MODULARITY AND HIERARCHY IN COMPLEX SYSTEMS: RELATING NETWORK STRUCTURE TO DYNAMICS

By

Raj Kumar Pan

The Institute of Mathematical Sciences, Chennai.

A thesis submitted to the
Board of Studies in Physical Sciences

In partial fulfillment of the requirements

For the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

August 2009

Homi Bhabha National Institute

Recommendations of the Viva Voce Board

As members of the Viva Voce Board, we recommend that the dissertation prepared by **Raj Kumar Pan** entitled "Modularity and Hierarchy in Complex Systems: Relating network structure to dynamics" may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

---------- **Date :**
Chairman : R. Jagannathan

---------- **Date :**
Convener : Sitabhra Sinha

---------- **Date :**
Member : Gautam I. Menon

---------- **Date :**
Member : Sudeshna Sinha

---------- **Date :**
Member : C. R. Subramanian

---------- **Date :**
External : Bikas K. Chakrabarti

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to HBNI.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

---------- **Date :**
Guide : Sitabhra Sinha

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and the work has not been submitted earlier as a whole or in part for a degree/diploma at this or any other Institution or University.

Raj Kumar Pan

To my parents

ACKNOWLEDGEMENTS

During my stay at IMSc, I was fortunate to be in the company of some wonderful people who have directly or indirectly contributed to this thesis. First, I would like to thank my advisor, Sitabhra Sinha, whose energy and involvement has made my research experience especially rewarding. It was he who introduced me to the fascinating subject of complex systems and helped me to understand many of the problems. I couldn't imagine an advisor more encouraging, creative and supportive.

I have had the opportunity to discuss many of the research problems that I worked on as part of this thesis with a number of people, both from IMSc and elsewhere. They have generously spent time reading the early drafts of many of my papers and their critical comments were essential in improving my work. Among others, I would like to offer sincere thanks to S. Sridhar, Sudeshna Sinha, Gautam Menon, Subinay Dasgupta, R. Rajesh, Purusattam Ray, Ronojoy Adhikari, Ramesh Anishetty, Deepak Dhar, M. S. Santhanam and Sanjay Jain. I would also like to thank M. Krishna and Anindya Sinha for providing me data on stock price variation in NSE (India) and on social interactions in macaque troupes, respectively.

I also thank my officemates, Mithun Kumar Mitra and Bireswar Das, for our fun time together. Lots of my free time was spent in the enjoyable company of many friends in IMSc and with our football team. I also like to thank all the staff members in the IMSc office for their sincere help and co-operation in academic as well as non-academic matters.

Finally, I would like to thank my parents and my wife, without whose love and support, none of this would have been possible.

Abstract

Complex systems, whether integrated circuits, food webs, transportation networks, social systems, or the biochemical interactome of a living cell, all behave in ways that cannot be fully explained by analyzing their constituent parts in isolation. Understanding the emergent behavior of such nonlinear systems, which is more than just an aggregate of the properties of their components, require novel integrative approaches. Many of these systems can be represented as networks, consisting of a large number of nodes connected via directed or undirected links. The recent discovery of the existence of universal principles underlying these complex networks that occur across widely differing domains in the biological, social and technological arenas have spurred the interest of physicists in trying to understand such principles using techniques from statistical physics and non-linear dynamics. In this thesis we look at how the structure of a network, as characterized by the connection topology, governs its dynamical behavior, and conversely, how the dynamical processes taking place on the network affects its structure (e.g., stability considerations constraining the evolution of the network towards specific topologies). In particular, we focus on modularity, i.e., the existence of groups whose nodes are more densely connected to each other than to nodes in other groups, and hierarchy, i.e., the nested arrangement of connection topology into several layers. Both of these mesoscopic organizational structures are observed in many complex networks that occur in reality.

We begin with a short overview of the physics of complex networks in **Chapter 1**. In the first few sections, we introduce important concepts and definitions that are used throughout the thesis. This is followed by a brief discussion of some of the commonly used network models found in the literature. Next, we analyze a simple model of modular random networks in **Chapter 2** and show that it has structural properties similar to many real-world networks. We also examine the effect of modular structure on dynamics occurring over the network by studying the phenomena of synchronization, diffusion and spin-ordering on the network model. We show that for all these different varieties of dynamical processes modularity gives rise to the same characteristic signature of multiple distinct time scales. In **Chapter 3**, we explore how modularity can arise in networks as an outcome of evolution in the presence of multiple co-existing constraints. As an example, we

demonstrate the emergence of modular organization upon simultaneous optimization of several structural and dynamical constraints to which many real networks are subject. The minimal number of such constraints is found to be three, e.g., minimizing (a) average path length, (b) total number of links, and (c) probability of local instability. The well-known connection topologies of star, chain and random networks appear as limiting cases when one of these constraints is relaxed. In **Chapter 4**, we introduce a model for hierarchical random networks and analyze the effect of having multiple structural levels (or hierarchies) on different dynamical processes. In general, making the previously introduced modular network more hierarchical, has effects similar to increasing the number of structural modules. We show that a generalization of the evolutionary model for modularity introduced in the preceding chapter can also give rise to simple hierarchical ordering in a network. In **Chapter 5**, we show that the modular structure of networks occurring in reality can be uncovered from empirical data, namely the dynamical time-series obtained from the component nodes. In particular, we have analyzed the personal ties between individuals in primate troops, in terms of allogrooming behavior, in order to deduce their social organization. Next, we reconstruct the network of interactions among stocks in the Indian financial market by using spectral techniques on the cross-correlations in their price variations. We identify modules corresponding to groups of strongly interacting stocks. Using a multi-factor model, we show that the emergence of such structures is an outcome of relatively stronger mutual interactions between nodes belonging to the same business sector, as compared to their susceptibility to common signals that affect the entire market. In **Chapter 6** we look at the reverse problem of how network structure reflects the dynamics or function of the system. We analyze the somatic neuronal network of the nematode *C. elegans*, the only organism whose entire nervous system has been completely mapped. We demonstrate that the network structure cannot be fully explained on the basis of exclusively structural considerations, e.g., minimization of wiring cost or maximization of communication efficiency, and give examples of how the functional role of the system as an information processor in a noisy environment can result in specific structural patterns. Finally, in **Chapter 7**, we conclude with a general overview of our results on how the dynamics occurring on a network is governed by specific structural features of the system, and in turn, affects the evolution of ubiquitous structural patterns such as modularity and hierarchy.

Contents

List of Figures

List of Tables

List of Publications

- S. Sinha and R.K. Pan, **The Power (Law) of Indian Markets: Analysing NSE and BSE trading statistics** in *Econophysics of Stock and Other Markets* (Eds. A. Chatterjee and B.K. Chakrabarti), Springer-Verlag Italia, Milan (2006) pp. 24-34.

- S. Sinha and R.K. Pan, **Uncovering the Internal Structure of the Indian Financial Market: Cross-correlation behavior in the NSE** in *Econophysics of Markets and Business Networks* (Eds. A. Chatterjee and B.K. Chakrabarti), Springer-Verlag Italia, Milan (2007) pp. 3-20.

- R.K. Pan and S. Sinha, **Self-organization of price fluctuation distribution in evolving markets**, *Europhys. Lett.* **77**, 58004 (2007).

- R.K. Pan and S. Sinha, **Collective behavior of stock price movements in an emerging market**, *Phys. Rev. E* **76**, 046116 (2007).

- R.K. Pan and S. Sinha, **Modular networks emerge from multi-constraint optimization**, *Phys. Rev. E* **76**, 045103(R) (2007).

- R.K. Pan and S. Sinha, **Inverse cubic law of index fluctuation distribution in Indian markets** *Physica A* **387**, 2055-2065 (2008).

- R.K. Pan and S. Sinha, **Modular networks with hierarchical organization: The dynamical implications of complex structure**, *Pramana - journal of physics* **71**, 331-340 (2008).

- R.K. Pan and S. Sinha, **Modularity produces small-world networks with dynamical time-scale separation**, *Europhys. Lett.* **85**, 68006 (2009).

- S. Dasgupta, R.K. Pan and S. Sinha, **Phase of Ising spins on modular networks analogous to social polarization**, accepted for publication in *Phys. Rev. E*, (arXiv:0903.4529).

- R.K. Pan, N. Chatterjee and S. Sinha, **Mesoscopic organization reveals the constraints governing C. elegans nervous system**, *arXiv:0905.3887*.

1 Introduction

In recent years, there has been a growing interest in the study of complex systems which pervade all of science, from cell biology to ecology, and from computer science to sociology [1, 2]. Instead of being a simple aggregation of a limited set of linearly interacting units, most real world systems are made up of large number of components, or agents, which interact in such a way that their collective behavior is not a simple combination of the individual properties of their components. Such emergence of system-level features is one of the characteristic indicators of a complex system. Moreover, these systems are often seen to be self-organized as a result of mutual interactions between their components rather than being ordered by some external agency. Consequently, understanding the behavior of such a system requires integrative approaches. One must understand not only the behavior of its parts, but also, how they act together to give rise to the collective behavior of the whole. One way of describing complex systems is modeling them mathematically by using the framework of networks. In this approach, one focuses only on the topology of interactions between the elements, providing a systems-level perspective to the example under study [3, 4]. Such an analysis helps to reveal the universal principles underlying their organization and function, despite the great complexity and variety of these systems [5, 6].

