p_\alpha), \text{ by (6.10).} \end{aligned} \tag{6.11}$$

Similarly, for any $q_\alpha \in \Sigma'$

$$\frac{1}{n_{q_\alpha}} \sum_{s \sim q_\alpha} f^{\Gamma^\Sigma}(s) = (1-\lambda) f^{\Gamma^\Sigma}(q_\alpha). \tag{6.12}$$

Now, according to equation (6.10), $f^{\Gamma^\Sigma}(p_\beta) = 0$ and

$$\begin{aligned} \frac{1}{n_{p_\beta}} \sum_{s \sim p_\beta} f^{\Gamma^\Sigma}(s) &= \frac{1}{n_{p_\beta}} \sum_{p \notin \Sigma, \Sigma'; p \sim p_\beta} f^{\Gamma^\Sigma}(p) + \frac{1}{n_{p_\beta}} \sum_{p_\alpha \in \Sigma; p_\alpha \sim p_\beta} f^{\Gamma^\Sigma}(p_\alpha) + \frac{1}{n_{p_\beta}} \sum_{q_\alpha \in \Sigma'; q_\alpha \sim p_\beta} f^{\Gamma^\Sigma}(q_\alpha) \\ &= 0 + \frac{1}{n_{p_\beta}} \sum_{p_\alpha \in \Sigma; p_\alpha \sim p_\beta} f(p_\alpha) + \frac{1}{n_{p_\beta}} \sum_{q_\alpha \in \Sigma'; q_\alpha \sim p_\beta} -f(p_\alpha) \text{ by (6.10)} \\ &= 0. \end{aligned} \tag{6.13}$$

So

$$\frac{1}{n_{p_\beta}} \sum_{s \sim p_\beta} f^{\Gamma^\Sigma}(s) = (1-\lambda) f^{\Gamma^\Sigma}(p_\beta). \tag{6.14}$$

Any other vertices $s \notin \Sigma, \Sigma'$ only have neighbors that do not lies in Σ or Σ', so for those vertices, $f^{\Gamma^\Sigma}(s)$ will be 0 and f^{Γ^Σ} will satisfy the eigenvalue equation (6.1). Thus f^{Γ^Σ} will be an eigenfunction correspoding to the eigenvalue λ of the graph Γ^Σ. □

So here we can construct a localized eigenfunction $f_\lambda^{\Gamma^\Sigma}$ for the eigenvalue λ after motif duplication. The simplest motif is an edge connecting two vertices p_1, p_2. The corresponding relations (6.9) are then

$$\frac{1}{n_{p_1}} f(p_2) = (1-\lambda) f(p_1), \quad \frac{1}{n_{p_2}} f(p_1) = (1-\lambda) f(p_2) \tag{6.15}$$

which admit the solutions

$$\lambda = 1 \pm \frac{1}{\sqrt{n_{p_1} n_{p_2}}}. \tag{6.16}$$

Thus, edge doubling leads to eigenvalues which approach 1 as the degree of p_1 or p_2 gets larger. In any case, the two values are symmetric about 1.

We can also double the entire graph:

Theorem 6.2.4. *Let Γ_1 and Γ_2 be isomorphic graphs with vertices $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$, respectively, where p_i corresponds to q_i for $i = 1, \ldots, n$. We then construct a graph Γ_0 by connecting p_i with q_j whenever $p_j \sim p_i$. If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of Γ_1 and Γ_2, then Γ_0 has the same eigenvalues, as well as the eigenvalue 1 with multiplicity n.*

Proof. The degree of every vertex p in Γ_0 is $2n_p$, where n_p is its original degree in Γ_1. Let f_λ be an eigenfunction of Γ_1 (which is therefore also an eigenfunction of Γ_2). So f_λ will satisfy the eigenvalue equation (6.1) on Γ_1 (and on Γ_2). Hence

$$\frac{1}{n_p} \sum_{s_1 \in \Gamma_1; s_1 \sim p} f_\lambda(s_1) = (1-\lambda) f_\lambda(p), \text{ for all } p \in \Gamma_1. \tag{6.17}$$

According to the construction of Γ_0,

$$f_\lambda(p_\alpha) = f_\lambda(q_\alpha) \text{ for all } p_\alpha \in \Gamma_1, q_\alpha \in \Gamma_2. \tag{6.18}$$

Now our claim is that f_λ is an eigenfunction of Γ_0 for the eigenvalue λ.

Now, for any vertex $p \in \Gamma_0$,

$$\begin{aligned}\frac{1}{2n_p}\sum_{s\in\Gamma_0;s\sim p} f_\lambda(s) &= \frac{1}{2}\left[\frac{1}{n_p}\sum_{s_1\in\Gamma_1;s_1\sim p} f_\lambda(s_1) + \frac{1}{n_p}\sum_{s_2\in\Gamma_2;s_2\sim p} f_\lambda(s_2)\right] \\ &= \frac{1}{2}\left[2\frac{1}{n_p}\sum_{s_1\in\Gamma_1;s_1\sim p} f_\lambda(s_1)\right], \text{ using (6.18)} \\ &= (1-\lambda)f_\lambda(p), \text{ by (6.17)}.\end{aligned} \tag{6.19}$$

Thus, by (6.1), it is an eigenfunction of Γ_0.
Finally, similarly to the proof of Theorem 6.2.1, we obtain the eigenvalue 1 with multiplicity n: for each $p_\alpha \in \Gamma_1$, we construct an eigenfunction with value 1 at p_α, -1 at its double, $q_\alpha \in \Gamma_2$, and 0 elsewhere. □

Here all eigenfunctions corresponding to the new n eigenvalue 1 are localized, and the rest of the eigenfunctions have the same nature as before.

6.2.2 Graph splitting

We now turn to a different operation. Let Γ be a graph with an eigenfunction f_1. We arbitrarily divide Γ into subgraphs Σ_0, Σ_1 and Σ_2 such that there is no edge between an element of Σ_1 and an element of Σ_2. We then construct the graphs $\Gamma_1 = \Sigma_1 \cup \Sigma_0$ and $\Gamma_2 = \Sigma_2 \cup \Sigma_0$ in such a manner that each edge between two elements of Σ_0 is contained in either Γ_1 or Γ_2, but not in both of them, and form a connected graph Γ_0 by taking an additional vertices w for each vertex $q \in \Sigma_0$ and connect it with the two copies of q in Γ_1 and Γ_2. We call this process *graph splitting*. Now, we are interested in investigating the change in m_1 after graph splitinig.

Theorem 6.2.5. *Γ_0 possesses the eigenvalue 1 with an eigenfunction that agrees with f_1 on Γ_1.*

Proof. Since f_1 is an eigenfunction of the graph Γ, for each vertex $q \in \Sigma_0$,

$$\sum_{s\in\Gamma;s\sim q} f_1(s) = \sum_{s\in\Gamma_1;s\sim q} f_1(s) + \sum_{s\in\Gamma_2;s\sim q} f_1(s) = 0. \tag{6.20}$$

Hence

$$\sum_{s\in\Gamma_1;s\sim q} f_1(s) = -\sum_{s\in\Gamma_2;s\sim q} f_1(s). \tag{6.21}$$

Now, we construct a function

$$f_1^{\Gamma_0}(p) = \begin{cases} f_1(p) \text{ for } p \in \Gamma_1 \\ -f_1(p) \text{ for } p \in \Gamma_2 \\ -\sum_{s\in\Gamma_1, s\sim q} f_1(s) \text{ when } p = w \text{ is one of the added vertices connected to } q \in \Gamma_1 \end{cases} \tag{6.22}$$

Our claim is that $f_1^{\Gamma_0}$ is an eigenfunction of the graph Γ_0. To prove it, we need to check the balanced condition (6.3) for the vertices q that were in Γ_1 and Γ_2 before splitting, as well as for the vertices w. It is clear that the balanced condition (6.3) will be satisfied for all other vertices.

Now, for any q that was in Γ_1,

$$\sum_{s\in\Gamma; s\sim q} f_1^{\Gamma_0}(s) = \sum_{s\in\Gamma_1; s\sim q} f_1(s) + f_1^{\Gamma_0}(w) = 0, \text{ by using (6.22).} \tag{6.23}$$

For any q that was in Γ_2,

$$\begin{aligned} \sum_{s\in\Gamma; s\sim} f_1^{\Gamma_0}(s) &= \sum_{s\in\Gamma_2; s\sim q} f_1^{\Gamma_0}(s) + f_1^{\Gamma_0}(w) \\ &= -\sum_{s\in\Gamma_2; s\sim q} f_1(s) - \sum_{s\in\Gamma_1; s\sim q} f_1(s), \text{ using (6.21) and (6.22)} \\ &= 0, \text{ using (6.22).} \end{aligned} \tag{6.24}$$

For any w,

$$\sum_{s\sim w} f_1^{\Gamma_0}(s) = \sum_{q\in\Gamma_1; q\sim w} f_1^{\Gamma_0}(q) + \sum_{q\in\Gamma_2; q\sim w} f_1^{\Gamma_0}(q) = 0, \text{ using (6.22).} \tag{6.25}$$

Thus $f_1^{\Gamma_0}$ is an eigenfunction of the graph Γ_0 corresponding to the eigenvalue 1. □

So the unsigned value of any eigenfunction f_1, localized or global, does not change, and neither does the multiplicity of the eigenvalue 1. A simple and special case consists of taking a node p and joining a chain of length 2 to it, that is, connecting p with a new node p_1, and that node in turn with another new node p_2, and setting the function equal to 0 at p_1 and equal to $-f_1(p)$ at p_2. This case was obtained in (Bevis et al., 1997).

6.2.3 Motif joining

The next operation, graph or motif joining, works for any eigenvalue, not just 1:

Theorem 6.2.6. *Let Γ_1 and Γ_2 be graphs with common eigenvalue λ and corresponding eigenfunctions f^1_λ, f^2_λ. Assume that $f^1_\lambda(p_1) = 0$ and $f^2_\lambda(p_2) = 0$ for some $p_1 \in \Gamma_1$ and $p_2 \in \Gamma_2$. Then the graph Γ obtained by joining Γ_1 and Γ_2 via identifying p_1 with p_2 also has the eigenvalue λ with an eigenfunction given by f^1_λ on Γ_1 and f^2_λ on Γ_2.*

Proof. Define a function f^Γ_λ by

$$f^\Gamma_\lambda(p) = \begin{cases} f^1_\lambda(p) & p \in \Gamma_1 \\ f^2_\lambda(p) & p \in \Gamma_2 \end{cases} \tag{6.26}$$

Now our claim is that f^Γ_λ is an eigenfunction of the graph Γ.

We need to check the eigenvalue equation (6.1) for the new vertex p obtained by identifying p_1 and p_2. It is clear that for all other vertices, equation (6.1) will be satisfied. We observe from (6.1) that for an eigenfunction f_λ, whenever $f_\lambda(q) = 0$ at some q, then $\sum_{s\sim q} f_\lambda(s) = 0$ also. This applies to p_1 and p_2, i. e. ,

$$\sum_{s_1\in\Gamma_1; s_1\sim p_1} f^\Gamma_\lambda(s_1) = \sum_{s_2\in\Gamma_2; s_2\sim p_2} f^\Gamma_\lambda(s_2) = 0. \tag{6.27}$$

Now, $f^\Gamma_\lambda(p) = 0$ (since $f^1_\lambda(p) = f^2_\lambda(p) = 0$) and

$$\begin{aligned} \frac{1}{n_p}\sum_{s\sim p} f^\Gamma_\lambda(s) &= \frac{1}{n_p}\sum_{s_1\in\Gamma_1; s_1\sim p_1} f^\Gamma_\lambda(s_1) + \frac{1}{n_p}\sum_{s_2\in\Gamma_2; s_2\sim p_2} f^\Gamma_\lambda(s_2) \\ &= 0 + 0, \text{ by (6.27)}. \end{aligned} \tag{6.28}$$

This proves the claim. □

This includes the case where either f^1_λ or f^2_λ is identically 0. That is, we join a graph Γ_0, with vertex $j_0 \in \Gamma_0$, to an arbitrary vertex i of Γ. If Γ_0 has an eigenfunction u^λ corresponding to the eigenvalue λ of Γ_0, vanishing at j_0 (i.e., $u^\lambda(j_0) = 0$), then the new graph $\bar{\Gamma}$ will also have the eigenvalue λ, with an eigenfunction that agrees with u^λ on Γ_0 and vanishes at all other vertices.

Thus, a motif Γ_0 can be joined to an existing graph in a manner that preserves an eigenvalue and a localized eigenfunction, provided the joining occurs at one (or several) vertices where that eigenfunction vanishes.

Also, when the condition of Theorem 6.2.6 is satisfied at several pairs of vertices, we can form more bonds by vertex identifications between the two graphs.

For the eigenvalue 1, the situation is even better: We need not require that $f^1_\lambda(p_1) = 0$ and $f^2_\lambda(p_2) = 0$, but only that $f^1_\lambda(p_1) = f^2_\lambda(p_2)$ to make the joining construction work.

6.2.4 Examples

By the lemma and node doubling, a chain of m vertices (that is, where we have an edge between p_j and p_{j+1} for $j = 1, \dots, m-1$) possesses the eigenvalue 1 (with multiplicity 1) iff m is odd, with eigenfunction $f_1(p_1) = 1, f_1(p_2) = 0, f_1(p_3) = -1, f_1(p_4) = 0, \dots$. Similarly, a closed chain (that is, where we add an edge between p_m and p_1) possesses the eigenvalue 1 (with multiplicity 2) iff m is a multiple of 4.

Local operations like adding an edge may increase or decrease m_1, or leave it invariant. Adding a pending vertex to a chain of length 2 increases m_1 from 0 to 1, adding a pending vertex to a closed chain of length 3, a triangle, leaves $m_1 = 0$, adding a pending vertex to a closed chain of length 4, a quadrangle, reduces m_1 from 2 to 1 (see (Bevis et al., 1997) for general results in this direction). Similarly, closing a chain by adding an edge between the first and last vertex may increase, decrease or leave m_1 the same.