In this chapter, we introduced the relevant concepts and definitions from network science that we shall be using throughout the thesis. In Sec. 1.1, we first define a set of terms to describe the complex networks and models to study such systems. In Sec. 1.2, we give a brief overview of several types of complex networks seen in real-world. In Sec. 1.3, the measurable properties of networks, such as,

path length, clustering and degree distribution, are introduced. This is followed by Sec. 1.4, in which we discuss models used for describing and analyzing complex networks that are already extant in the literature. Finally, in Sec. 1.6, we present an overview of the thesis, and the principal results of each chapter.

1.1 Complex networks

The different components or interacting units of a complex system when described as a network, are represented by nodes or vertices, and the interactions or connections between the units are represented by edges or links between pairs of nodes. Such networks provide a concise mathematical representation of the topology of interactions between the components. Thus, understanding how social, biological, and economic systems work may often depend partially on understanding their patterns of interactions, i.e., the underlying networks. The graph theoretical framework has provided the potential synergies among researchers across different multidisciplinary fields to come and work together to solve apparently unrelated problems.

Network architecture may have important functional consequences for the whole system. For example, the topology of the network controls the rate at which information or diseases propagate through it [7, 8], its robustness under attack or failure of individual components [9], as well as, adaptation and learning processes on it [10]. Recent work has pointed out the crucial role played by the network structure in determining the emergence of collective dynamical behavior, such as, synchronization of nodal activity. Hence, studying these patterns of interactions between the components of a complex system can lead to a better understanding of its dynamical and functional behavior, in addition to throwing light on the evolutionary mechanism leading to it.

There are several reasons for the emergence and rapid development of this field. Many of the insights and advances in this field are due to the recent availability of large quantities of high resolution data from different systems. Obtaining such empirical data has become possible because of technological advances. For example, the network of social interactions among individuals can be constructed from information about the calls they make using their mobile phones [11], leading to

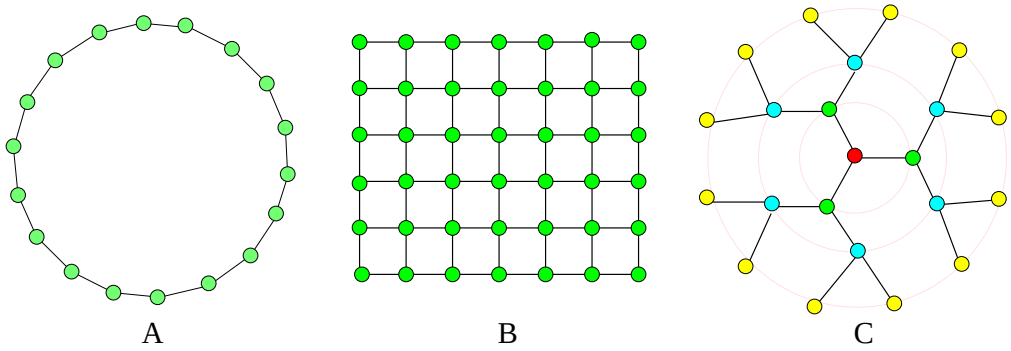


Figure 1.1: Representations of regular graph models: Nodes on (A) a 1-dimensional lattice, (B) a 2-dimensional lattice, and, (C) a Bethe lattice or Cayley tree with nearest neighbor connections.

better understanding of human social dynamics. There has also been remarkable increase in computational power, using which regularities and patterns in large data-sets can be determined. Another reason for the involvement of a large number of physicists in this field is that, statistical physics and non-linear dynamics can be used to develop methods and techniques for analyzing and modeling complex networks [12, 13].

The traditional approach in physics for describing an interacting system is to use a lattice embedded in d-dimensional space. Each elements of the system is located on a lattice site and interacts with neighboring sites within a range r $(= 1, 2, \cdots)$. On such a regular network, all nodes have the same number of neighbors $(= (2r)^d)$, where r is the range of interaction and d is the dimension of the space on which the lattice is embedded [Fig. 1.1]. Another commonly used graph in the literature is the Cayley tree or Bathe lattice, where each node has the same number of neighbors but there are no cycles in the structure. At the other extreme, we have the homogeneous random graph (also referred to as Erdos-Renyi or ER graphs) where the edges between any pair of nodes are randomly placed with probability p. However, networks occurring in the world around us have structures that occur between these extremes and have properties that often differ remarkably from both regular and random graphs [Fig. 1.2].

1.2 Examples of complex networks

The empirical data for connectivity in real-world complex networks span several disciplines. For the purpose of summarization, we loosely divide them into four categories: biological networks, social networks, technological networks and financial networks.

Biological networks

A number of biological systems can be usefully represented as networks. Examples of such networks occur at many different length scales. At the molecular level, protein structure can be considered as a network where the residue atoms of two amino acids are said to be connected if the Euclidean distance between them is less than a threshold, so that there is a significant van der Waals interaction between them [14]. On a slightly large length scale, we have the example of intra-cellular signaling networks that allow extra-cellular stimuli at cell surface receptors to be relayed to the nucleus by a sequence of enzyme catalyzed reactions [15]. Such a network allows the cell to respond to specific stimuli with appropriate actions, such as cell division, apoptosis, etc. Another example of an intra-cellular network that is defined in terms of the existence of physical contacts between the constituents of the network (rather than functional relations), is the protein interaction network [16, 17]. At the inter-cellular scale, the most prominent example is that of neuronal networks, involved in processing information vital to the survival of the organism. Here, the nodes are neurons, and the links are electrical (gap junction) or chemical (synapse) connections [18]. At even larger length scales, there are ecological networks such as food webs [19], where the links correspond to trophic relations between species (represented by the nodes).

Social networks

Social networks are probably the earliest empirical networks that have been analyzed in detail. Much prior to the recent excitement (starting from 1998) among physicists about networks, sociologists had been constructing networks of social contacts within small groups (such as ties of friendship within a school) in order to understand how and why relations that define a society develop [23]. However,

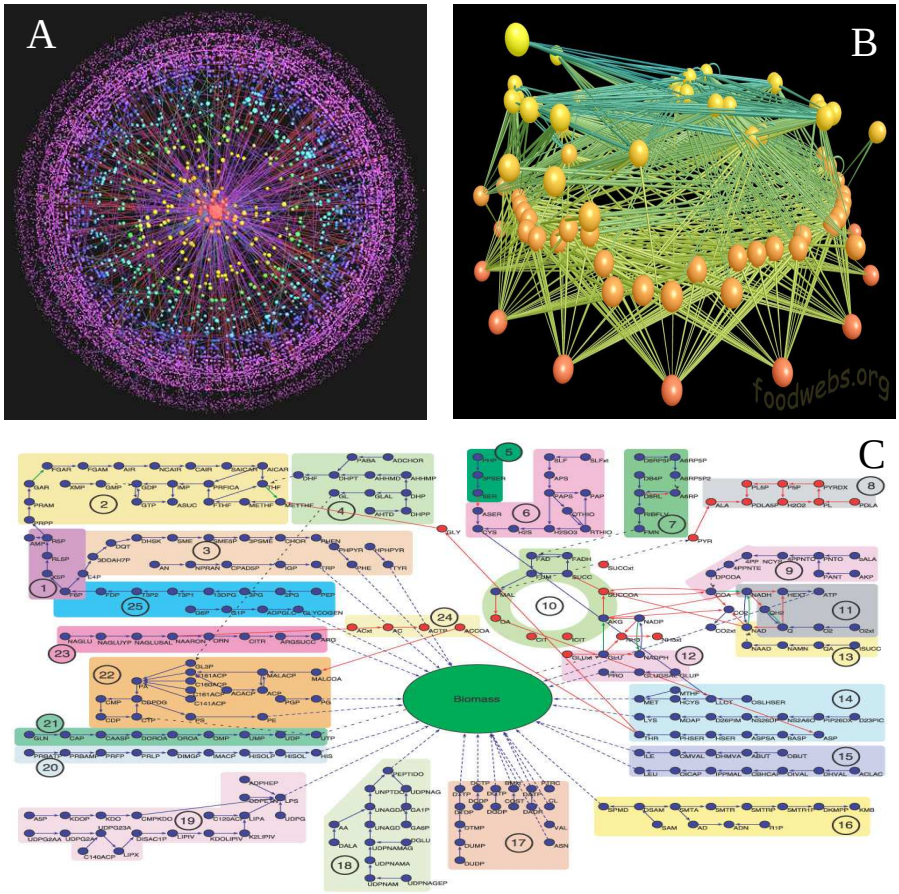


Figure 1.2: Examples of complex networks: (A) Internet at the Autonomous Server level [20], (B) Food web at East River Valley [21] (nodes: species, links: trophic relation), and (C) Metabolic network of *E. coli* [22] (nodes: metabolite, links: reaction).

such studies could not be extended to groups exceeding a few tens of individuals owing to limitations in the way data was collected, e.g., using questionnaires. With the advent of online sites, such as Facebook, Orkut, etc., where information about links between different individuals can be electronically gathered, it is now possible to study contact networks in populations numbering thousands or more. Moreover, such data also allows us to see how patterns in social ties develop over time [24]. Similar detailed analysis has also been done for friendship networks, reconstructed on the basis of frequency and duration of mobile phone conversations between individuals subscribing to the same phone company [11]. The availability of large computer databases have also allowed looking at other relational networks

between people, such as those formed by collaborations between authors of scientific papers [25].

Technological networks: Information transmission & Resource transportation

One of the networks that has been the subject of numerous studies in recent times is the Internet, which is composed of servers around the world exchanging enormous volumes of information packets regularly [26]. It can be studied at the level of individual routers, i.e., special purpose computers on the network that control the movement of data. Alternatively, it can be analyzed at the level of autonomous systems (AS), i.e., groups of computers within which communication is handled by a local internal network, but between AS, data is transmitted over the Internet. This physical network forms the backbone of another technological graph, the World Wide Web [5]. This is a network of web-pages which are linked together by hyper-links from one page to another.

Another class of technological networks is that formed by networks which allow transportation, either of resources or of people. One of the most important examples is the electrical power grid, whose nodes are generators, transformers and substations that are linked together by high-voltage transmission lines [27]. This network has been the focus of several studies which look at how local failures can lead to cascading failures resulting in overall or global catastrophic breakdowns. Transportation networks can also be defined in terms of the distinct modes by which movement of individuals occur between different geographical locations. The nodes are cities and towns, while the links may correspond to highways and smaller roads (for the road transportation network [28]), trains (for the railway network [29]) or flights (for the airline network [30]).

Financial networks

In the financial domain, one has the scope of looking at different types of networks including flow systems, such as the credit transfer network between banks, where the nodes are financial institutions that are linked by exchange of loans and debts [31]. Other examples include the graph of interacting stocks, where two stocks in a financial market are connected if their price fluctuations are signifi-

cantly correlated [32]. This network has often been used for the classification of stocks into different business sectors and the identification of unrelated stocks for the purpose of portfolio management [33].

1.3 Fundamental network concepts

Graph theory is the natural framework for the exact mathematical treatment of complex networks. Formally, a complex network can be represented as a graph which is defined in terms of a set of N vertices (or nodes) and E edges (or links). Every edge corresponds to a specific pair of nodes in the graph. We shall use the terms network and graph interchangeably in this thesis.

Adjacency Matrix

Any two nodes that are joined by a link are referred to as being adjacent or neighboring. A complete description of the connection topology of a graph is provided by a tabulation of every connected pair of nodes in it. Alternatively, this information can be gleaned from its adjacency matrix. A matrix $A = \{a_{ij}\}_{N\times N}$ is called the adjacency matrix of a graph G with N nodes, if the elements of A have the following property:

$$a_{ij} = \begin{cases} 1 & \text{if nodes } i \text{ and } j \text{ are adjacent in } G, \\ 0 & \text{otherwise.} \end{cases} \tag{1.1}$$

This matrix is symmetric if the network is undirected, i.e., if a link between nodes i and j exists, so does a link between j and i. On the other hand, if the network is directed, i.e., each link has an associated direction, then the matrix is asymmetric.

1.3.1 Measures for complex networks

Various properties of the connection topology for a complex network can be used to characterize the system. Indeed, many local and global measures have been introduced in the literature over the years, in order to unveil the organizational principles of networks. Below, we describe some of the most commonly used measures.

Degree

The simplest local characteristic of a node i is its degree, k_i, which is the total number of connections it has to other nodes. It can be calculated from the adjacency matrix as

$$k_i = \sum_{j=1}^{N} A_{ij}. \tag{1.2}$$

In the case of directed networks, the number of incoming (outgoing) edges of a vertex is called its in-degree (out-degree). The mean degree $\langle k \rangle$ is the average of k_i over all nodes $i = 1, \ldots, N$ in the graph. In an undirected graph, each edge contributes to the degree of two nodes, so that $\langle k \rangle = \frac{2E}{N}$, where E and N are the total number of links and nodes in the network, respectively. A node whose degree is significantly large compared to the average degree of the network is termed as a *hub*. A fully connected graph of N nodes with $k_i = N - 1 \; \forall i$ is called a *clique*. Most real-world networks are *sparse* such that their average degree is much smaller compared to the corresponding clique [3, 4].

Degree distribution

Although degree is a local parameter, we can obtain information about the global topology of the network by looking at its degree distribution, p_k, which is the set of probabilities that a vertex has degree $k = 1, 2, \ldots, N - 1$. A network having a narrow degree distribution with a well defined mean and a small variance indicates that all its nodes are similar in terms of structural importance, and that the network can be well described by its average properties. However, many networks occurring in reality are characterized by a degree distribution which decays as a power law:

$$p_k \sim k^{-\gamma}, \tag{1.3}$$

with an exponent γ whose value is typically seen to range between 2 and 3. Thus, there is a significantly high probability of observing vertices with large degree relative to the network size [34]. The power-law distribution implies that there is no characteristic scale for the degree of the nodes, so that this class of networks is also termed as *scale-free networks*. In addition to power laws, degree distributions that follow truncated power law or exponential distributions are also observed in

many networks occurring in nature and society [35].

Path length

A global measure of a network is provided by the shortest path length or distance between any pairs of nodes i and j. This is defined as the number of links that must be traversed to go from one node to another using the shortest route. The average of shortest path lengths over all pairs of nodes in the graph, also known as the characteristic path length, is an indicator of compactness of the network. It is defined as

$$\ell = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i \geq j} d_{ij}, \tag{1.4}$$

where d_{ij} is the shortest path length from vertex i to j and N is the number of nodes in the network. However, if the network consists of disconnected parts, the above definition gives infinite ℓ. To avoid this problem one can define ℓ on such networks to be the harmonic mean of the shortest distance between all pairs:

$$E \equiv \ell^{-1} \equiv \frac{1}{\frac{1}{2}N(N-1)} \sum_{i > j} \frac{1}{d_{ij}}. \tag{1.5}$$

This is also termed as *efficiency* of the network and is a measure of the speed with which information propagates over the network [36].

Most real-world networks have been seen to exhibit the small world property, which is related to the observation that one can reach a given node from the other nodes in a very small number of steps, on average. In recent years, the term *small-world effect* has taken on a more precise meaning: networks are said to show the small-world effect if ℓ scales logarithmically or slower with network size for fixed mean degree, $\langle k \rangle$ [4].

Diameter

Another related measure for compactness of the network is its diameter D, which is defined as the longest of all the shortest paths in the network.

$$D = \max\{d_{ij}\}, \ \forall \ i\text{-}j \text{ pairs of shortest paths.} \tag{1.6}$$

Network	N	$\langle k \rangle$	ℓ	C
Protein interaction	2115	2.12	6.80	0.071
Physics co-authorship	52909	9.27	6.19	0.56
Internet	10679	5.98	3.31	0.39
Marine food web	135	4.43	2.05	0.23

Table 1.1: Properties of some real-world complex networks: size (N), average degree ($\langle k \rangle$), characteristic path length (ℓ) and average clustering coefficient (C). (From Ref. [4])

As the diameter and characteristic path length are related properties, sometime these measures are used interchangeably to measure the network compactness.

Clustering

Many real networks have been shown to have a significant transitivity in the pattern of their connections, such that, if the pairs of nodes i, j are connected and the pair j, k are also connected, then so is the pair i, k. This is equivalent to having a significantly high frequency of triangular structures in the network [37]. In such circumstances, the nodes of the network are said to be clustered. The compactness of the local neighborhood for a node i is measured by the clustering index:

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, \tag{1.7}$$

where, E_i is the number of edges among the k_i neighbors of node i. Note that, $C_i = 1$ if the neighbors of node i are fully interlinked, and $C_i = 0$ if none of its neighboring nodes share any links with each other. The average clustering coefficient for the entire network, C, is defined as the average of C_i over all the nodes in the network, i.e., $C = \frac{1}{N} \sum_{i=1}^{N} C_i$.

This average clustering coefficient is a measure of the “cliquishness" or local compactness of a network. For different real networks, C takes values which are orders of magnitude larger than that of an equivalent random graph with the same number of nodes and edges. If, in addition to the small world property, a network also possesses a high clustering coefficient C, then it is termed as a *small-world network* (SWN). Many of the real world network are seen to belong to this class [37, 38].

1.4 Models of complex networks

One way of understanding complex networks observed in nature and society is to construct a minimal model that exhibits properties which are similar to those of empirical networks. Such a network model can help to explain processes by which such systems evolve and also shed light on the function of the network. Further, a network model can be used for studying the dynamics on such networks, e.g., to understand how the processes of synchronization and diffusion are affected by different network topologies [39, 40].