In any case, the question of the eigenvalue 1 is not a local one. Take closed chains of lengths $4k - 1$ and $4\ell + 1$. Neither of them supports the eigenvalue 1, but if we join them at a single point (that is, if we take a point p_0 in the first and a point q_0 in the second graph, then form a new graph by identifying p_0 and q_0), the resulting graph has 1 as an eigenvalue. An example of an eigenfunction has the value 1 at the joined node, and the values ± 1 alternate along neighboring pairs in the rest of the chains, where the two neighbors of p_0 in the first chain both get the value -1, while the ones in the second chain get the value 1. Here the eigenfunction f_1 has to be global. As another simple example of a global eigenfunction, we can take any two connected graphs Γ_1, Γ_2. Their disjoint union then has two components, and therefore, the multiplicity of the eigenvalue 0 is 2. One eigenfunction u_0 is $\equiv 1$ on Γ_1 and $\equiv 0$ on Γ_2, and for the other one, v_0, the roles of the components are reversed. If we now form a graph Γ by connecting some vertex $i_0 \in \Gamma_1$ to some vertex $j_0 \in \Gamma_2$ by an edge, the multiplicity of the eigenvalue $\lambda_0 = 0$ becomes 1 because Γ is connected; the corresponding eigenfunction u is

$\equiv 1$. However, when both Γ_1 and Γ_2 are large, the next smallest[1] eigenvalue λ_1 of Γ is very small, and a corresponding eigenfunction is well approximated by one, v, that equals a positive constant on Γ_1 and a negative constant on Γ_2 (satisfying $\sum_{i\in\Gamma} n_i v(i) = 0$). Thus, u is a symmetric linear combination and v an antisymmetric one of the original eigenfunctions u_0, v_0, and the eigenvalues are also close.

6.3 Construction of graphs with eigenvalue 1 from given data

Let f be an integer-valued function on the vertices of the graph Γ. We define the excess of $p \in \Gamma$ as

$$e(p) := \sum_{q\sim p} f(q). \tag{6.29}$$

Thus, f is an eigenfunction for the eigenvalue 1 iff $e(p) = 0$ for all p.
We will show that we can construct graphs Γ and functions f with the property that $e(p) = 0$, except for a single vertex p_0 where the pair $(f(p), e(p))$ assumes any prescribed integer values (n, m). These will be assembled from elementary building blocks.

1. A triangle with a function f that takes the value -1 at two vertices and the value 1 at the third vertex, our p_0, realizes the pair $(1, -2)$.

2. The same triangle, with a pending vertex, our new p_0, connected to the vertex with value 1, and given the value 2, realizes $(2, 1)$.

3. Joining instead ℓ triangles at a single vertex, our p_0, with value 1, and assigning the value -1 to all the other vertices as before, yields $(1, -2\ell)$.

4. A pentagon, i.e., a closed chain of 5 vertices, with value -1 at two adjacent vertices and 1 at the remaining three, the middle one of which is our p_0, realizes $(1, 2)$.

5. Similarly, adding a pending vertex, again our new p_0, connected to the former p_0 in the pentagon, and assigning it the value -2, realizes $(-2, 1)$.

6. Likewise, joining instead ℓ such pentagons at p_0 yields $(1, 2\ell)$.

[1] Assuming, for simplicity, that Γ_1, Γ_2 do not have small nonzero eigenvalues themselves.

7. In general, connecting a pending vertex as the new p_0 to the former p_0 changes (n, m) to $(-m, n)$.

8. In general, joining the p_0s from graphs with values $(n, m_1), \dots (n, m_k)$ yields $(n, \sum_1^k m_j)$.

Thus, from the triangle and the pentagon, by adding pending vertices and graph joining, we can indeed realize all integer pairs (n, m).

Theorem 6.3.1. *Let Σ be a graph and f be an integer valued function on its vertices. We can then construct a graph Γ containing the motif Σ with eigenvalue 1 and an eigenfunction coinciding with f on Σ.*

Proof. At each $p \in \Sigma$, we attach a graph realizing the pair $(f(p), -e(p))$. This ensures (6.1) at p. □

6.4 Graph operations and changes of m_1

The preceding constructions also tell us how m_1, the multiplicity of the eigenvalue 1, behaves when we modify a graph Γ', consisting possibly of two disjoint components Γ_1 and Γ_2, by either identifying vertices or by joining vertices with new edges. The graph resulting from these operations will be called Γ. We consider two cases:

1. We identify the vertex p_j with q_j for $j = 1, \dots, m$, assuming that they do not have common neighbors. Then

   (a) We can generate an eigenfunction on Γ whenever we find a function g on Γ' with vanishing excess, except possibly at the joined points, where we require

   $$g(p_j) = g(q_j) \text{ and } e_g(p_j) = -e_g(q_j) \text{ for } j = 1, \dots, m. \tag{6.30}$$

   (b) As a special case of (6.30), an eigenfunction $f_1^{\Gamma'}$ produces an eigenfunction f_1^{Γ} whenever

   $$f_1^{\Gamma'}(p_j) = f_1^{\Gamma'}(q_j) \text{ for } j = 1, \dots, m. \tag{6.31}$$

   In the case where Γ' consists of two disjoint components Γ_1 and Γ_2, this includes the case where both side of 6.31 are 0 for all j and $f_1^{\Gamma'}$ vanishes identically on one of the components. In other words, we can extend an

eigenfunction from Γ_1, say, to the rest of the graph by 0 whenever the function vanishes at all joining points.
Since, in general, (6.31) cannot be satisfied for a basis of eigenfunctions, we can not generate all $m_1^{\Gamma'}$ linearly independent eigenfunctions on Γ by this process.

Whether m_1^Γ is larger or smaller than $m_1^{\Gamma'}$ then depends on the balance between these two processes, that is, how many eigenfunctions satisfy (6.31) vs. how many new eigenfunctions can be produced by functions satisfying (6.30) with nonvanishing excess at some of the joined vertices.

2. We connect the vertices p_j and q_j by an edge for $j = 1, \ldots, m$. Then

   (a) We can generate eigenfunctions on Γ whenever we can find a function g on Γ' with vanishing excess, except possibly at the connected points, where we require

   $$g(p_j) = -e_g(q_j) \text{ and } g(q_j) = -e_g(p_j) \text{ for } j = 1, \ldots, m. \tag{6.32}$$

   (b) Again, as a special case of (6.32), an eigenfunction $f_1^{\Gamma'}$ produces an eigenfunction f_1^Γ whenever

   $$f_1^{\Gamma'}(p_j) = 0 = f_1^{\Gamma'}(q_j) \text{ for } j = 1, \ldots, m. \tag{6.33}$$

   This imposes a stronger constraint than in (6.31) on eigenfunctions to yield an eigenfunction on Γ.

6.4.1 Changes of m_1 by vertex deletion: Observations

Here we will see how the deletion operation on a vertex affects m_1^Γ. It is easy to determine from the rank of the adjacency matrix whether deletion or insertion of a vertex will increase or decrease m_1^Γ by 1 or leave it fixed (Bevis et al., 1997). Now, it is clear that if we delete a vertex which is a duplicate of some other vertex, then m_1^Γ decreases by 1. And if the deletion (of one vertex that is not a duplicate of some other vertex) makes two other vertices into a duplicate pair, then m_1^Γ increases by 1.

Now we can make another observation related to vertex deletion. It is known from the rank of the adjacency matrix that deletion of an edge changes m_1^Γ to $m_1^{\Gamma\setminus e} \in \left[m_1^\Gamma - 2, m_1^\Gamma + 2\right]$ (Bevis et al., 1997). In a graph Γ, let i and j be two such (non-adjacent) vertices. Connecting them with an edge increases (decreases) m_1

by 2. If now instead of joining them by an edge, we delete one of them, m_1 will be increased (decreased) by 1. The reason is as follows:

Denote the graph obtained by deletion of one (vertex i or j) of the vertices by Γ', and the graph obtained by connecting i and j by an edge by Γ''. Now, it is easy to see that we can always obtain Γ'' from Γ' by adding one extra vertex. So $m_1^{\Gamma''} - m_1^{\Gamma'} \leq 1$. But $m_1^{\Gamma''} = m_1^{\Gamma} + 2$. Hence $m_1^{\Gamma'} = m_1^{\Gamma} + 1$. In the same way, we can prove the other case.

6.5 The evolutionary hypothesis and the spectrum

What could be the evolutionary process behind the formation of similar structures? There is an interplay between the dynamics of the network and inheritance structure. The evolutionary processes that are responsible for the construction of the network could be studied via the spectrum of the connectivity matrix. Different graph operations, like motif joining or duplication, produce specific eigenvalues. Constructions with these operations describe certain processes of graph formation that leave characteristic traces in the spectrum. So a useful and plausible hypothesis about evolutionary process can be developed, and it is easy to find the evolutionary assumption that is most relevance for the evolution of the system by investigating the spectrum of a graph constructed from actual data. Let us consider some examples.

- The simplest version of motif duplication is the doubling of a single vertex $j_0 \in \Gamma$. That is, we add a new vertex i_0 and connect i_0 with all neighbors of j_0. This generates an eigenvalue 1, with an eigenfunction u_1 that is localized at j_0 and i_0, specifically $u_1(j_0) = 1$ and $u_1(i_0) = -1$.

  Thus, if the spectral plot of a graph has a high peak at the eigenvalue 1, a natural hypothesis is that this graph evolved via a sequence of vertex doubling.

- In a similar vein, doubling an edge that connects the vertices j_1 and j_2 produces the eigenvalues
$$\lambda = 1 \pm \frac{1}{\sqrt{n_{j_1} n_{j_2}}},$$
which are symmetric about 1 and close to 1 when the degrees of j_1 and/or j_2 are sufficiently large.

Thus, when the spectral peak at 1 is high, but not too sharp, and symmetric about 1, this is an indication that edge duplication has played some role in the evolution of the structure.

As a secific case, doubling an edge that connects the vertices j_1 and j_2 with $n_{j_1} n_{j_2} = 4$ produces eigenvalues 3/2 and 1/2. Three different situations of this kind of edge doubling are described in Fig.6.1.

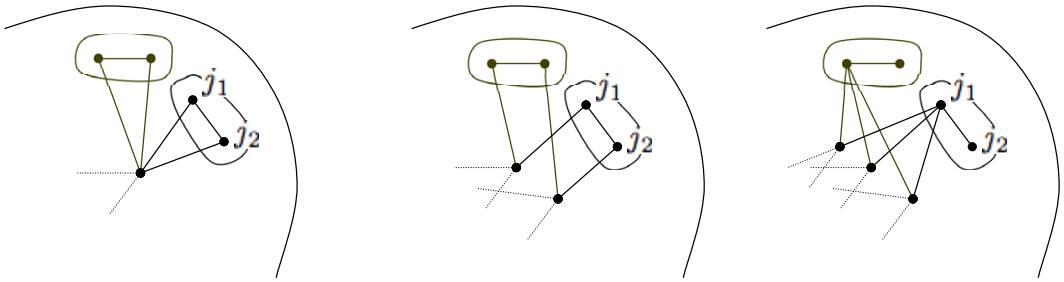


Figure 6.1: Three different edge doubling that produce the eigenvalues 1/2 and 3/2.

- A triangle, that is, a complete graph of 3 vertices, possesses the eigenvalue 3/2 with multiplicity 2.

Next, we connect an edge between the vertices j_1 and j_2 to an existing graph Γ by connecting both j_1 and j_2 via an edge to some vertex $i_0 \in \Gamma$, or equivalently, we join a triangle with vertices j_0, j_1 and j_2 to Γ by identifying j_0 with $i_0 \in \Gamma$ (Fig.6.2(a)). In that case, we produce the eigenvalue 3/2. An eigenfunction u for the eigenvalue 3/2 satisfies $u(j_1) = 1, u(j_2) = -1$, and vanishes elsewhere. Thus, again, it is localized.

The same result is obtained when we join the triangle by connecting j_0 and i_0 by an edge instead of identifying them (Fig.6.2(b)).

A high multiplicity of the eigenvalue 3/2 may then generate the hypothesis that such triangle-joining processes repeatedly occurred in the evolution of the structure.

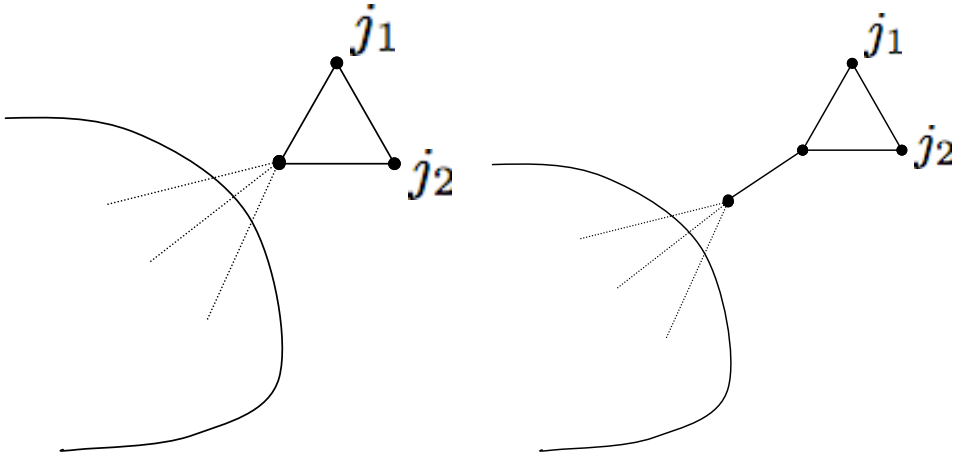


Figure 6.2: Triangle joining. (a) Joined via indentifying the vertices. (b) Joined by an edge.

- Thus, when a triangle is joined at one vertex to another graph, the eigenvalue 3/2 is kept. For instance (see (Chung, 1997)), the petal graph, that is, a graph where m triangles are joined at a single vertex, has the eigenvalue 3/2 with multiplicity $m+1$ (here, m of these eigenvalues are obtained via the described construction, and the remaining eigenfunction has the value -2 at the central vertex where all the triangles are joined and 1 at all other ones). The same result could also be obtained by applying our edge doubling method in a triangle, and then we can deduce $m_{3/2} = m+1$ and $m_{1/2} = m-1$ from equation (6.16).

- The described operations can also be of a global nature.
  For example, we can double the entire graph Γ; when Γ consists of the vertices $p_1, \ldots, p_N$, we take another copy Γ' with vertices $q_1, \ldots, q_N$ and the same connection pattern, and the connect each q_α to all neighbors of p_α. From the graph doubling theorem (6.2.4), the new graph $\bar{\Gamma}$ then has the same eigenvalues as Γ, plus the eigenvalue 1 with multiplicity N.

  This is biologically relevant, because there is some evidence for whole genome duplication. However, protein-protein interaction networks do have a high multiplicity, but not of the order of half the system size (Banerjee & J.Jost,

b). This is readily explained by subsequent mutations after the genome duplication that destroy the symmetry and thereby reduce the multiplicity of the eigenvalue 1[2].

[2]Also, since graph duplication does not change λ_1 and λ_{N-1}, the synchronization properties are not affected.