1.4.1 Erdos-Renyi random network

The earliest mathematically analyzed non-trivial network model in the literature is that for an ensemble of homogeneous random graphs introduced by Erdos and Renyi [41]. Starting from a set of N disconnected nodes, each pair of nodes is connected with a probability p. This simple model leads to a surprising list of properties, many of which can be computed exactly in the limit of large N. For a sparse graph, if the average number of edges in the graph is a fraction p of the $N(N-1)/2$ possible edges, then the average degree

$$\langle k \rangle = \frac{2E}{N} = p(N-1). \tag{1.8}$$

The degree distribution can also be computed, with the probability of a vertex having degree k being

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{\langle k \rangle^k \exp^{-\langle k \rangle}}{k!}. \tag{1.9}$$

The approximate equality, i.e., binomial distribution being approximated by a Poisson distribution, becomes exact in the asymptotic limit of large network size. These graph are therefore also known as *Poisson random graph.*

The expected structure of the random graph varies with the connection probability p. For $p = 0$, there are no edges and the graph is termed an empty graph, whereas for $p = 1$, all possible edges exist and we get a complete graph. As p increases from 0, the edges join nodes together to form components, i.e., subsets of

nodes that are connected by paths through the network. Erdos and Renyi demonstrated that the random graph undergoes a phase transition at a critical value of $p_c = 1/N$, from a low-density state in which there are few edges and all components are small to a high-density state in which an extensive [i.e., O(n)] fraction of all nodes are joined together in a single *giant component*. This component is a set of mutually reachable nodes, whereas the remainder of the nodes occupy smaller components. With increasing p, the giant component captures more and more nodes of the graph. Another important feature is the occurrence of a second connectivity transition at $p_{c1} = \ln N/N$. For $p > p_{c1}$, all sites belong to a single component (in the limit $N \to \infty$), while for $p < p_{c1}$ disjoint clusters can exist.

These graphs have a low clustering coefficient as the probability of connection between two nodes is p regardless of whether they have a common neighbor or not. Hence,

$$C = p = \frac{\langle k \rangle}{N-1}, \tag{1.10}$$

which goes to zero as N^{-1} in the limit of large system size [37]. To get an idea of the average path length for the graph, note that the mean number of neighbors at a distance ℓ away from a vertex in a random graph is $\langle k \rangle^\ell$, so that the value of ℓ needed to encompass the entire network is $\langle k \rangle^\ell \simeq N$. Thus a typical characteristic distance for the network is

$$\ell = \log N / \log \langle k \rangle. \tag{1.11}$$

This scaling is much slower than that of a d-dimensional regular lattice where $\ell \sim N^{1/d}$. If the growth of $\ell(N)$ is slower than any positive power of N, it is referred to as *small-world effect* [42].

The ease of analysis for random graphs has proven to be very useful in the early development of the field. Although the average path length scales logarithmically with graph size and therefore, shows the small-world effect, in almost all other respects the properties of random graphs do not match those of networks in the real world. Their degree distribution is Poisson, whereas most real-world graphs seem to exhibit broader degree distributions. Also, the random graph lacks clearly defined communities and the clustering coefficient is usually far smaller than that in comparable real-world graphs. The basic Erdos-Renyi model has been extended in several ways, e.g., to exhibit a power law degree distribution pattern [43, 25].

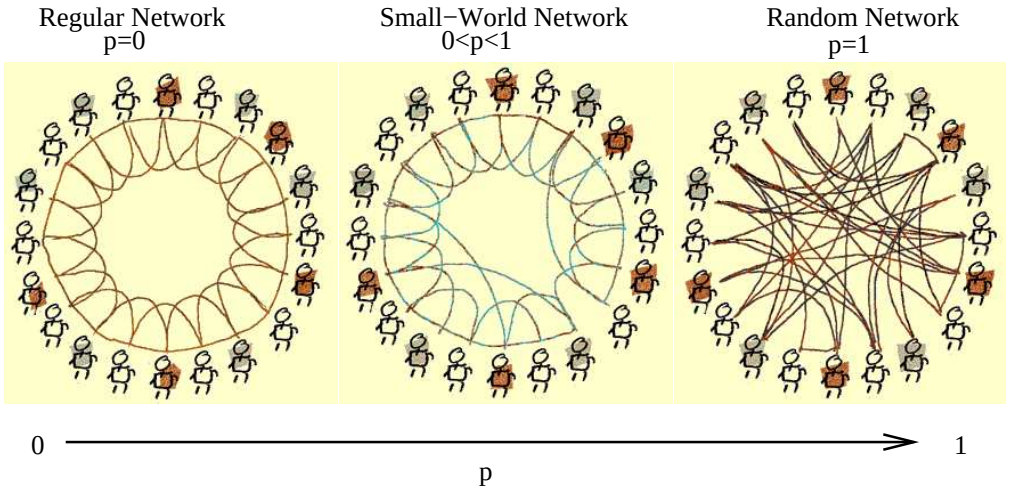


Figure 1.3: The Watts-Strogatz (WS) small-world network model, seen as an interpolation between a regular and a random network.

However, these models do not describe how real-world properties evolve dynamically, thus making them less useful in understanding the processes of network formation in the real world.

1.4.2 Watts-Strogatz small-world network

Social networks often show a high tendency of being transitive, that is two people who are friends have a high probability of having one or more *mutual* friends. This kind of clustering is not seen in random graphs, as mentioned previously. In 1998, Watts and Strogatz proposed a mechanism for generating small-world networks with high clustering [37]. This model is often termed as the WS-model and the generative mechanism is as follows: A regular network is first constructed by arranging N nodes on a $1-$d periodic lattice. Each vertex is connected to $k = 2z$ nearest neighbors within the range z, so that all nodes have the same initial degree. Next, one goes through each edge, and with rewiring probability p, detaches the far side of the edge and reconnects it to a randomly chosen vertex (excluding self and multiple connections).

Changing the rewiring probability p allows us to investigate the transition from a regular graph ($p = 0$) to a random graph ($p = 1$) (Fig. 1.3). Let us consider first the limit $p = 0$, where the network is regular and arranged on a ring. The shortest

average path length for this system is $\ell \sim N/4z$ for large N, and this grows linearly with N. The clustering coefficient $C_{\rm reg} = (3z-3)/(4z-2)$ is constant and tends to $3/4$ for large z. This large value indicates the presence of a significant number of triangular structures in the network. On the other hand, for $p = 1$ we have the random graph for which $\ell \sim \ln N/\ln z$ and $C \sim 2z/N \to 0$ as N increases. In the WS model, by changing the rewiring probability one finds that there is a broad range of p, where $\ell \approx \ell_{\rm rand}$ and $C = C_{\rm reg}$. Thus, globally the network has the small-world property of a random graph, while locally it is clustered like a regular graph. This is because the diameter ℓ drops rapidly when p increases, as adding even a few short-cuts during the rewiring process reduces the average distance between any pair of nodes significantly. However, the clustering coefficient

$$C = \frac{3(k-1)}{2(2k-1)}(1-p)^3, \tag{1.12}$$

of the network decreases very slowly with increasing p [44].

The WS-model was one of the first models that could explain the co-existence of high clustering and small-world effects. Further, this model introduced the concept of physical distance constraints in network formation. For example, it is easy to form a link between nodes which are geographically close to each other. Although other variations of the WS network have been proposed, in all these models the signature of a physical d-dimensional lattice is still observed, so that, shortcuts occur with higher probability between nodes that are physically closer. However, the conventional WS model does not exhibit a broad degree distribution, and the discovery of this latter feature in several real-world networks led to the next breakthrough in the physics of complex networks [34, 45].

1.4.3 Barabasi-Albert scale-free network

First proposed to explain the degree distribution in citation networks [46], the idea of preferential attachment has been rediscovered recently by A-L. Barabasi and R. Albert (BA) in a network model that shows broad degree distributions described by a power law [34]. They showed that the scale-free nature of these networks can originate from two generic features seen in many real-world networks,

1. *Growth*: Networks are open systems with the number of nodes growing with

time (i.e., N increases), and

2. *Preferential attachment*: New nodes in the graph are not connected randomly but preferentially attach to existing nodes which have high degree, thereby making the degree of the latter even higher. This process is sometimes referred to as the *rich getting richer* phenomenon.

If Π, the probability that the new node will be connected to node i, depends linearly on the degree k_i of node i, i.e.,

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}, \tag{1.13}$$

then, it was shown that the model network evolves into a system with a scale-invariant degree distribution having an exponent $\gamma = 3$.

As the degree distribution of the preferential attachment models match with those occurring in real-world graphs, it suggests that real networks might have been generated by similar processes. However, many networks in nature with a broad degree distribution show deviations from a pure power-law, typically exhibiting an exponential cutoff at high degrees:

$$p_k = k^{-\gamma}\phi(k/\xi) \tag{1.14}$$

where $\phi(k/\xi)$ is the cutoff at some scale. In the context of the growing BA model, this phenomenon can be explained due to aging and saturation effects that limit the number of links a node can acquire. Thus, the preferential attachment function, $\Pi(k_i)$ is nonlinear, following $\Pi(k_i) = f(k_i)/\sum_j f(k_j)$, where $f(k)$ is an arbitrary function, resulting in deviations from the power-law [47].

The average path length ℓ of the BA network ($\gamma = 3$) grows as

$$\ell(N) \sim \frac{\ln(N)}{\ln\ln(N)} \tag{1.15}$$

with N slower than $\ln N$, which is also termed as *ultra-small-world effect* [48]. This indicates that the heterogeneous scale-free topology is more efficient in bringing the nodes closer than the homogeneous topology of random network. Other scale-free networks with $2 < \gamma < 3$ have a much smaller diameter, with $\ell \sim \ln\ln(N)$,

while for networks with $\gamma > 3$, the shortest path length $\ell \sim \ln(N)$ [49]. The clustering coefficient of the BA model decreases with the network size, following approximately a power law, $C \sim N^{-0.75}$. While being slower than the $1/N$ decay observed for C in random graphs, this is still different from the behavior of small-world network models and real world networks, where C is independent of N [3]. Further there is a strong correlation between age and degree in this model which is rarely seen in real-world systems. Moreover, only linear preferential attachment gives a power-law degree distribution, that brings into question the general validity of this process.

1.5 Importance of mesoscopic organization in complex networks

It has now been known for some time that the topological structure of a network can affect the function of the system [6]. E.g., it has been shown that the connection architecture has important consequences on the functional robustness of the network and its response to external perturbations [50]. This has led to a series of studies pointing to the crucial role played by the network topology in determining the emergence of collective dynamical behavior [39, 40], such as synchronization, diffusion, the spreading of contagion such as epidemics, information and rumors, etc. To study this we need to go beyond the properties of single nodes and pairs of nodes, and consider the mesoscopic properties of networks (i.e., properties of groups or local clusters in the network).

Motifs

Network motifs are patterns (sub-graphs) that occur within a network much more often than expected in corresponding randomized versions. Most networks studied in biology, ecology and other fields have been found to show a small set of network motifs which occur again and again. Each class of networks seems to display its own set of characteristic motifs, e.g., motifs that are commonly seen in food webs are distinct from the motifs seen in the genetic networks of different species. However, similar motifs are found in networks that have similar function, such as information processing, even though they describe elements as different as biomolecules within

a cell and synaptic connections between neurons. These small circuits therefore can be considered as simple building blocks from which the network is composed [51].

Modularity

Looking beyond small micro-level motifs, it has been observed that, at the mesoscopic level many of the networks in real-world have modular structure [52]. Modules or communities are subnetworks within the network, where connections are more frequent between nodes within the same subnetwork than between nodes of different subnetworks. The presence of modular structure may also alter the way in which dynamical processes (e.g., spreading processes, synchronization) unfold on the network. With this realization many of recent studies have focused on models of modular networks and their inter-relation with the dynamical processes taking place on the network [53].

Hierarchy

Further, these networks have also been shown to have *hierarchical* organization, i.e., they are composed of successive interconnected layers or inter-nested communities [54]. Hierarchy describes the organization of elements in a network: how nodes link to each other to form communities and how communities are joined to form the entire network. E.g., the metabolic network of several organisms can be organized into highly connected modules that hierarchically combine into larger units [55]. The observed hierarchy also coincides with known metabolic functions, indicating that there may be a functional basis for such meso-level organization.

1.6 Overview of thesis

The aim of the present thesis is to look at the mesoscopic organization of complex networks. This is viewed from three perspectives: (i) the structural properties of such an organization, (ii) their role in dynamical processes defined on such networks, and (iii) the possible origin and evolution of such structures. These are complemented by empirical analysis of networks occurring in reality that show similar organizational features.

In **Chapter 2**, we investigate the structural and dynamical consequences of modular organization in networks. Using a simple model, we show that small-world networks can arise as an immediate result of modular configuration. We demonstrate a distinct dynamical signature for such modular networks, namely, the existence of multiple characteristic time scales in processes as different as synchronization, diffusion and spin-ordering. The dichotomy between fast intra-modular dynamics and slow inter-modular dynamics is directly related to the topological structure of the model through the spectral properties of the network Laplacian. By verifying the existence of similar features in the empirically determined cortico-cortical networks in cat and macaque brains, we propose that the modular network model may better represent certain natural systems reported to have small-world properties [56, 57].

To understand the process by which networks evolve towards modular organization, we note that they are subject to multiple structural and functional constraints. In **Chapter 3**, we consider the particular examples of (i) minimizing the average path length, (ii) minimizing the total number of links, while (iii) maximizing robustness against perturbations in node activity. We show that the optimal network satisfying these three constraints is modular, characterized by the existence of multiple sub-networks sparsely connected to each other. In addition, these modules have distinct hubs resulting in an overall heterogeneous degree distribution, as seen in many real networks [58].

In addition to the existence of modular structures, several networks in nature also have these modules arranged in a hierarchical fashion. Therefore, we next consider a model for such hierarchical modular networks in **Chapter 4**. We show that a scaling relation between the clustering and degree of the nodes is not a necessary property of such networks, contrary to what has been claimed recently. We investigate the dynamical properties of such networks, in particular, the stability of (i) equilibria of network dynamics, and (ii) synchronized activity. For both these cases, we find that increasing modularity or the number of hierarchical levels tend to increase the probability of instability. As both hierarchy and modularity are seen in natural systems, which necessarily have to be robust against environmental fluctuations, we show using a generalization of the model used in Chapter 3, how constraints on communication efficiency and maximum degree can result in the emergence of hierarchical structures [59].

After having analyzed network models for understanding the dynamical consequences of modularity and hierarchy in the preceding chapters, in **Chapter 5** we consider how their existence in complex systems occurring in reality, can be uncovered from a knowledge of the collective dynamics of the component nodes. We first demonstrate the possibility of reconstructing a network through an analysis of the time-series data of its components, by using the behavioral data of individuals belonging to a troupe of macaque monkeys. To reconstruct a much more complex network from dynamical information about its components, we consider the example of financial markets. These complex systems have many interacting elements (traders and stocks) and exhibit large fluctuations in their associated observable properties, such as stock price or market index. By analyzing the cross-correlation matrix of stock price fluctuations through spectral techniques, we reveal the underlying network of interactions between stocks in different markets. We observe the existence of modules which approximately correspond to specific business sectors. Using a multi-factor model, we suggest that the gradual emergence of modules, indicating the strengthening over time of direct interactions between related stocks, is a signature of market development [33].

In **Chapter 6**, we consider the reverse problem, i.e., we try to understand the functional significance (arising out of the dynamical consequences) of the observed structural features. As an example, we consider the somatic nervous system of the nematode *C. elegans*. We determine the structural modules of the neuronal network, and show that such an organization can only be explained if one consider constraints that are possibly related to the information processing function of the system apart form static considerations. A detailed analysis of the intra-module degree and participation coefficient allows us to identify key neurons involved in information processing tasks which are verified from earlier reports of experimental studies. We also show that the existence of a hierarchical structure in the nervous system has the functional benefit of reducing diffusive spread of activity throughout the network (thus, acting as a noise filter), while maintaining high communication efficiency between neurons [60].

We conclude with a general discussion on how the dynamics occurring on a network is governed by mesoscopic structural features of the system, and in turn, affects the evolution of ubiquitous structural patterns, such as modularity and hierarchy.

2

Modularity in complex networks

In many natural situations, dynamics at the local level may occur over a very different time-scale compared to processes at the global level. Such a temporal separation is often desirable functionally, e.g., for information processing in the brain. It requires synchrony between local areas processing specific stimuli [61], but, global or very large scale synchrony is considered to be pathological, as in epilepsy [62]. Many systems in nature have network descriptions, with the connection topology playing a crucial role in determining their dynamical behavior [12]. Therefore, it is of considerable interest to understand how the structural organization in complex networks can give rise to dynamics at multiple discrete time-scales.

As discussed in Chapter 1, a large class of networks in nature have also been reported to be small-world networks (SWN) [37], which are characterized by the coexistence of very high clustering among neighboring nodes and short average path length. The clustered structure of SWN distinguishes them from networks with "small-world property" [38], whose average path length increases slower than any polynomial function of the system size. This latter feature is seen in random graphs, as well as, in most complex networks [4]. SWN have been reported in a variety of contexts, including the brain [63], human society [64] and cellular metabolism [65]. Several models for SWN have been proposed [66], beginning with a simple interpolation scheme between regular and random structure through rewiring of links (the WS model) [37] [Fig. 2.1 (a)].

In this chapter, we relate the independent properties of dynamical time-scale separation and the clustered small-world property of many complex networks, with the crucial observation that such systems often manifest modular structure [67].

Modules are defined as subnetworks comprising of nodes connected to each other with a density significantly higher than that of the entire network. Modular structures have been observed in a wide variety of contexts, from cellular metabolism [68] and signalling [69] to social communities [70], internet [71] and foodwebs [19]. Our results, therefore, suggest that the large number of instances of SWN in the real world is related to the ubiquity of modular structures in complex systems.