Chapter 7

Reconstucton of Protein-Protein Interaction Network from Graph Spectrum

Using the spectral plot of the (normalized) graph Laplacian, the essential qualitative properties of a network can be simultaneously deduced. In our study, instead of looking at certain important parameters for empirical networks, we focus on an essentially complete set of graph variables, given by the spectrum of the normalized Laplacian (see, Chapter 4 and Chapter 6). On this basis, we can then develop algorithms that construct networks with all the qualitative properties of those in a given data set. These algorithms consist of reconstruction procedures for a graph, using plausible hypotheses deduced from its spectrum (see Chapter 6.5). For biological networks, we can thereby retrace the regularities in their evolutionary history. Here, we demonstrate this principle and apply this method for protein-protein interaction networks (PPIN for short). We detect indications of evolutionary duplication and divergence, as argued in (Huynen & Bork, 1998; Wagner, 2001).

This approach then also sheds light on a somewhat different issue, namely that of which features and properties are distinctive for networks from particular empirical classes, as opposed to universal features shared across classes.

7.1 Protein-protein interaction networks (PPINs)

Activation times and actions of proteins can vary within the cell (e.g., many proteins act as enzymes, catalysts of metabolic reactions, signal-components for cell signaling, ionchannels in the cellular membrane, etc.). Most of the, time proteins

don't act alone at the biochemical level, rather they interact with other proteins as an assembly to perform particular cellular tasks. In protein-protein interaction networks, we consider a single protein to be a node and a direct interaction of two proteins to be an (undirected) edge between them. Fig. 7.1 gives an impression of the topological structure of this kind of network. In general, these networks consist of one giant component, many small components and many isoloated proteins. Our analysis will always be performed on the giant components of these networks, so as to work with connected graphs, and we will neglect all the small components and isolated proteins.

Figure 7.1: General topological structure of protein-protein interaction networks.

7.2 Spectral plot and structural analysis of PPIN

In spite of their rather wide range of sizes, the spectral plots of different PPINs[1] share a particular pattern (Fig. 7.2; the spectral density is given as a sum of Lorentz distributions, $\rho(\lambda) = \sum_{k=1}^{N-1} \frac{\gamma}{(\lambda_k-\lambda)^2+\gamma^2}$, with width $\gamma = .08$). The most prominent feature is the sharp peak around the eigenvalue 1.[2] In spite of possible

[1] See the data source for details.

[2] A high multiplicity of eigenvalue 1 has also been observed in other networks, like the Internet (Vukadinovic, Huang, & Erlebach, 2002).

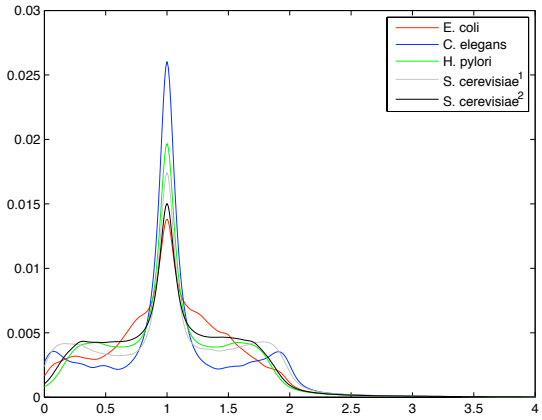


Figure 7.2: Spectral plots of all protein-protein interaction networks. Spectral density is given as a sum of Lorentz distributions, $\rho(\lambda) = \sum_{k=1}^{N-1} \frac{\gamma}{(\lambda_k-\lambda)^2+\gamma^2}$, with width $\gamma = .08$.

statistical fluctuations affecting the acccuracy of the underlying data, we can also determine the ranges of various structural parameters (Fig. 7.3) and motif numbers (Fig. 7.4): (N below is the size of the network)

$Maximum\ Degree < \frac{N}{10}$
$1.56N < Number\ of\ edges < 1.97N$
$0.307N < m_1 < 0.445N$
$0.015 < Transitivity < 0.028$
$0.017 < QF_1 = \frac{4M_6}{M_3+4M_6} < 0.067$
$0.0345 < QF_2 = \frac{4M_6+6M_7}{M_3+4M_6+6M_7} < 0.0837$

In particular, the multiplicity m_1 of the eigenvalue 1, the transitivity, QF_1 and QF_2 are much larger than in random graphs of Erdős-Rényi type with a similar number of vertices and edges. (Also, the numbers of 4-motif structures are higher, but, for example, the many M_4 motifs are readily explained by a power-law degree distribution.)

The multiplicity of the eigenvalue 1, m_1 of Δ is particularly significant. Node duplication increases m_1 by 1[3] (see Chapter 6.2.1 for details). For this reason, it constitutes an important invariant for our investigation of protein-protein interaction networks. Also, if we duplicate a particular node m times, then the number of specific motifs containing that node will grow like $\binom{m}{2}$ (see, Theorem 6.2.2); again,

[3]One should note, however, that the determination of m_1 is not a local question in general.

Name	*Size*	*Number of Edges*	*Max. Degree*	m_1	*Transitivity*	QF_1	QF_2
Escherichia coli	230	695	36	57	0.1552452	0.297609	0.4083
Caenorhabditis elegans	314	363	28	190	0.01808067	0.067327	0.0837
Helicobacter pylori	710	1396	55	316	0.01523657	0.042377	0.047
Saccharomyces cerevisiae^1	1458	1948	56	564	0.05177614	0.017670	0.0529
Saccharomyces cerevisiae^2	3930	7725	282	1206	0.02821804	0.0179	0.0345

Figure 7.3: Parameter values

that is then something that can easily be detected in given network data.

More generally, the effects of a motif on the spectrum of the Laplacian have been systematically investigated in Chapter 6.2 (see also (Banerjee & J.Jost, c)), and conversely, the analysis of the spectrum then gives indications about the evolution of a network via such processes as motif duplication, joining or splitting. Also, the largest eigenvalue characterizes how different the graph is from a bipartite graph (see Chapter 4.2.2).

A few comments on the motif structures of PPINs: Some observations could be made from the number of motif structures of PPINs (Fig. 7.4), e.g., that the number of motifs M_1 is less than the number of motifs M_3 for all PPINs. The graph could contain some structures that would cause this phenomenon, e.g., simple cycles (possibly triangles) with higher degree nodes that have many neighbors of degree 1. Another possibility could be that many triangles (motifs M_2) share only one common node. In the latter situation, there could be a higher multiplicity of the eigenvalue 3/2 ($m_{3/2}$), e.g., the spectra of *Escherichia coli*, *Saccharomyces cerevisiae*^{1}, and *Saccharomyces cerevisiae*^{2} contain the eigenvalue 3/2.

In a few PPINs (*Escherichia coli*, *Helicobacter pylori*, *Saccharomyces cerevisiae*^{2}), number of motifs M_7 is higher than the number of motifs M_2. Many triangles sharing a common edge is one of the possible structural conformations behind this phenomenon. In this case, though the graph contains many triangles, but many eigenvalues 3/2 will be lost. For instance, in *Saccharomyces cerevisiae*^{2}, $m_{3/2} = 4$, in *Escherichia coli*, $m_{3/2} = 6$ (which is relatively low compared to the high numer of triangles), and in *Helicobacter pylori*, there is no eigenvalue 3/2.

Motifs \ Names	*Escherichia coli*	*Caenorhabditis elegans*	*Helicobacter pylori*	*Saccharomyces cerevisiae*[1]	*Saccharomyces cerevisiae*[2]
M_1 	7803	2118	14736	11318	161378
M_2 	478	13	76	260	1562
M_3 	54613	2826	93825	30908	1117123
M_4 	42464	10354	120585	71905	6459415
M_5 	12612	139	3328	2547	68576
M_6 	5785	51	1038	139	5089
M_7 	2425	9	79	195	3256
M_8 	230	0	0	39	590

Figure 7.4: Number of 3-motif and 4-motif structures

7.3 Data sources

The protein-protein interaction data sets for *Saccharomyces cerevisiae*[1] (yeast) are from http://www.nd.edu/∼networks/, used in (Jeong et al., 2001) [download date: 17 September 2004]. The ones for *Escherichia coli* are as used in (Butland et al., 2005), *Caenorhabditis elegans, Helicobacter pylori* and, as a check, a second data set for *Saccharomyces cerevisiae*[2] is taken from http://www.cosin.org [download date: 25 September, 2005].

7.4 Previous models

First, our spectral analysis reveals that a PPIN is rather different from a generic network model, like an Erdős-Rényi random graph, a small world Watts-Strogatz graph, or a Barabasi-Albert scale-free graph, as a comparison with the corresponding spectral plots directly shows (see Fig. 5.5). This implies that for understanding the specific features of PPINs, more specific models also are needed. Two main basic evolutionary processes are important for the growth and evolution of PPINs: duplication of a protein (node) and mutation of the connections (edges). Protein duplication, which is the outcome of gene duplication, plays a major role in increasing the size of the network. PPIN evolution via gene duplication has been

modelled previously, with different assumptions. Still, spectral analysis shows that these models are able to recover some, but not all structural properties of real PPINs.

In (Ispolatov, Krapivsky, & Yuryev, 2005), a randomly selected protein (node) is duplicated and the connections (edges) of the new protein are subsequently deleted with the same probability. In (Pastor-Satorras et al., 2003), after a duplication step, new connections to other proteins are made with equal probability. In (Kim, Krapivsky, Kahng, & Redner, 2002), with similar assumptions, the emergence of the giant component in a PPIN was investigated. All these models succeed in constructing a network with a degree distribution similar to a real PPIN. But these models did not aim at certain other structural parameters, and therefore produce a very small *transitivity* or *clustering coefficient*, essentially because the schemes do not readily produce triangles. More generally, the spectral plot for a network produced by the method of (Pastor-Satorras et al., 2003) (Fig. 7.5) is rather different from the real data (although definitely not as bad as the generic models), and the same holds true for various structural parameters:

Maximum degree ≈ 56
Number of edges ≈ 1297
Multiplicity of eigenvalue $1 \approx 141$
Transitivity ≈ 0.0077
$QF_1 \approx 0.1503$
$QF_2 \approx 0.1533$
Here the network size is 500. Parameters: $\delta = 0.58$ *and* $\alpha = (2\delta - 1)/N_t$, *where* N_t *is the size of the network at time step* t.

Networks with different *transitivity* or *clustering coefficients* can, however, be constructed by tuning the probability p in the model of (Vázquez, Flammini, Maritan, & Vespignani, 2003). That parameter incorporates the probability of cross interactions[4] between the old protein and its duplicated copy, for example, resulting from self-interactions of the old one. A realistic value of p can then be determined from the data in (Wagner, 2001; Wagner, 2003) and is smaller than 0.018. That

[4] There are two biological reasons behind the cross interaction between the protein that is duplicated and the protein that is produced from duplication. The first is simple mutation; a new interaction could be formed due to mutation of either of those two proteins. The second is the existence of self-interactions of the protein which is duplicated and in that case the two proteins will have crossinteractions (in the second case, of course, the crossinteraction could be deleted afterwards because of a mutation of the interaction).

upper bound is the value employed in (Vázquez et al., 2003), but this turns out to lead to too small a value for the transitivity of the giant cluster. Therefore, in our model we assumed that, with some low probability, there is a preference for a protein to make new connections with its 2nd neighbors.

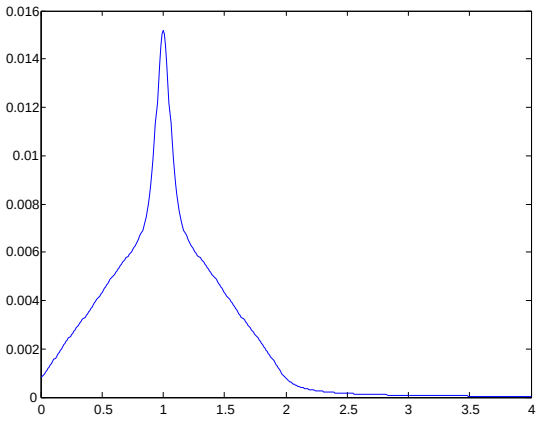


Figure 7.5: Spectral plot for a protein-protein interaction network (model) produced by the method of (Pastor-Satorras et al., 2003). Spectral density is given as a sum of Lorentz distributions, $\rho(\lambda) = \sum_{k=1}^{N-1} \frac{\gamma}{(\lambda_k - \lambda)^2 + \gamma^2}$, with width $\gamma = .08$.

7.5 Biological processes in PPINs

All the above models are based on the duplication and mutations of the connections of duplicated proteins. They all assume that mutation occurs right after the duplication. It is true that genome evolution analysis (Wagner, 2001; Huynen & Bork, 1998) supports the idea that the divergence of duplicated genes takes place shortly after the duplication, but only indirect evidence is available for rapid functional divergence after gene duplication (Wagner, 2001). So we cannot say that a divergence process always occurs just before any new duplication takes place. There is no prominent time scale separation for the processes of duplication and mutation. Furthermore, these two processes take place simultaneously, but at different rates. The overall rate of link dynamics, the mutation of interactions between existing proteins, is at least an order of magnitude higher than the growth rate of the network due to gene duplication (Berg, Lassig, & Wagner, 2004). In link dynamics, attachment (creation of a new edge due to mutation) occurs preferentially towards partners of high connectivity (Berg et al., 2004).

7.6 New model and network reconstruction

Our constructive model for PPINs is inspired by general evolutionary considerations. The basic evolutionary processes for the growth and evolution of PPINs are duplication of a protein (node) and mutation of the connections (edges).

Instead of cross links between the old protein and its duplicated copy – which would produce values for the transitivity that take too small – a low probability preference for 2nd order neighbors as recipients of new connections is assumed. New connections with other proteins then occur with a different probability. Since in link dynamics, attachment occurs preferentially towards partners of high connectivity (Berg et al., 2004), some preferential attachment to proteins with higher connections is included. In contrast, deletion is random with a uniform probability.

Since on the one hand, genome evolution analysis (Wagner, 2001; Huynen & Bork, 1998) supports the idea that the divergence of duplicated genes takes place shortly after the duplication, but on the other hand, only indirect evidence is available for rapid functional divergence after gene duplication (Wagner, 2001), we have considered two different mutation processes:

1. A random deletion process that is independent of the duplication process and occurs uniformly with probability δ, as well as two different kinds of addition processes with preference towards a partner with high degree:

   (a) Connection with protein i at distance 2 with probability $\frac{d_i}{\sum_i d_i}\alpha_1$, where d_i is the degree of protein i and α_1 is a parameter.

   (b) Connection with another protein i (which could even be in another component) with probability $\frac{d_i}{\sum_i d_i}\alpha_2$, with a parameter α_2.