In Sec. 2.1, we introduce a simple model of modular networks. In Sec. 2.2, we show that these networks exhibit all the structural characteristics of SWN. Such modular networks, in sharp contrast to previous models of SWN, exhibit distinct time-scale separation in their dynamics, corresponding to fast intra-modular and slow inter-modular processes. In Sec. 2.3, we show the universality of this behavior by using three very different types of dynamics, viz., (i) the ordering of spins through exchange interactions, (ii) synchronization among relaxation oscillators and (iii) diffusion. In all cases, the modular configuration allows coordination within local clusters to occur much more rapidly than global ordering. The occurrence of multiple discrete timescales in such a wide variety of systems highlights the role of modularity in the dynamics on complex networks. In Sec. 2.4, we show that these multiple timescales can be related to the Laplacian spectra of the network. We conclude by discussing in Sec. 2.5 how identifying modular structures is crucial for designing intelligent intervention strategies for complex systems, e.g., controlling epidemics.

2.1 Modular random networks: A model

The network model considered in this chapter follows directly from the definition of modular networks and consists of N nodes arranged into m modules (similar to the construction used, e.g., in Ref [52]). Each module contains the same number of randomly connected nodes [Fig. 2.1(b)]. The connection probability between nodes belonging to the same module is ρ_i, and for those belonging to different modules is ρ_o. Thus, one of the key parameters defining the model is the ratio of inter- to intra-modular connectivity $\frac{\rho_o}{\rho_i} = r \in [0, 1]$. For $r \to 0$, the network gets fragmented into isolated clusters, while as $r \to 1$, the network approaches a homogeneous or Erdos-Renyi (ER) random network. The other parameter that

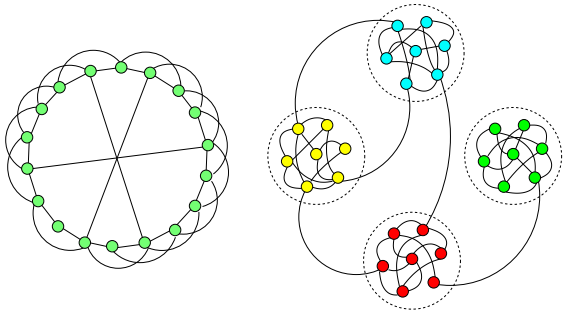


Figure 2.1: Schematic diagrams of (left) Watts-Strogatz model and (right) modular network, with modules in the latter indicated by broken circles.

together with r completely defines the modular network is its average degree (i.e., the number of links per node),

$$\langle k \rangle = \frac{\rho_i}{m} \left[(N - m) + rN(m - 1) \right]. \tag{2.1}$$

2.2 Static properties of modular networks

To look at the structural properties of the model, we first consider the communication efficiency E for the entire system. This is a measure of the information propagation speed over the network and is defined as [36],

$$E \equiv \ell^{-1} \equiv \frac{1}{\frac{1}{2}N(N-1)} \sum_{i>j} \frac{1}{d_{ij}}, \tag{2.2}$$

where, d_{ij} is the shortest distance between nodes i and j. Note that, E is related to the harmonic mean distance, ℓ, which is a measure of the average path length. We also quantify the clustering within local neighborhoods by measuring the coefficient $C = (1/N) \sum_i 2n_i/k_i(k_i - 1)$, where k_i and n_i are the degree and the number of links between the neighbors of node i, respectively. For the modular random network,

$$C = \rho_i(d_1^2 + (m-1)d_2^2) + (m-1)\rho_o(2d_1d_2 + (m-2)d_2^2), \tag{2.3}$$

where, $d_1 = (\frac{N}{m} - 1)\rho_i/\langle k \rangle$ and $d_2 = \frac{N}{m}\rho_o/\langle k \rangle$ are the probabilities that a node has a neighbor in the same or a different module, respectively. Thus, if the number of

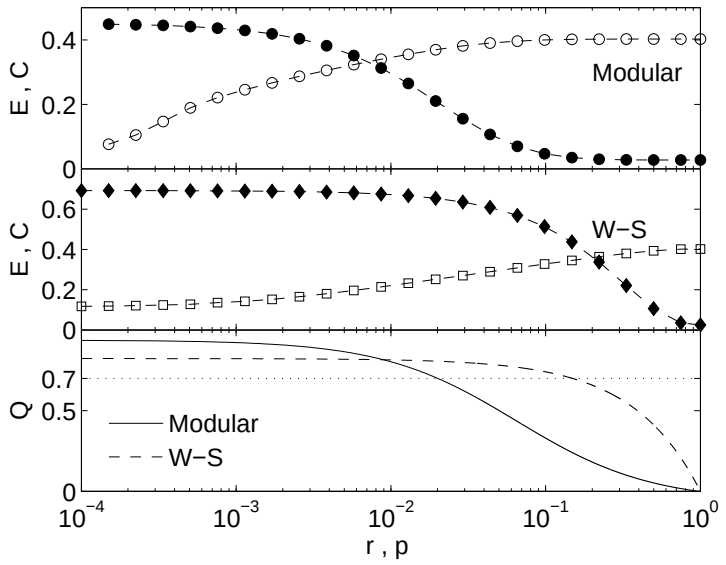


Figure 2.2: Communication efficiency E (empty circle) and clustering coefficient C (filled circle) for (a) modular random network with $m = 16$ modules as a function of r and (b) Watts-Strogatz (WS) network as a function of rewiring probability p ($N = 512$ and $\langle k \rangle = 14$). The data points are obtained by averaging over 100 realizations. Error bars are in all cases smaller than the symbols used. (c) The variation of modularity measure, Q_M, with r for modular random networks (solid line) and with p for WS network (broken line). The dotted line indicates $Q_M = 0.7$ and its intersection with the other two curves gives a pair of r and p values at which we can compare the two model networks.

modules is large then, clustering is high at low values of r. As r is increased in our model, we observe an increase in E while simultaneously C decreases [Fig. 2.2 (a)]. The small-world property is associated with high values of both E and C, which is indeed what is observed in our model for an intermediate range of r, exactly as in the WS model [Fig. 2.2 (b)].

Next, we characterize the model using a measure of modularity, Q [72]. For a given partition of the nodes of a network into m modules,

$$Q \equiv \sum_{s=1}^{m} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right] , \tag{2.4}$$

where the total number of links in the network is L, and l_s and d_s are the links between nodes and the total degree of all nodes belonging to module s, respectively. The largest modularity that is obtained from all possible partitions of the

network is denoted by $Q_M = \max\{Q\}$. A high value for Q_M is a necessary but not sufficient condition for a network to be modular, as there can be various regular graphs having high Q_M value for which the modules cannot be identified unambiguously [73, 74]. In particular, for the WS small-world model, calculating Q_M yields high values although the modules are not defined in an unique manner. For a WS network defined on a ring of N nodes (each connected to $2z$ nearest neighbors) where a fraction p of the links have been rewired,

$$Q = (1-p)\left(\frac{2N - zm - m}{2N} - \frac{1}{m}\right). \tag{2.5}$$

Here, the existence of m modules of equal size $n = N/m$ were assumed for the calculation of Q. The maximum value $Q_M = (1-p)\left[1 - \sqrt{2(z+1)/N}\right]$, occurs for $m^* = \sqrt{\frac{2N}{z+1}}$ and can be very high for low p. Similar high values of Q_M are obtained for modular random networks at low r, the modularity measure for such a system with N nodes being

$$Q = \frac{(m-1)[N(1-r) - m]}{m[N(1-r+rm) - m]}. \tag{2.6}$$

Unlike the WS model, here the modules are pre-defined and Q does not need to be maximized with respect to different choices for partitioning the network. Fig. 2.2 (c) shows the variation of Q with r and p for the two classes of small-world network models.

Note that, WS networks are parametrized with respect to the rewiring probability p, while modular random networks are defined in terms of r, the ratio of inter- to intra-modular connectivity. Therefore, in order to circumvent the difficulty in directly comparing these two types of networks, in subsequent work we have considered networks having the same N, $\langle k \rangle$ and Q. We observe that it is difficult to differentiate between WS and modular random networks from their structural information only, by using any of the commonly used static measures. For example, on applying the k-clique (complete subgraphs with k nodes) percolation cluster technique used for detecting overlapping communities [75, 76], we found large clusters to appear in both types of networks. This is because the local link density in both systems are much higher than their overall connectivity.

Other measures such as betweenness centrality (BC), edge clustering (EC), etc., also gave similar results for the two network models. The betweenness centrality of an edge (i.e., link) is defined as the number of shortest paths between pairs of vertices that go through it. If there are more than one shortest path between a pair of vertices, each path is given equal weight such that the total weight of all the paths is unity. The edges with maximum betweenness often act as the bridges between different community. So by removing the edges with maximum betweenness centrality the modular structure can be determined [52]. We find that this is indeed what happen for modular network. However, for Watts-Strogatz (WS) network, the shortcuts or rewired links also have high edge betweenness. Thus, the above algorithm which removes edges with high betweenness centrality cause the removal of these shortcuts, so that only the regular chain structure is left (Fig. 2.3). When further links are removed then the chain structure is divided into group of nodes (which are roughly of equal size). Hence, applying this method to determine modules in a network gives community structure in both the cases.

Instead of BC, other parameters like edge clustering can also be used to determine the modular structure in a network [77]. Edge clustering is defined, analogous to the node-clustering coefficient, as the ratio of the number of triangles to which a given edge belongs to the total number of potential triangles that might include it given the degrees of the adjacent nodes. More formally, for the edge connecting node i to node j, the edge clustering coefficient is

$$C_{ij}^{(3)} = \frac{z_{ij}^{(3)} + 1}{\min(k_i - 1, k_j - 1)},$$

where $z_{ij}^{(3)}$ is the number of triangles to which that edge belong and $\min(k_i - 1, k_j - 1)$ is the maximal possible number of triangles. As edges that link different communities are unlikely to belong to many short loops, these edges have low EC. Therefore, removing the edges with low EC will reveal the community structure as disjoint subsets if the underlying network structure is modular. However, in a WS network the shortcuts have low EC, and hence the algorithm of removing low EC links will again remove the shortcuts resulting in a lattice structure. As above, on further removal of links, the chain structure gets disrupted into disjoint groups of nodes. Hence, the WS network also appears to have a community structure.

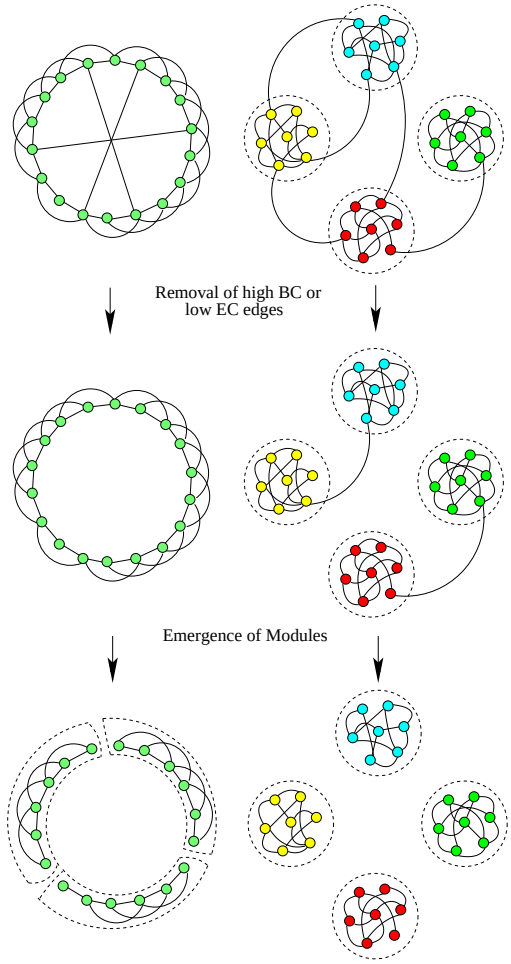


Figure 2.3: Schematic diagram showing the effect of removing edges with high betweenness centrality (BC) or low edge clustering (EC). For both WS (left) and modular (right) networks, the system gets divided into isolated communities (groups of nodes).

2.3 Dynamics on modular networks

So far we had been considering exclusively the structural aspects of small world networks. However, apart from topological structure, networks are often associated with certain dynamics [6]. As dynamics is often crucial for the functioning of many systems, we now examine very different dynamics on network models having the clustered small-world property. These dynamics range from nonlinear interactions (representative of collective ordering in a network) to strongly nonlinear local dynamics at each node (as in relaxation oscillators) with diffusive coupling.

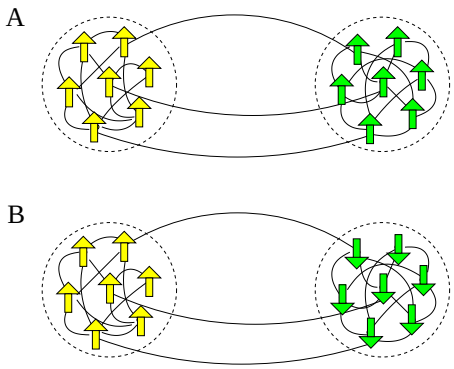


Figure 2.4: Schematic diagram of (a) global ordering ($M = 1$, $M_m = 1$) and (b) modular ordering ($M = 0$, $M_m = 1$) in a modular network of Ising spins.

2.3.1 Spin-Spin interaction dynamics

We first consider the effect of modular organization on the emergence of collective behavior, a simple model of which is the ordering of Ising spins arranged on a network. This system is described by the Hamiltonian,

$$H = -\sum_{i,j} J_{ij}\sigma_i\sigma_j, \tag{2.7}$$

where, $\sigma_i, \sigma_j = \pm 1$ are spins placed on nodes i, j, and J_{ij} is the ferromagnetic coupling between them ($= J > 0$ if i, j are connected and 0 otherwise). Starting from an initial random configuration of spins on a modular random network with average degree $\langle k \rangle$, the system is allowed to evolve to its ground state using Glauber dynamics. It corresponds to a globally ordered state [Fig. 2.4 (a)] if $T < T_c (= \langle k \rangle)$, the mean-field critical temperature measured in units of J/k_B (k_B: Boltzmann constant). We observe that the time (τ_{gm}) needed for magnetization $M = \sum_{i=1}^{N} \sigma_i / N$ to reach its high asymptotic value, diverges as r decreases. This is because, at low r, the system remains for a long time in a state of modular ordering [Fig. 2.4 (b)], where the spins in each module are ordered but aligned in opposite directions in different modules resulting in an absence of global ordering. The local order parameter, modular magnetization $M_m = \frac{m}{N}\langle | \sum_{i \in k} \sigma_i^k | \rangle$, where σ_i^k is the $i-$th spin in the k-th module and the averaging is over all modules, exhibits convergence to its asymptotic value over a time-scale τ_{mm}, which is almost

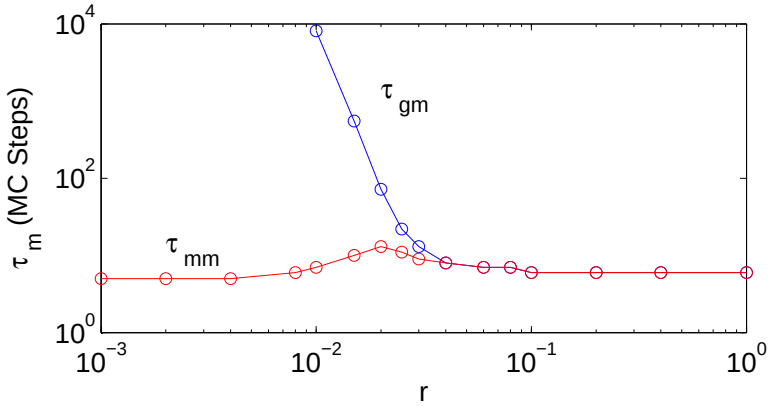


Figure 2.5: The two time-scales corresponding to local ordering within a module (τ_{mm}) and global ordering over the entire network (τ_{gm}) for a modular random network of Ising spins ($m = 16$) at $T = 6$ as a function of r. (In all cases $N = 512$, $\langle k \rangle = 14$).

independent of r.

Fig. 2.5 shows the existence of two time-scales which diverge at low r indicating the ordering process within modules to be much faster compared to between modules. At low temperatures, as the spins within each module get ordered, different modules may get aligned in opposite directions. To achieve global order, some of the modules need to turn all their spins, a process that has a considerable energy barrier. To cross this with thermal energy takes extremely long times, resulting in divergence of τ_{gm}. A similar investigation of the WS network shows only global ordering, with τ_{gm} diverging as p decreases. Related dynamical processes where the appearance of distinct time-scale events as a consequence of modular network structure have important functional significance, include the adoption of innovations [78], spread of epidemics [8] and consensus formation [79].

2.3.2 Synchronization

Next, we compare the dynamics of synchronization in modular random and WS networks. We consider a population of N coupled relaxation oscillators (described

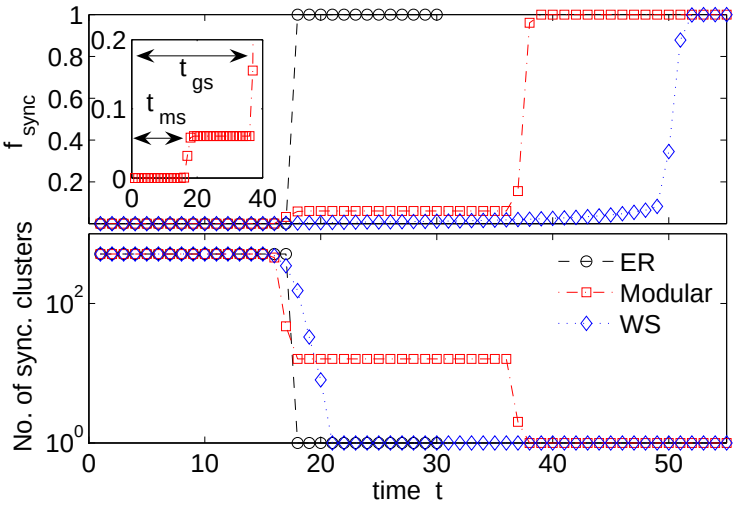


Figure 2.6: Time evolution of (top) the fraction of synchronized nodes f_{sync} and (bottom) the number of synchronized clusters for ER random, WS small-world ($p = 0.2$) and modular random networks ($r = 0.02$). For all cases, $N = 512$ and $\langle k \rangle = 14$. Unlike the ER and WS networks, the synchronization in modular networks ($m = 16$) occur in two distinct steps. Local synchronization within nodes belonging to the same module is achieved relatively fast and is then followed by global synchronization. The two time scales corresponding to synchronization within modules (t_{ms}) and synchronization over the entire network (t_{gs}) are shown in the magnified view (inset).

by a fast variable x and a slow variable y) which evolve as

$$\dot{x}_i = c\left[y_i - x_i + \frac{x_i^3}{3}\right] + \sum_{j=1}^{N} \frac{K_{ij}}{k_i}(x_j - x_i); \tag{2.8}$$

$$\dot{y}_i = \frac{-x_i}{c}. \tag{2.9}$$

Here, c is the ratio between time-scales of x and y. $K_{ij} = \kappa A_{ij}$ is the coupling between a connected pair of oscillators with strength κ, and $\mathbf{A}$ is the network adjacency matrix, i.e., $A_{ij} = 1$ if i, j are connected and 0 otherwise. For networks of simple oscillator models, the approach to synchronization exhibits temporally varying patterns that are intrinsically related to the underlying connection topology [53].