2. A deletion with probability δ' that occurs for $\frac{1}{3}$ of the duplications and shortly after such a duplication. This process operates by elimination of one of the two interactions in each redundant interaction pair of two duplicate proteins with equal probability. For simplicity, there is no edge addition for this mutation process.

To make the duplication process independent of the first mutation process and to make the duplication rate lower than the mutation rate, duplication occurs with probability P_{dup} and with a preference that is the inverse of the square-root of the degree of the protein.

A component of the network can grow by duplication of proteins within that component, or by attachment of other components or isolated proteins. Here, we have neglected isolated proteins, but the model can be readily extended by attachment of isolated proteins with some probability P_{add}. One might also include a mechanism for cross link connections between duplicate protein pairs with some probability P_{CLink}, but the same effect can be achieved by tuning the other parameters.

The algorithm starts with a small seed network of two linked proteins. The growth procedure is run until the giant component reaches the desired size of our network. 100 repetitions are performed with parameter values
$P_{dup} = 0.15$
$\delta' = 0.7$
$\delta = 0.00025$
$\alpha_1 = 0.00008$
$\alpha_2 = 0.0002$
$P_{add} = 0.025$
$P_{CLink} = 0.008$

The structural properties of the resulting giant component (size ≈ 500) are:
$Maximum\ degree \approx 43.69$
$Number\ of\ edges \approx 712.97$
$m_1 \approx 161.07$
$Transitivity \approx 0.02793$
$QF_1 \approx 0.07199$
$QF_2 \approx 0.08016$

Thus, the spectral plot (Fig. 7.6) and the structural properties of the giant component of the simulated network match the real PPIN data closely.[5]

[5]The spectrum of the Laplacian is always confined between 0 and 2. This is not quite exhibited by our spectral plots, due to the positive width of the kernel employed in our visualization.

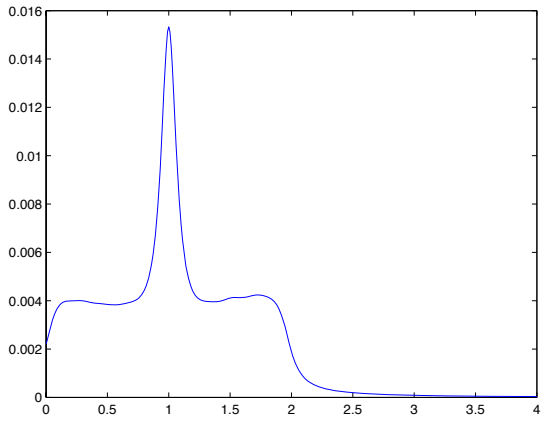


Figure 7.6: Spectral plot for a protein-protein interaction network (model) produced by the method described above. Spectral density is given as a sum of Lorentz distributions, $\rho(\lambda) = \sum_{k=1}^{N-1} \frac{\gamma}{(\lambda_k - \lambda)^2 + \gamma^2}$, with width $\gamma = .08$.

Chapter 8

Qualitative Classification of Real Networks by Spectral Plot Properties

8.1 Introduction

Real networks have very complicated and irregular structures. In Chapter 3.2, we have discussed the difficulties to analyze and visualize the graph structure. To capture the qualitative aspects and characteristics of a graph, we have therefore introduced a spectral method that have been discussed in Chapter 5 and Chapter 6. The spectrum of the normalized Laplacian captures the properties that can characterize specific classes of networks, and the spectral plot, which can be easily visualized (see, Chapter 5), reflects the characterization of a network (see also (Banerjee & J.Jost, a; Banerjee & J.Jost, 2007)). Here, we are introducing a tentative classification scheme[1] for empirical networks based on global qualitative properties detected through the spectrum of the Laplacian of the graph underlying the network. This classification is not only robust towards fluctuations and perturbations within a given class, but can also readily distinguish different types. It can also be easily and directly visually inspected.
Even though empirical networks typically have directed and weighted edges, we consider here only the underlying undirected and unweighted graphs. The methods utilized, however, easily extend to the directed and weighted case, but it turns out that the reduced graph already carries a lot of structural information about the network. This, as well as space constraints, is our rationale for that simplification.

[1]It has been also discussed in (Banerjee & J.Jost, d).

8.2 Network classes

In order to gain some orientation, we start with some spectral plots of artificial, that is, simulated networks. Our first examples come from two classes of regular networks. The first one consists of regular 2d square grids, with 10,000 nodes. As we see in Fig. 5.8, when we make the grid narrower and longer, the spectrum shows characteristic side peaks. The spectral plot is symmetric about 1, as all these graphs are bipartite. When we add one of the two possible diagonals (always choosing the same) in each square, we destroy bipartiteness and get a systematic shift in the spectral plot (Fig. 5.9), again with the side peaks when the grid gets narrower. The other regular graphs originate from a circular arrangement of 1,000 nodes, where we connect each node with the 2, 4, 6, 10, 20, or 50 closest nodes on the circle (Fig. 5.4). When thus progressing to higher degrees, we see an eventual merging at 1 of the two peaks that start out at 0 and 2 for small degrees.

We next turn to stochastically constructed graphs, an Erdős-Rényi random network, a Watts-Strogatz small-world network and a Barabási-Albert scale-free network (Fig. 5.6) (see Chapter 5.2 for discussions on these plots). Fig. 5.10 shows a small-world graph constructed by rewiring a square grid, with rewiring probability 0.3. The spectral plot becomes characteristically different from the regular one.

We now turn to empirical networks and compare their spectral plots both with each other and with the model types presented above. Bellow, we shall have to keep in mind below, however, that some of the empirical networks are quite small, on the order of 100 nodes only, and so obviously random fluctuations may have stronger effects, suggesting some caution concerning the robustness of our classification. The first type comprises several classes of biological networks at the molecular level, including metabolic, transcription, signal transduction, and protein-protein interaction networks, as well as word adjacency and internet topology graphs (Fig. 5.11, Fig. 5.12, Fig. 5.13, Fig. 5.16 and Fig. 5.17]). The characteristic features are the very high peak at or near 1, the shallow rest with two secondary peaks, and the high degree of symmetry about 1. As we recall from the mathematical discussion in chapter 6, these graphs then come close, in spectral terms, to a complete bipartite graph which, as we discussed, arises through repeated node duplication. Simulations that we present elsewhere (Banerjee & J.Jost, b) indicate that the secondary peaks arise from the random deletion of edges after the node duplications. For each of these empirical classes, one can then try to find an explanation of their evolution or construction through such processes, like gene duplication in the biological case. Our second class contains weblog hyperlink graphs (in US pol-

itics), conformation spaces of polypeptides, food webs, and, with less confidence, email interchanges (Fig. 5.18, Fig. 5.19, Fig.5.15, Fig. 5.20]). Neuronal connectivity graphs of *C. elegans* constitute a borderline case (Fig. 5.14). This second class is characterized by a concentration near 1, though not as sharply peaked of one as in the first case, and, except for the neuronal network, again symmetry about 1. This class is different from all the model types, but shows a little similarity with the scale-free type. The third class contains power grids, coauthorships between scientists, copurchasing of books, and US football games (Fig. 5.21, Fig. 5.22, Fig. 5.24, Fig. 5.25). They all resemble the class of square grids with diagonals, moving from the less narrow to the very narrow ones. Finally, the electronic circuit graph spectra (Fig. 5.23) resemble those of a narrow square grid without diagonals.

8.3 Conclusion

In this chapter, we have presented a scheme for the rough classification of empirical networks in terms of their qualitative spectral properties. Since we can also understand from mathematical theory how some of those characteristic spectral properties are caused by topological properties of the underlying graph or can emerge from processes like node duplication, random rewiring, random edge deletion etc., this scheme also offers the potential for systematic insights into the evolution or the emergence of global properties of specific classes of empirical networks. As usual with mathematical structures, structural similarities can be shared across empirical domains.

Of course, this represents at best the first step towards a systematic theory of complex networks. Perhaps the current state is somewhat similar to that of cellular automata about 25 years ago, when classifications in terms of visually representable global features were also proposed. Not all of what was proposed then could be consolidated by subsequent research, but it nevertheless opened up a fruitful perspective.

Chapter 9

Graph Coarsening and Spectral Plots

9.1 Introduction

With ever-increasing computational capacity and experimental technology, we can capture larger networks in many systems. Due to limited computation power, it sometimes becomes hard to calculate the whole spectrum of a large graph. Most real networks are sparse, but there are some giant networks, and because of their huge sizes, the number of edges (hence the number of nonzero elements in the connectivity matrices) is large and the computation of the spectrum needs a lot of time and space. For example, a WWW network of size 325729 has 2505945 nonzero elements, which is very large for computational purposes. So we need to find some procedure for generating the spectral plot without spending much time and space on the computation. If it is not possible to produce the plot for the original graph, at least we can try to generate a similar one.

In our case, we can also exploit the eigenvalue equation of the normalized Laplacian, $\lambda u(i) = u(i) - \frac{1}{n_i}\sum_{j,j\sim i} u(j)$, with the constraint $\sum_i n_i u(i) = 0$ for the eigenvector, as well as convolution with a kernel for plotting the spectrum, e.g., $f(x) = \sum_{\lambda_j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{|x-\lambda_j|^2}{2\sigma^2})$ (here the example is with Gaussian kernel). There are many ways to deal with the problem of space, but it is hard to reduce the computational complexity. We can apply the spectral mapping theorem and expand the convolution equation with Gaussian kernel in an exponential series, as

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{\lambda\in\sigma(\Delta)} \exp(-\frac{|\lambda - x|^2}{2\sigma^2}) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{\lambda'\in\sigma\left(\exp(-\frac{(\Delta - Ix)^2}{2\sigma^2})\right)} \lambda' \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{T}race\big[\exp(-\frac{(\Delta - Ix)^2}{2\sigma^2})\big]
\end{aligned}
\tag{9.1}
$$

(here $\sigma(\Delta)$ be the spectrum of the matrix Δ). Now, without expanding the exponential series, we can also calculate the value of $\exp(-\frac{(\Delta-Ix)^2}{2\sigma^2})$ as a solution of a first order linear differential equation: $y' = By$, where $B = -\frac{(\Delta-Ix)^2}{2\sigma^2}$. This method gives an accurate spectral plot without calculating the eigenvalues of the matrix[1], but the computational complexity for getting a good result is very high. To deal with this problem, we propose here a method for coarsening, or reducing the size of a graph while keeping the pattern of structural connectivity similar and therefore also the spectral plot.

9.2 Coarsening scheme

First we need to index all vertices, keeping in mind that any two vertices that are at short distance[2] from each other should also have close-together indices. We can formalize the indexing scheme by introducing a good algorithm to minimize the difference between indices of two vertices according to the distance between them. But in real network data, available from various sources, it is already minimized to a greater extent, because it is intuitive to index the vertices sequentially according to their distance. For example, while indexing the vertices of a friendship network, we usually give the next index to a vertex which is a friend of the previously indexed vertex. So for the moment, for the sake of simplicity, we are not much concerned about the indexing scheme for a real network, but of course things will work better if we reindex the vertices of a graph with a good algorithm.

Scheme: Create a new graph by giving a single index, i_α, to each pair of consecutive vertices of the old graph, $i_\alpha \equiv (j_{2\alpha-1}, j_{2\alpha})$, where j_β are vertices of the old graph and i_α are vertices of the new graph. Now, if there exists at least one

[1]This method is proposed by Ulrich Steinmetz

[2]Here we always mean Hamming distance between two vertices.

edge between two vertices belonging to two different pairs, then we place an edge between the vertices represent those two pairs in the new graph.

As an example, let $i_\alpha \equiv (j_{2\alpha-1}, j_{2\alpha})$ and $i_\beta \equiv (j_{2\beta-1}, j_{2\beta})$ be two vertices in the new graph after coarsening. Now, if any element of the set $\{j_{2\alpha-1}, j_{2\alpha}\}\times\{j_{2\beta-1}, j_{2\beta}\}$ belongs to the edge set of the old graph, then there will be an edge between i_α and i_β in the new graph.
One iteration of this process reduces the size of the graph by half. Instead of grouping two old vertices into one vertex in the new graph, we can choose groups with any fixed number d of old vertices, and then the size of the new graph will be reduced by $1/d$.
Also, we can make groups of vertices that are less than a fixed distance apart from one another (Song, Havlin, & Makse, 2005). Here, the sizes of the groups can be very different. This is an efficient method of renormalization, but to avoid the computational complexity involve and to maintain simplicity of arguments, we will not considering this process now.

We now explore the efficiency of our coarsening scheme with some examples (Fig. [9.1 - 9.6]; All examples are with one iteration. All plots are with Gaussian kernel, $\sigma = 0.03$. Blue line shows the original plot and the red line shows the plot after one iteration.). Surprisingly, for many cases, this simple process works fine. E.g., for a random graph (Fig. 9.1), neuronal network, power-grid network and two of the food-web networks (see, Fig. 9.3, Fig. 9.6, Fig. 9.5(a) and Fig. 9.5(b)), it works very well, but this is not the case for all networks. Specially, for graphs which have sharp peak at 1 in their spectral plot (See, Fig. 9.4(a) and 9.4(b)), precisely the graphs belonging to the first category (see chapter 8), the scheme fails. Also, for the scale-free network (Fig. 9.2(b)), small-world graph Fig. 9.2(a) and one of the food-web networks (Fig. 9.5(c)) it doesn't work well.

What would happen if we pass to further iterations? Lets take the examples of the neuronal conncetivity and power-grid networks (Fig. 9.7 and Fig. 9.9). Since the sparsity pattern of the connectivity matrix has a great influence on the spectrum, we consider here the sparsity pattern of the connectivity matrices at each iteration (see, Fig. 9.8 and Fig. 9.10). The qualitative characteristice of the spectral plots of the neuronal connectivity graph does not change much with higher iterations (Fig. 9.7). But for the power-grid graph, it becomes more symmetric around 1 (see, Fig. 9.9). If, after one iteration, the number of nonzero elements does not decrease by half, as the size of of the network does, the pattern of the plot changes. Bad indexing is responsible for this situation. To get a similar pattern

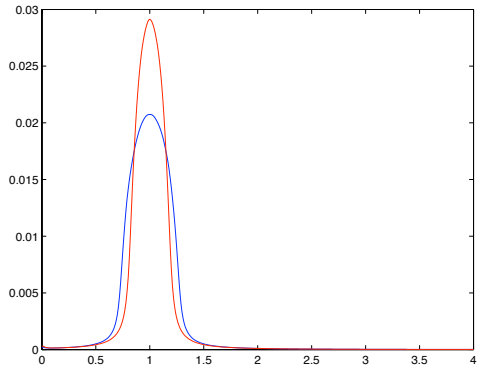


Figure 9.1: Spectral plots of random network from Erdős and Rényi's model (Erdős & Réanyi, 1959) with $p = 0.05$.

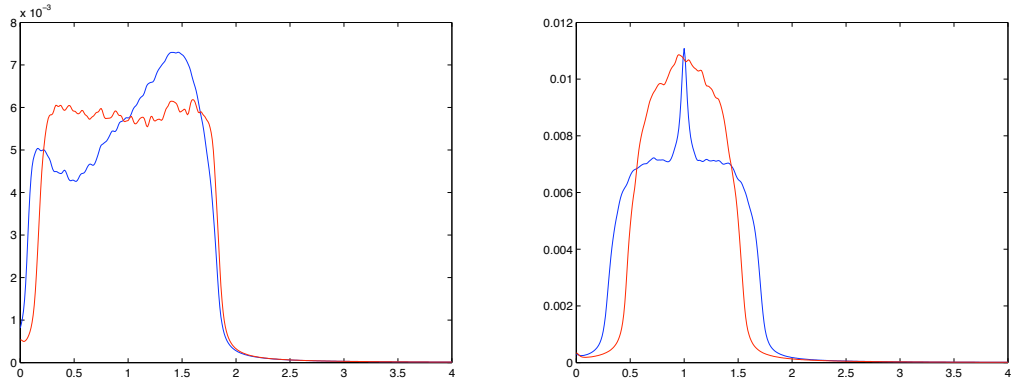
after many iterations, it is better to use an optimizing indexing strategy to keep the rate of decrease of the number of nonzero elements same. Rather than randomizing the graph, this scheme finds the pattern of the backbone of the graph. Hence, with increasing iterations, a regular ring lattice with higher-degree vertices converges to a simple cycle (Fig. 9.11).

9.3 Spectral plot of WWW network

As we discussed before, the problem of computing the entire spectrum of the WWW network is the enormous size of this network[3]. But we are nevertheless interested in knowing what the spectral plot of this network looks like? One can also ask whether this graph belongs to the first category, whose evolutionary dynamics is more understandable than others.

We shall apply our coarsening scheme to find the answers to these questions. We have iteratively applied our coarsening scheme 7 times (Fig. 9.12). In the first iteration, the number of nonzero element decreases by more than one half, and in the fourth to seventh iterations, it decreases by less than one half (see, Fig. 9.13 and Fig. 9.14). Hence, it is clear that the original spectral plot has changed to some extent after many iterations. But the spectral plots in 5th to 7th iterations

[3]The particular network we are analyzing here has been downloaded from http://www.nd.edu/∼networks/resources.htm.



(a) (b)

Figure 9.2: Spectral plots of (a) Small-world network from the Watts-Strogatz's model (Watts & Strogatz, 1998) (rewiring a regular ring lattice of average degree 4 with rewiring probability 0.3). (b) Scale-free network from the Albert-Barabási's model (Barabási & Albert, 1999) ($m_0 = 5$ and $m = 3$).

show that WWW network cannot belong to the first category, yet there is a high possibility that it is in the second category (see Fig. 9.12).

9.4 Conclusion

Our coarsening scheme (without reindexing) does not always produce the same spectral plot as the original graph, but it captures the basic pattern of the plot. The scheme is very simple and easy to implement. It reduces the size of the graph, hence also the computational complexity and the memory required to compute the whole spectrum. The scheme itself has the computatonal complexity of $O(N^2)$, where N is the number of vertices of the graph. To make the scheme more accurate we can use a good indexing strategy, with computational complexity of order not higher than N^2 to keep the total complexity for the entire process same.

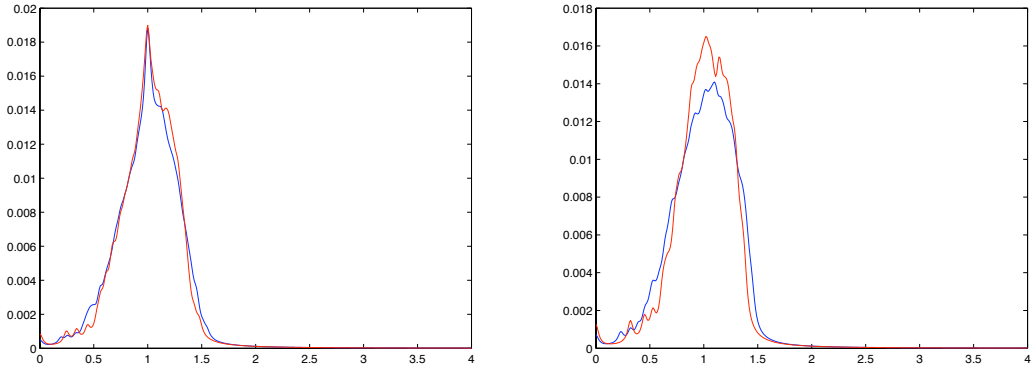


Figure 9.3: Spectral plots of neuronal connectivity. (a) *Caenorhabditis elegans*. Network size: 297. Data obtained from (Watts & Strogatz, 1998; White et al., 1986). Data Source: http://cdg.columbia.edu/cdg/datasets [Download date: 18 Dec. 2006]. (b) *Caenorhabditis elegans* (animal JSH, L4 male) in the nerve ring and RVG regions. Network size: 190. Data source: Data assembled by J. G. White, E. Southgate, J. N. Thomson, S. Brenner (White et al., 1986) and revisited by R. M. Durbin (Ref. http://elegans.swmed.edu/parts). [Download date: 27 Sep. 2005].

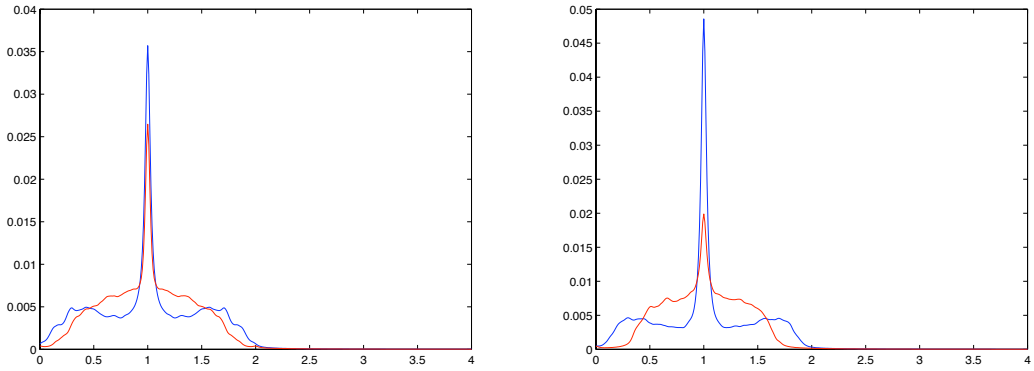


Figure 9.4: Spectral plot of (a) A metabolic network of *Caenorhabditis elegans*. Size of the network is 1173. Nodes are substrates, enzymes and intermediate complexes. Data obtained from (Jeong et al., 2000). Data Source: http://www.nd.edu/∼networks/resources.htm. [Download date: 22 Nov. 2004]. (b) A protein-protein interaction network of *Helicobacter pylori*. Network size: 710. Data downloaded from http://www.nd.edu/∼networks and data obtained from (Jeong et al., 2001) [download date: 17 September, 2004].

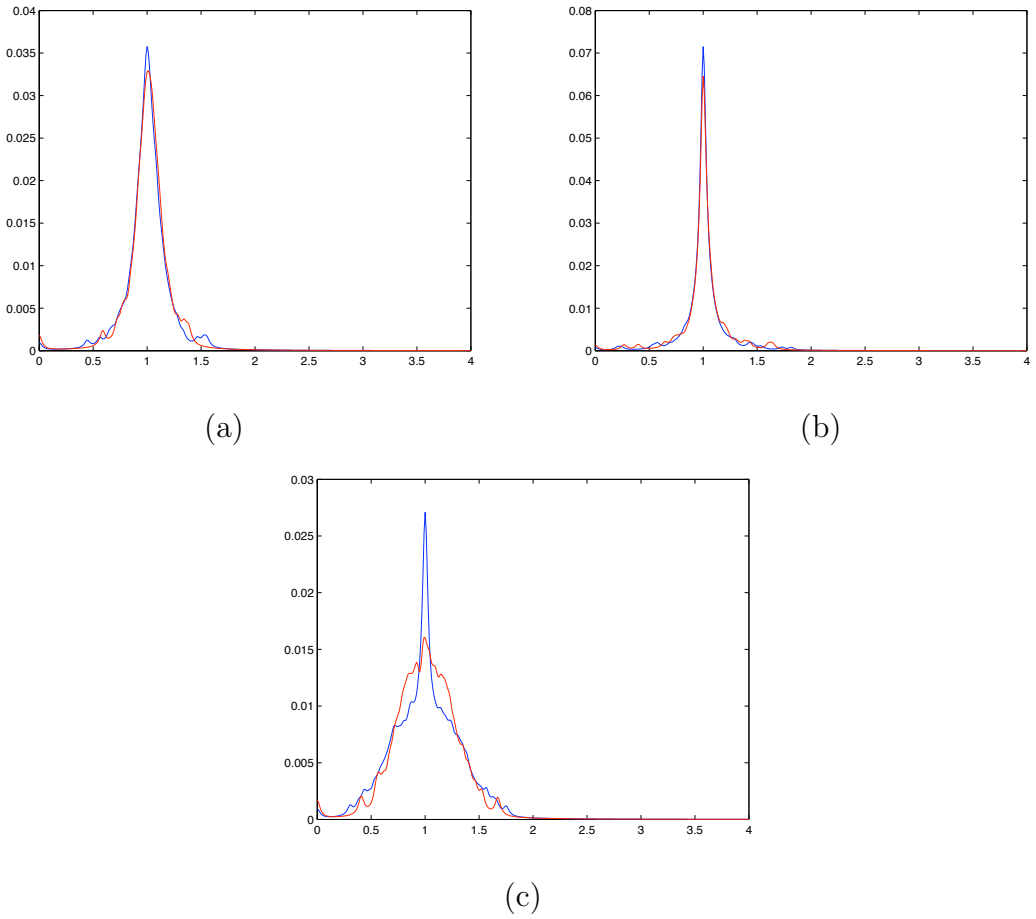


Figure 9.5: Spectral plots of Food-web. (a) From "Florida bay in dry season". Data downloaded from http://vlado.fmf.uni-lj.si/pub/networks/data (main data resource: Chesapeake Biological Laboratory. Web link: http://www.cbl.umces.edu). [Download Date 21 December, 2006]. Network size: 128. (b) From "Little rock lake". Data downloaded from http://www.cosin.org. [Download Date 21 December, 2006]. Size of the network is 183. (c) From "Ythan estuary". Data downloaded from http://www.cosin.org. [Download Date 21st December, 2006]. Network size: 135.

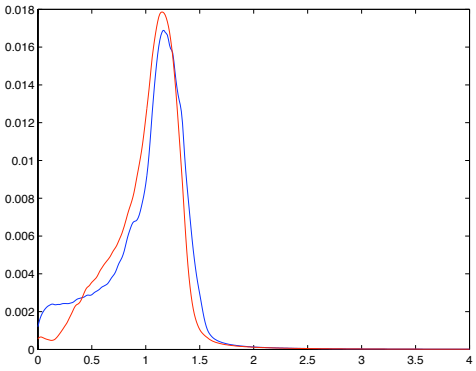


Figure 9.6: Power-Grid network of the Western States of the United States (Watts & Strogatz, 1998). Network size: 4941. Data downloaded from http://cdg.columbia.edu/uploads/datasets [Download date: 1 March, 2007.].

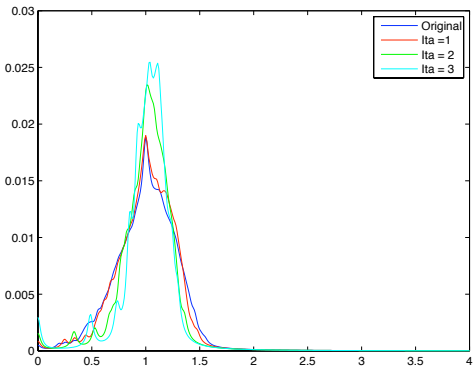


Figure 9.7: Spectral plots of neuronal connectivity of $Caenorhabditis\ elegans$. Size of the original network is 297. Data obtained from (Watts & Strogatz, 1998; White et al., 1986). Data Source: http://cdg.columbia.edu/cdg/datasets [Download date: 18 Dec. 2006]. 'Original' corresponds to the plot of original network. 'Ita' shows plots after a particular number of iterations. All plots are with Gaussian kernel, $\sigma = 0.03$.

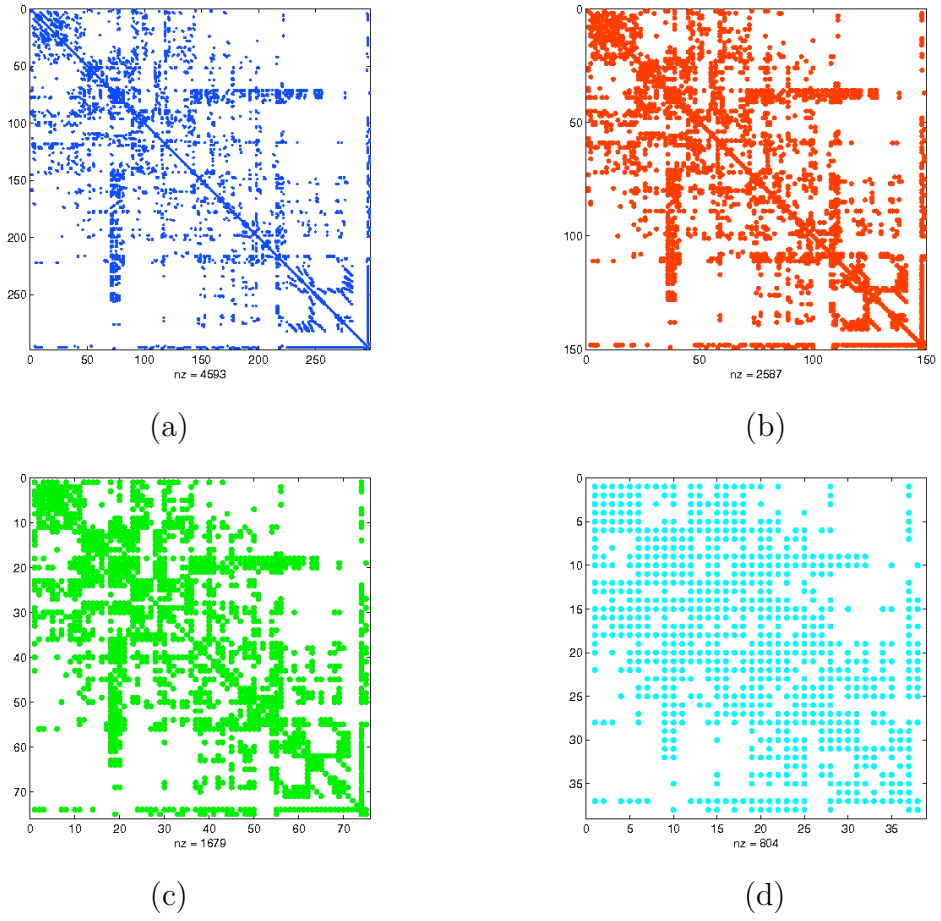


Figure 9.8: Sparsity patterns of the connectivity matrix of neural network in Fig. 9.7. (a) Original network. (b) After one iteration. (c) After two iterations. (d) Afther three iterations. 'nz' is the number of nonzero elements in the matrix.

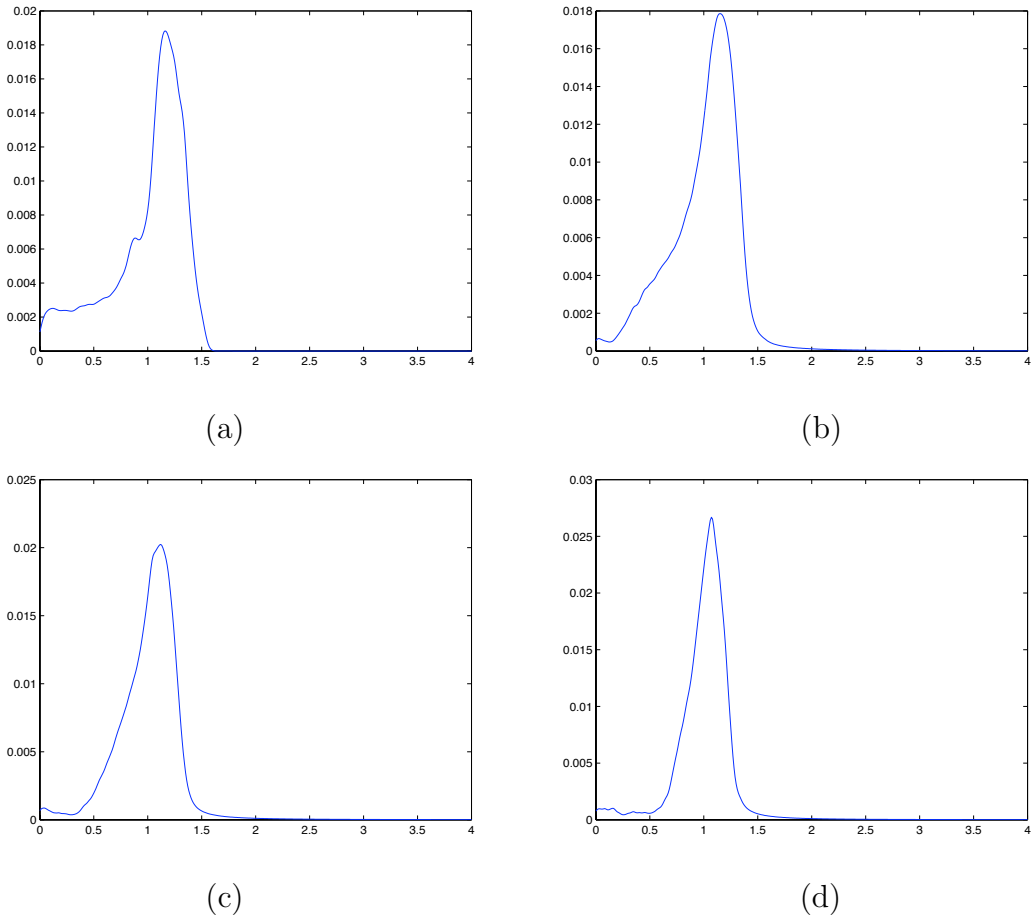


Figure 9.9: Spectral plots of power-grid network of the Western States of the United States (Watts & Strogatz, 1998). Size of the original network is 4941. Data downloaded from http://cdg.columbia.edu/uploads/datasets [Download date: 1 March, 2007.]. (a) Original network. (b) After one iteration. (c) After two iterations. (d) Afther three iterations. All plots are with Gaussian kernel, $\sigma = 0.03$.

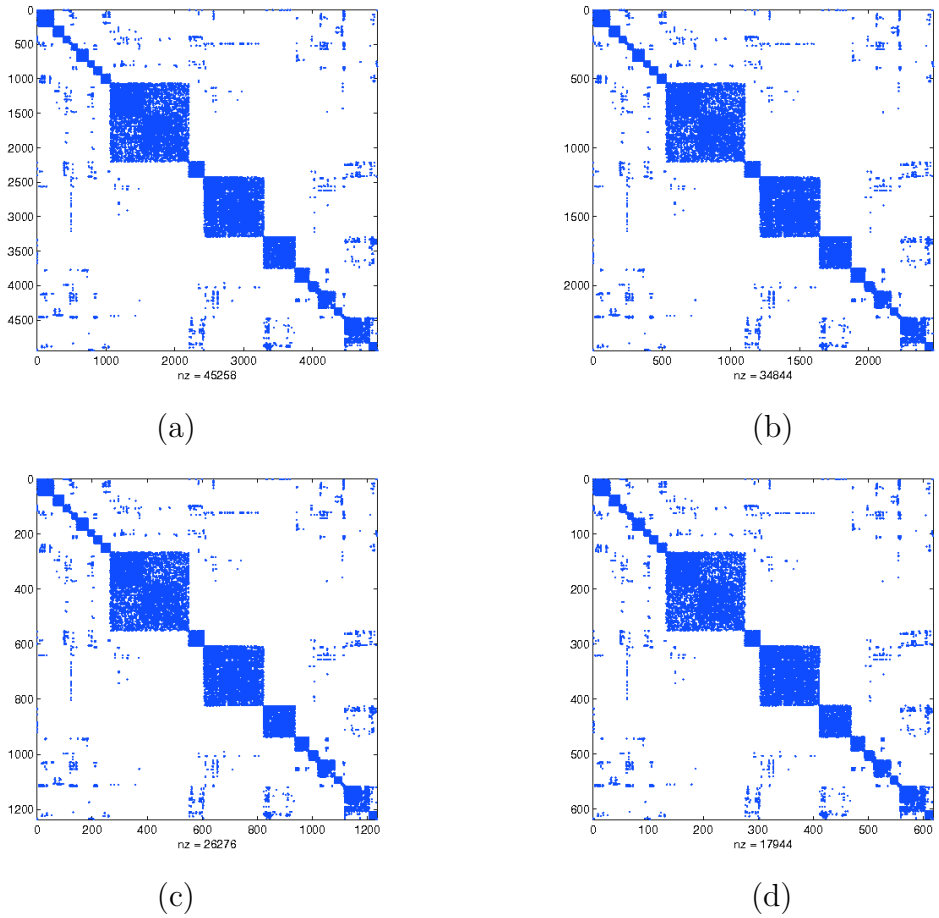


Figure 9.10: Sparsity pattern of the connectivity matrix of power-grid network in Fig. 9.9. (a) Original network. (b) After one iteration. (c) After two iterations. (d) After three iterations. 'nz' is the number of nonzero elements in the matrix.

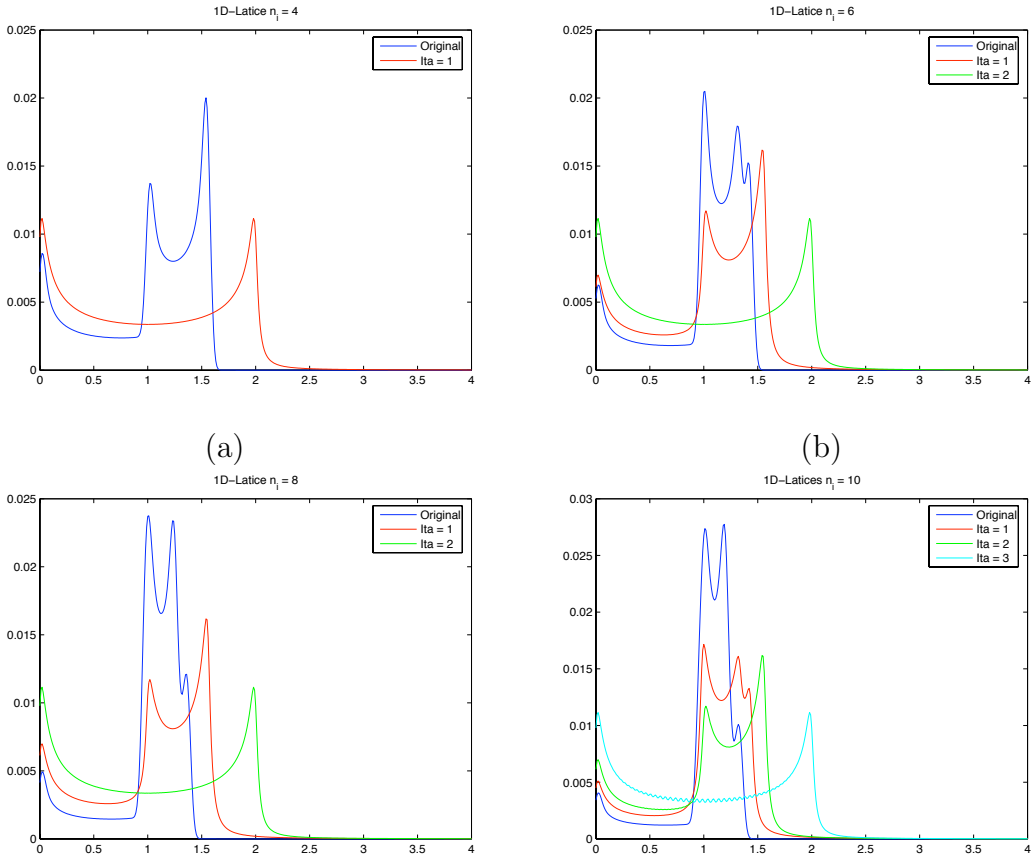
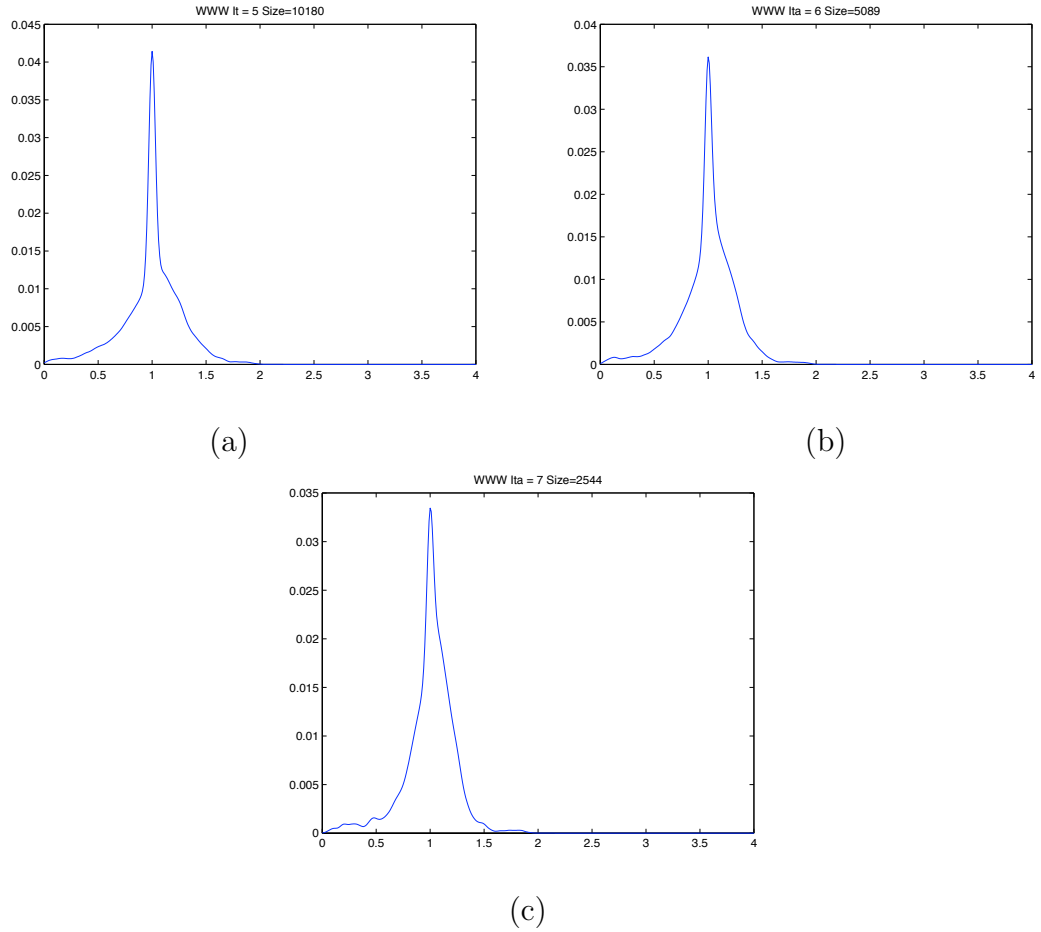


Figure 9.11: Spectral plots of 1D-regular ring lattice (a) with degree of each node = 4. (b) with degree of each node = 6. (c) with degree of each node = 8. (d) with degree of each node = 10. 'Original' corresponds to the plot of original network. 'Ita' shows plots after a particular number of iterations. All plots are with Gaussian kernel, $\sigma = 0.03$.



(a) (b)

(c)

Figure 9.12: Spectral plots WWW netork. Size of the original graph is 325729. Data obtained from (Albert et al., 1999). Data Source: http://www.nd.edu/∼networks [Download date: 17 Feb. 2004]. (a) Network after 5 iterations. (b) Network after 6 iterations. (c) Network after 7 iterations. All plots are with Gaussian kernel, $\sigma = 0.03$.

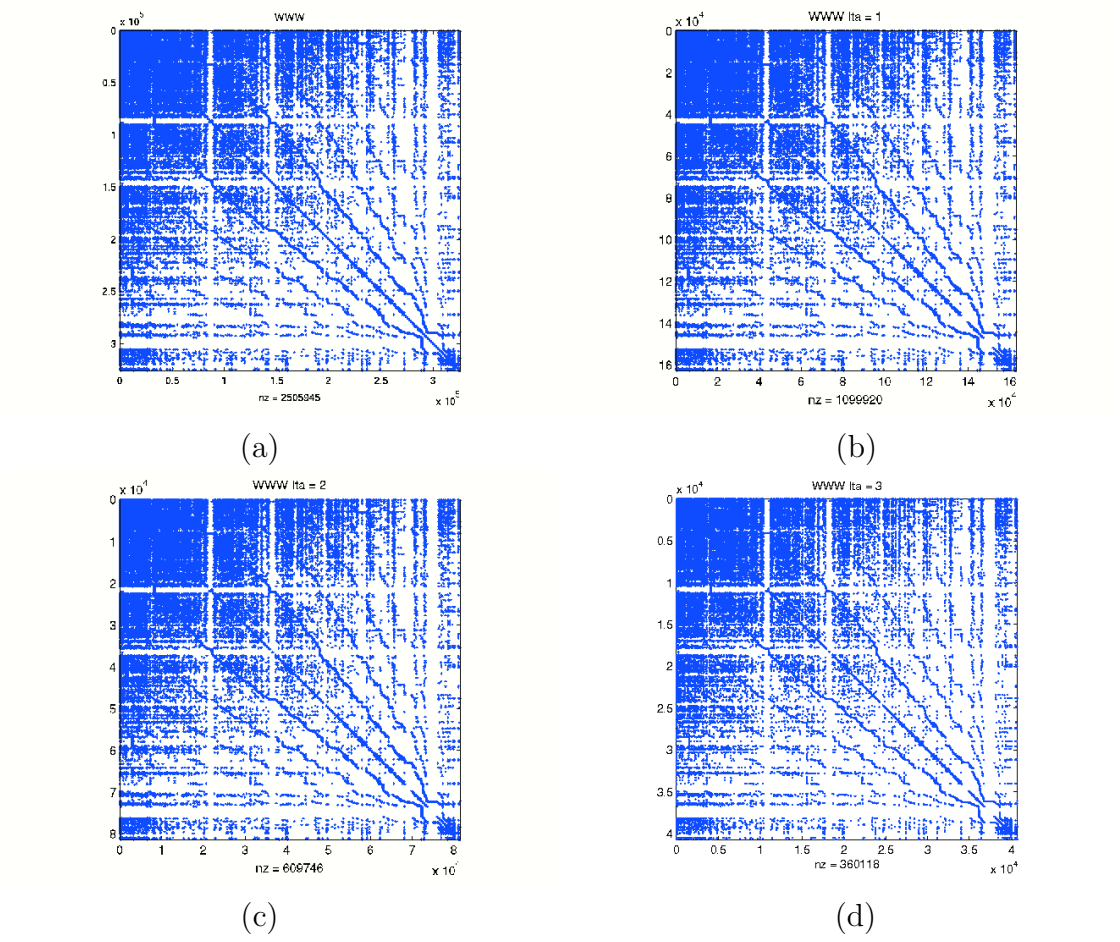


Figure 9.13: Sparsity pattern of the connectivity matrix of WWW network in Fig. 9.12. (a) Original network. (b) After one iteration. (c) After two iterations. (d) After three iterations. 'nz' is the number of nonzero elements in the matrix.

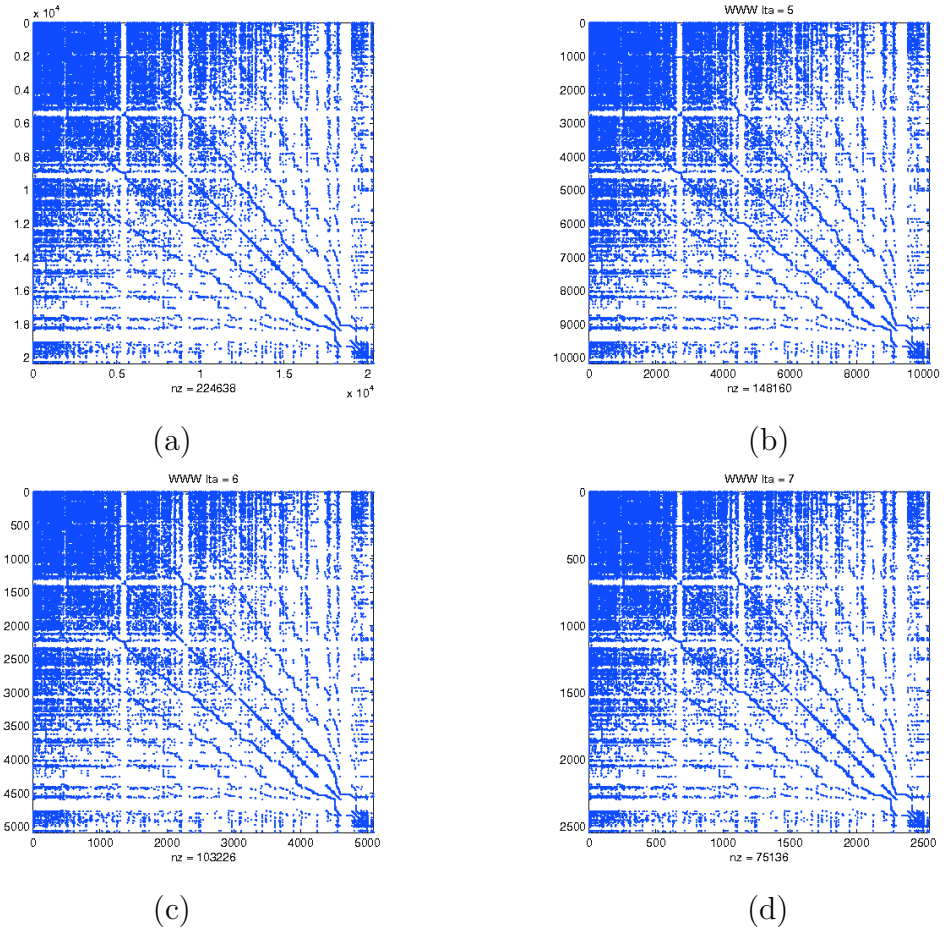


Figure 9.14: Sparsity pattern of the connectivity matrix of WWW network in Fig. 9.12. (a) After four iterations. (b) After five iterations. (c) After six iterations. (d) After seven iterations. 'nz' is the number of nonzero elements in the matrix.

Chapter 10

Conclusions

In this final chapter, we summarize the results and applications presented in the thesis and give an overview of some interesting open problems connected to the discussions of the previous chapters, as well as possible directions for future research.

10.1 Summary

In this thesis, we have proposed the spectrum of the normalized graph Laplacian, as an excellent tool for analyzing network structure. We have explored the information about different topological properties of a graph carried by the complete spectrum. We have investigated how and why structural properties are reflected by the spectra and how spectra change under different graph operations. We have develop a theoretical scheme and applied a general method, based on the spectral plot, that is easily visually analyzed and serves as an excellent diagnostic to categorize networks from different sources. We have presented a scheme for the rough classification of empirical networks based on qualitative global properties detected through the spectrum of the Laplacian of the graph underlying the network. Constructions with different graph operations related to the evolution of a network produce specific eigenvalues and describe certain processes of graph formation that leave characteristic traces in the spectrum. We have shown how useful and plausible hypotheses about evolutionary process can be made by investigating the spectrum of a graph constructed from actual data. Based on this idea, we have reconstructed protein-protein interaction networks based on their spectra. We infered that the spectral distribution is a complete qualitative characterization of a graph. At the end, we also proposed a tentative scheme for reducing the size of large graph while keeping the basic pattern of the spectral plot the same. Now we

briefly summarize our work chapter wise.

In Chapter 5, in order to get an impression about how the patterns of spectral plots differ among different networks, we have presented spectral plots of different networks.

- We have discussed the different ways to produce the spectral plot.
- With many examples, we have shown how this simple visualization method is efficient in capturing the qualitative structural properties of a graph.

In Chapter 6, we have introduced new results which help to explain how the spectrum changes with respect to different structures of graphs and under different graph operations.

- Since the multiplicity of the eigenvalue 1 plays an important role (many spectral plots of real networks have a very sharp peak around 1), we have explained the situations that produce a eigenvalue 1 and proved how this multiplicity changes according to different evolutionary graph operations, like motif duplication, graph joining and splitting.
- We have explained edge duplication, which can produce high but not-too-sharp peak around 1 in the spectral plot. Conformation helps in understanding the underlying structures (like triangle joining) that produce the eigenvalues 3/2 and 1/2, which are also remarkable in the spectrum.

In Chapter 7, using the reconstruction scheme based on the spectrum, we have reproduced protein-protein interaction networks.

- We have shown how one can identify the duplication and divergence processes in their evolutionary history from the spectrum.
- We have also identified typical specific features that robustly distinguish protein-protein interaction networks from other classes of networks, in spite of possible statistical fluctuations in the underlying data.
- We have shown how protein duplication increases the number of a particular type of motif structure.

In Chapter 8, we have introduced a qualitative classification scheme for real networks which uses the properties of the spectral plot.

- We have shown that networks constructed from three generic models – a Erdős-Rényi random network, a Strogatz-Watts small-world network, and a Barabási-Albert scale-free network – have spectral plots that are very different from those of real networks.

- We have roughly divided different real networks into four classes.

In Chapter 9, we have introduced a graph coarsening scheme that reduces the size of a graph, yielding a new reduced graph that produces a spectral plot similar to that of the original graph.

- We have discussed the situations where this scheme fails and the reasons for this.

- We have applied this method to the WWW network to find its rough category class.

10.2 Future research

Finally we propose the following possible directions for future research on the spectrum of the normalized Laplacian.

- In this work, we have mostly emphasized the processes that produce the eigenvalues 1/2, 1, 3/2, which are most common in many real networks. We are interested in exploring different graph operations or conformations that produce other specific eigenvalues and the evolutionary significance behind those processes.

- Spectral plots of many real networks have a small peak around a particular value which does not correspond any eigenvalue of that graph. E.g., spectral plots of different metabolic networks have their second highest peak around .3 or 1.7, which are not eigenvalues of the network. Conformation created by processes like edge duplication can produce the eigenvalues that are very close to 1, but not exactly 1. One can also other seek particular conformations that produce eigenvalues which are very close to .3 or 1.7.

- Our method of reconstruction of a graph based on the spectrum has been successfully applied to protein-protein interaction networks. We can carried out similar studies to explore the evolutionary processes in other networks.

- Compared to Erdős-Rényi's random graph, protein-protein interaction networks are closer to a bipartite graph, but have higher clustering coefficient than these random graphs. It could be interesting to intensively study the linking dynamics of the evolutionary processes of these networks.

- We can make our graph coarsening scheme more efficient by introducing good indexing method. We can also find other appropriate renormalization grouping procedures to get a more accurate result.

- All of our analysis has been carried out for unweighted and undirected graphs. We are interested in extending our study to any generalized graph. But since the in-degree or out-degree of a vertex in a directed sparse graph can be zero, we need to carefully choose the normalization factor of the graph Laplacian.

- From the application of the nodal domain theorem on a graph, it seems that there could be some relation between the multiplicity of a certain eigenvalue and the number occurrences of a specific motif in the graph.

References

Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 us election. In *in proceedings of the www-2005 workshop on the weblogging ecosystem (2005).*

Aiello, W., Chung, F. R. K., & Lu, L. (2000). A random graph model for massive graphs. In *in proceedings of the 32nd annual acm symposium on theory of computing* (p. 171-180).

Aiello, W., Chung, F. R. K., & Lu, L. (2001). Random evolution in massive graphs. In *in proceedings of the 42nd ieee symposium on foundations of computer science* (p. 510-519).

Albert, R., Jeong, H., & Barabási, A. L. (1999). Internet - diameter of the world-wide web. *Nature*, *401*(6749), 130-131.

Alon, U., Surette, M. G., Barkai, N., & Leibler, S. (1999). Robustness in bacterial chemotaxis. *Nature*, *397*(6715), 168-171.

Amaral, L. A. N., & Ottino, J. M. (2004). Complex networks - augmenting the framework for the study of complex systems. *European Physical Journal B*, *38*(2), 147-162.

A.Pikovsky, M.Rosenblum, & J.Kurths. (2001). *Synchronization: A universal concept in nonlinear science.* Cambridge: Cambridge University Press.

Atay, F., T.Bıyıkoğlu, & J.Jost. *Network synchronization: Spectral versus statistical properties.* (Phys.D, to appear)

Atay, F. M., & Biyikoglu, T. (2005). Graph operations and synchronization of complex networks. *Physical Review E*, *72*(1), Art.

Atay, F. M., Biyikoglu, T., & Jost, J. (2006). Synchronization of networks with prescribed degree distributions. *IEEE Transactions On Circuits and Systems I-Regular Papers*, *53*(1), 92-98.

Atay, F. M., Jost, J., & Wende, A. (2004). Delays, connection topology, and synchronization of coupled chaotic maps. *Physical Review Letters*, *92*(14), Art.

Banerjee, A., & J.Jost. (a). *Graph spectra as a systematic tool in computational*

biology. (Discrete Applied Mathematics, to appear. Preprint. Available at arXiv:0706.0113v1)

Banerjee, A., & J.Jost. (b). *Laplacian spectrum and protein-protein interaction networks.* (Preprint. Available at arXiv:0705.3373v1.)

Banerjee, A., & J.Jost. (c). *On the spectrum of the normalized graph laplacian.* (Preprint. Available at arXiv:0705.3772v1)

Banerjee, A., & J.Jost. (d). *Spectral plot properties: Towards a qualitative classification of networks. preprint.* (Preprint.)

Banerjee, A., & J.Jost. (2007). Spectral plots and the representation and interpretation of biological data. *Theory in Biosciences*, *126*(1), 15-21.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509-512.

Barabási, A. L., Albert, R., & Jeong, H. (1999). Mean-field theory for scale-free random networks. *Phys. A*, *272*, 173-187.

Barabási, A. L., Albert, R., & Jeong, H. (2000). Scale-free characteristics of random networks: The topology of the world wide web. *Phys. A*, *281*, 69-77.

Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica a*, *311*(3-4), 590-614.

Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology*, *110*(1), 44-91.

Berg, J., Lassig, M., & Wagner, A. (2004). Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *Bmc Evolutionary Biology*, *4:*(4: Art. No. 51 Nov 27 2004), Art.

Bevis, J. H., Blount, K. K., Davis, G. J., Domke, G. S., & Miller, V. A. (1997). The rank of a graph after vertex addition. *Linear Algebra and Its Applications*, *265:*(265: 55-69 Nov 1 1997), 55-69.

Biggs, N. L. (1993). *Algebraic graph theory* (2nd ed.). Chambridge: Chambridge University Press.

Biggs, N. L., Lloyd, E. K., & Wilson, R. J. (1976). *Graph theory 1736-1936.* Oxford: Clarendon Press.

Bollobás, A. B. (1985). *Random graphs.* London: Academic Press.

Bollobás, A. B. (1998). *Modern graph theory.* New York: Springer.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., & Wiener, J. (2000). Graph structure in the web. *Computer Networks*, *33*, 309-320.

Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W. H., Yang, X. C., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M.,

Parkinson, J., Greenblatt, J., & Emili, A. (2005). Interaction network containing conserved and essential protein complexes in escherichia coli. *Nature*, *433*(7025), 531-537.

Caldarelli, G., Erzan, A., & Vespignani, A. (2004). Preface on "applications of networks". *European Physical Journal B*, *38*(2), 141.

Cancho, R. F. i, & Solé, R. V. (2001). *The small-world of human language.* Santa Fe Working Paper. (Paper Number 01-03-016)

Castro, R. D., & Grossman, J. W. (1999). Famous trails to paul erdos. *Mathematical Intelligencer*, *21*(3), 51-63.

Chen, G., Davis, G., Hall, F., Li, Z., Patel, K., & Stewart, M. (2005). An interlacing result on normalized laplacians. *Journal on Discrete Mathematics*, *18*(2), 353-361.

Chen, Q., Chang, H., Govindan, R., Jamin, S., Shenker, S. J., & Willinger, W. (2002). The origin of power laws in internet topologies revisited. In *in proceedings of the 21st annual joint conference of the ieee computer and communications societies.* Los Alamitos: IEEE Computer Society.

Chung, F., Lu, L. Y., & Vu, V. (2003). Spectra of random graphs with given expected degrees. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(11), 6313-6318.

Chung, F. R. (1997). *Spectral graph theory.* Providence, RI: American Mathematical Society.

Chung, F. R., & Lu, L. (2006). *Complex graphs and networks.* Providence, RI: American Mathematical Society.

Collatz, L., & Sinogowitz, U. (1957). Spektren endlicher grafen. *Abh. Math. Sem. Univ. Hamburg*, *21*, 63-77.

Cvetković, D. M., Doob, M., Gutman, I., & Torgašev, A. (1988). *Recent results in the theory of graph spectra.* Amsterdam: North Holland.

Cvetković, D. M., Doob, M., & Sachs, H. (1980). *Spectra of graphs, theory and applications.* Academic Press.

Cvetković, D. M., Doob, M., & Sachs, H. (1995). *Spectra of graphs theory and applications.* Heidelberg ; Leipzig Barth: Johann Ambrosius Barth Verlag.

De, P., Singh, A. E., Wong, T., Yacoub, W., & Jolly, A. M. (2004). Sexual network analysis of a gonorrhoea outbreak. *Sex Transm Infect*, *80*, 280-285.

Dorogovtsev, S. N., & Mendes, J. F. F. (2001). Language as an evolving word web. *Proceedings of the Royal Society of London Series B-Biological Sciences*, *268*(1485), 2603-2606.

Ellingham, M. (1993). Basic subgraphs and graph spectra. *Australasian J. Combinatorics*, *8*, 247-265.

Erdős, P., & Réanyi, A. (1959). On random graphs. *Publicationes Mathematicae*

Debrecen, *6*, 290-297.

Erdős, P., & Réanyi, A. (1960). On the evolution of random graphs. *Bulletin of the International Statistical Institute*, *38*(4), 343-347.

Erdős, P., & Réanyi, A. (1961). On the strength of connectedness of a random graph. *Acta Math. Acad. Sci.*, *12*, 261-267.

Euler, L. (1736). *Solution problematis ad geometriam situs pertinentis.*

Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). Applications, technologies, architectures, and protocols for computer communication. *ACM SIGCOMM Computer Communication Review (CCR)*, *29*, 251-262.

Farkas, I. J., Derenyi, I., Barabási, A. L., & Vicsek, T. (2001). Spectra of "real-world" graphs: Beyond the semicircle law. *Physical Review E*, *64*(2), Art.

Füredi, Z., & Komlós, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica*, *1*(3), 333-341.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, *99*, 7821-7826.

Godsil, C., & Royle, G. (2001). *Algebraic graph theory.* New York: Springer.

Goh, K. I., Kahng, B., & Kim, D. (2001). Spectra and eigenvectors of scale-free networks. *Physical Review E*, *64*(5), Art.

Greene, L. H., & Higman, V. A. (2003). Uncovering network systems within protein structures. *Journal of Molecular Biology*, *334*(4), 781-791.

Grossman, J. W., & Ion, P. D. F. (1995). On a portion of the well-known collaboration graph. *Congr. Numer.*, *108*, 129-131.

Guare, J. (1990). *Six degrees of separation: A play.* New York: Vintage Books.

Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, F., & Arenas, A. (2003). Self-similar community structure in a network of human interactions. *Physical Review E*, *68*(6), Art. No. 065103.

Guimera, R., Mossa, S., Turtschi, A., & Amaral, L. A. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(22), 7794-7799.

Huberman, B. A., & Adamic, L. A. (1999). Internet - growth dynamics of the world-wide web. *Nature*, *401*(6749), 131-131.

Hückel, E. (1931). Quantentheoretische beiträge zum benzolproblem. *Z. Phys.*, *70*, 204-286.

Huxham, M., Beaney, S., & Raffaelli, D. (1996). Do parasites reduce the chances of triangulation in a real food web? *Oikos*, *76*, 284-300.

Huynen, M. A., & Bork, P. (1998). Measuring genome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(11), 5849-5856.

Ipsen, M., & Mikhailov, A. S. (2002). Evolutionary reconstruction of networks. *Physical Review E*, *66*(4), Art.

Ispolatov, I., Krapivsky, P. L., & Yuryev, A. (2005). Duplication-divergence model of protein interaction network. *Physical Review E*, *71*(6), Art.

Jeong, H., Mason, S. P., Barabási, A. L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41-42.

Jeong, H., Tombor, B., Albert, R., Oltval, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*(6804), 651-654.

Jost, J. (2007). Formal aspects of networks: what perspectives can mathematics offer to biology and neurobiology. In J.F.Feng, J.Jost, & M.P.Qian (Eds.), *Networks: from biology to theory* (p. 37). Springer.

Jost, J. (to appear). *Mathematical methods in biology and neurobiology.* (monograph, to appear)

Jost, J., & Joy, M. P. (2001). Spectral properties and synchronization in coupled map lattices. *Physical Review E*, *65*(1), 016201.

Jost, J., & Joy, M. P. (2002). Evolving networks with distance preferences. *Physical Review E*, *66*, 36126-36132.

Kim, J., Krapivsky, P. L., Kahng, B., & Redner, S. (2002). Infinite-order percolation and giant fluctuations in a protein interaction network. *Physical Review E*, *66*(5), Art.

Koch, C., & Laurent, G. (1999). Complexity and the nervous system. *Science*, *284*(5411), 96-98.

Kretzschmar, M. (2000). Sexual network structure and sexually transmitted disease prevention: a modeling perspective. *Sexually transmitted diseases*, *27*, 627-635.

Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. (1999). Trawling the web for emerging cyber-communities. *Computer Networks*, *31*, 1481-1493.

Lago-Fernandez, L. F., Huerta, R., Corbacho, F., & Siguenza, J. A. (2000). Fast response and temporal coherent oscillations in small-world networks. *Physical Review Letters*, *84*(12), 2758-2761.

Lawrence, S., & Giles, C. L. (1998). Searching the world wide web. *Science*, *280*(5360), 98-100.

Lawrence, S., & Giles, C. L. (1999). Accessibility of information on the web. *Nature*, *400*(6740), 107-109.

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., & Aberg, Y. (2001). The web of human sexual contacts. *Nature*, *411*(6840), 907-908.

Lubotzky, A., Phillips, R., & Sarnak, P. (1988). Ramanujan grphs. *Combinatorica*, *8*, 261-278.

Margulis, G. A. (1984). Arithmetic groups and graphs without short cycles. In *6th internat. symp. on information theory, tashkent.*

Martinez, N. D. (1991). Artifacts or attributes? effects of resolution on the little rock lake food web. *Ecological Monographs*, *61*, 367-392.

Merris, R. (1994). Laplacian matrices of graphs - a survey. *Linear Algebra and Its Applications*, *198:*, 143-176.

Merris, R. (2001). *Graph theory.* New York: John Wiley.

Mihail, M., & Papadimitriou, C. H. (2002). On the eigenvalue power law. In *Random '02: Proceedings of the 6th international workshop on randomization and approximation techniques* (pp. 254–262). London, UK: Springer-Verlag.

Milgram, S. (1967). Small-world problem. *Psychology Today*, *1*(1), 61-67.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, *298*(5594), 824-827.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2004). Network motifs: Simple building blocks of complex networks. *Science*, *303*, 1538-1542.

Mohar, B. (1997). Some applications of laplace eigenvalues of graphs. In G. Hahn & G. Sabidussi (Eds.), *Graph symmetry: Algebraic methods and applications* (p. 227-277). Springer.

Molloy, M., & Reed, B. (1995). A critical-point for random graphs with a given degree sequence. *Random Structures & Algorithms*, *6*(2-3), 161-179.

Newman, M. E. J. (2001a). Scientific collaboration networks. i. network construction and fundamental results. *Physical Review E*, *64*(1), Art. No. 016131.

Newman, M. E. J. (2001b). Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical Review E*, *64*(1), Art. No. 016132.

Newman, M. E. J. (2001c). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 404-409.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167-256.

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, *101:*(101: 5200-5205 Suppl. 1, Apr 6 2004), 5200-5205.

Newman, M. E. J. (2005). Power laws, pareto distributions and zipf's law. *Contemporary Physics*, *46*, 323-351.

Newman, M. E. J., Strogatz, S. H., & Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*,

64(2), Art.

Oltvai, Z. N., & Barabási, A. L. (2002). Life's complexity pyramid. *Science*, *298*(5594), 763-764.

Pastor-Satorras, R., Smith, E., & Sole, R. V. (2003). Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology*, *222*(2), 199-210.

Pecora, L. M., & Carroll, T. L. (1990). Synchronization in chaotic systems. *Physical Review Letters*, *64*(8), 821-824.

Pimm, S. L., Lawton, J. H., & Cohen, J. E. (1991). Food web patterns and their consequences. *Nature*, *350*(6320), 669-674.

Price, D. J. D. (1965). Networks of scientific papers. *Science*, *149*(3683), 510-515.

Price, D. J. D. (1976). General theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*(5-6), 292-306.

Rao, F., & Caflsich, A. (2004). The protein folding network. *Journal of Molecular Biology*, *342*(1), 299-306.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barbasi, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science*, *297*(5586), 1551-1555.

Redner, S. (1998). How popular is your paper? an empirical study of the citation distribution. *European Physical Journal B*, *4*(2), 131-134.

Scala, A., Amaral, L. A. N., & Barthelemy, M. (2001). Small-world networks and the conformation space of a short lattice polymer chain. *Europhysics Letters*, *55*(4), 594-600.

Seidel, J. J. (1989). Graphs and their spectra. In Warsaw (Ed.), *Combinatorics and graph* (p. 147-162). PWN-Polish Scientific Publishers.

Shen-Orr, S. S., Milo, R., Mangan, S., & Alon, U. (2002). Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetics*, *31*(1), 64-68.

Simon, H. (1955). On a class of skew distribution functions. *Biometrika*, *42*, 425-440.

Solomonoff, R., & Rapoport, A. (1951). Connectivity of random nets. *Bulletin of Mathematical Biophysics*, *13*, 107-117.

Song, C., Havlin, S., & Makse, A. H. (2005). Self-similarity of complex networks. *Nature*, *433*(27), 392-395.

T.Bıyıkoğlu, J.Leydold, & P.Stadler. (2007). *Laplacian eigenvectors of graphs.* (to appear)

Vázquez, A., Flammini, A., Maritan, A., & Vespignani, A. (2003). Modeling of protein interaction networks. *ComplexUs*, *1*, 38-44.

Vukadinovic, D., Huang, P., & Erlebach, T. (2002). On the spectrum and structure of internet topology graphs. *Innovative Internet Computing Systems Lecture Notes in Computer Scienc*, *2346:*(2346: 83-95 2002), 83-95.

Wagner, A. (2001). The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Molecular Biology and Evolution*, *18*(7), 1283-1292.

Wagner, A. (2003). How the global structure of protein interaction networks evolves. *Proceedings of the Royal Society B: Biological Sciences*, *270*, 457-466.

Wagner, A., & Fell, D. A. (2001). The small world inside large metabolic networks. *Proceedings of the Royal Society of London Series B-Biological Sciences*, *268*(1478), 1803-1810.

Watts, D. J. (1999a). Networks, dynamics, and the small world phenomenon. *Amer. J. Sociol*, *105*, 493-592.

Watts, D. J. (1999b). *Small worlds*. Princeton, NJ: Princeton University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440-442.

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The structure of the nervous-system of the nematode caenorhabditis-elegans. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *314*(1165), 1-340.

Wigner, E. P. (1955). Characteristics vectors of bordered matrices with infinite dimension. *The Annals of Mathematics*, *62*(3), 548-564.

Wigner, E. P. (1957). Characteristics vectors of bordered matrices with infinite dimension ii. *The Annals of Mathematics*, *65*(2), 203-207.

Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrics. *The Annals of Mathematics*, *67*(2), 325-327.

Zhu, P., & Wilson, R. (2005, September). A study of graph spectra for comparing graphs. In P. T. William Clocksin Andrew Fitzgibbon (Ed.), *Proceedings of the british machine vision conference 2005.* Oxford, U.K.

Daten zum Autor:

Name Anirban Banerjee

geboren am 26. August 1978 in Bankura, Indien

2000-2002 Studium an der Fakultät für Mathematik, Indian Institute of Technology, Kharagpur, Abschluss: Master of Science.

2002-2003 Studium am Advanced Diploma in Bioinformatics, University of Calcutta, Abschluss: Advanced Diploma in Bioinformatics.

05/2004- Doktorand am Max-Planck-Institut für Mathematik in den Naturwissenschaften

Bibiliographische Daten

The Spectrum of the Graph Laplacian as a Tool for Analyzing Structure and Evolution of Networks
(Das Spektrum das Laplaceoperators eines Graphen als Methode zur Analyse der Struktur und Evolution von Netzwerken)
Anirban Banerjee
Universität Leipzig, Dissertation
146 Seiten, 60 Abbildungen, 126 Referenzen

Selbstständigkeitserklärung

Hiermit erkläre ich, die vorliegende Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet.

Leipzig, June 29, 2007

. . . . . .
(Anirban Banerjee)