We have analyzed the time-evolution to synchronization (i.e., $x_i = x$, $y_i = y$, $\forall i$) of these strongly non-linear oscillators using the pair-correlation function between

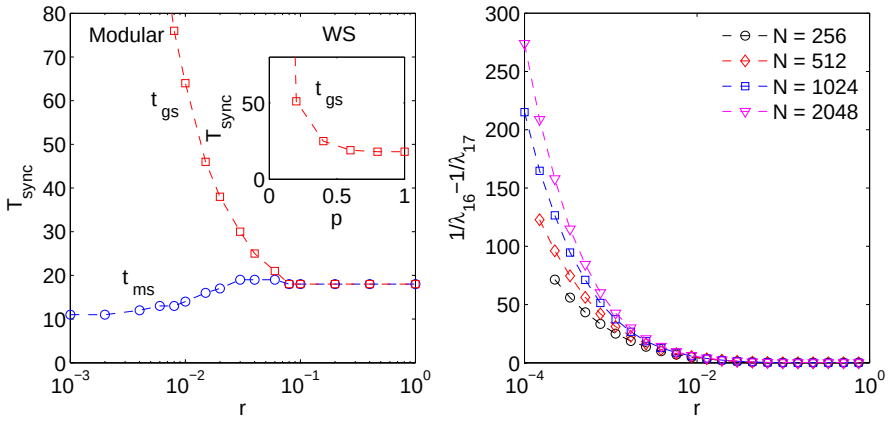


Figure 2.7: (a) Comparison of synchronization between modular random networks ($m = 16$) and WS networks of relaxation oscillators (Eq. 2.9) with $c = 2$ and $\kappa = 1.5$. ($N = 512$, $\langle k \rangle = 14$). In modular networks, the two time scales corresponding to intra-modular (t_{ms}) and global or inter-modular (t_{gs}) synchronization are shown as a function of r. The WS model exhibits only the time-scale corresponding to global synchronization (inset). Averaging has been done over random initial values and network realizations. (b) The Laplacian spectral gap between the m-th and $(m+1)$-th eigenvalues increases with decreasing r, shown for different system sizes with the number of modules $m = 16$.

oscillator phase angles θ [$= \arctan(y/x)$], $\rho_{ij}(t) = \langle \cos[\theta_i(t) - \theta_j(t)] \rangle$, where $\langle \ldots \rangle$ is an average over random initial conditions. By introducing a threshold T, the correlation matrix is converted into a dynamic connectivity matrix $\mathcal{D}_t(T)$ ($\mathcal{D}_{ij} = 1$, if $\rho_{ij} > T$, and $= 0$, otherwise). The ratio of non-zero elements of $\mathcal{D}$ to the total number of elements gives the fraction of synchronized nodes, f_{sync}, which increases to 1 with time as the system converges to global synchronization. Conversely, the number of distinct synchronized communities (i.e., the disconnected clusters in $\mathcal{D}$) decreases from N to 1 (Fig. 2.6).

As expected, we observe global synchronization to be extremely rapid in ER random networks, while, for WS networks it occurs relatively slowly. By contrast, in random modular networks, the synchronization occurs over two distinct time-scales, as reflected by the occurrence of a plateau with non-zero values of the two synchronization measures, f_{sync} and number of synchronized clusters. At the relatively shorter time scale of t_{ms}, disconnected clusters are observed to form in $\mathcal{D}$ corresponding to the structural modules of the network. Thus, local synchronization among the nodes belonging to the same module is achieved relatively quickly.

Global synchronization is a slower process, occurring over a time-scale t_{gs}, with the synchronized clusters remaining fairly stable in the intervening time-period. Fig. 2.7(a) shows the variation of these two time-scales with r, converging when the network becomes homogeneous (as $r \to 1$).

In the real world, for many systems the coupling strength between nodes within the same module may differ significantly from that between nodes belonging to different modules. For example, a recent study of tie strengths in mobile communication networks [11] observed that links connecting different communities tend to be weaker than links between members of the same community, supporting a well-known hypothesis for social networks [80]. Hence we look at the effect of different strengths for inter-modular coupling (κ_{inter}) and intra-modular coupling (κ_{intra}) on the synchronization behavior of oscillators on a modular network. As the inter-modular coupling strength becomes weaker relative to the intra-modular coupling, we observe the time-scales for modular and global synchronization to diverge (Fig. 2.8, a). Thus, in real systems where inter-community ties are relatively weaker, the time-scale separation between local and global events will be even more prominent. On the other hand, as the inter-modular coupling strength becomes large, the two time-scales gradually converge. As expected, at very large values of the ratio $\kappa_{inter} : \kappa_{intra}$, global and modular synchronization occur simultaneously.

We have also looked at the more general case of synchronization in the presence of delays in the coupling [39]. Even in the presence of delays, we observe distinct time-scales for modular and global events. If δt represents the delay period (i.e., the time required for signals to travel from one node to another through a link), the coupling terms of Eq (2.8) become:

$$\sum_{j=1}^{N} \frac{K_{ij}}{k_i}[x_j(t - \delta t) - x_i(t)]. \tag{2.10}$$

For constant delay (i.e., δt = constant, for all pairs of connected nodes), we observe in Fig. 2.8 (b) that the time required for modular synchronization (τ_{ms}) is shorter than that required for global synchronization (τ_{gs}), although in general both are longer than their corresponding values in the absence of any delay ($\delta t = 0$). We also consider the case where coupling delays are random and chosen from an uniform distribution. As in the case of coupling strengths κ, the delays may differ for

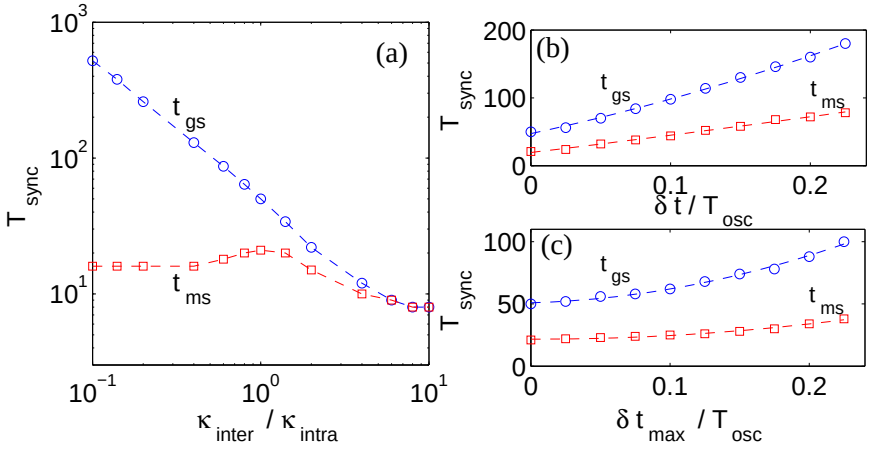


Figure 2.8: (a) Dependence of the two time-scales corresponding to modular (t_{ms}) and global synchronization (t_{gs}) on the ratio of the inter and intra modular coupling strengths ($\kappa_{inter}/\kappa_{intra}$). (b) The two synchronization time-scales shown as a function of a constant delay δt between any pair of connected oscillators. (c) Variation of t_{ms} and t_{gs} with random inter-modular coupling delays, that are distributed uniformly between $[0, \delta t_{max}]$. In this case, there is no delay for intra-modular couplings. Note that, T_{osc} is the time-period for an uncoupled relaxation oscillator. (In all cases $N = 512$, $\langle k \rangle = 14$, $m = 16$ and $r = 0.02$).

connections between nodes belonging to the same module as opposed to those belonging to different modules. For example, this may arise if nodes within a module are geographically closer to each other, relative to nodes in other modules. Therefore, we look at the case when there is no coupling delay within a module, while, the delay for connections between oscillators in different modules is distributed over the interval $[0, \delta t_{max}]$. In Fig. 2.8 (c), we observe that as in the case of constant delay, the inter-modular synchronization takes significantly longer time than intra-modular synchronization, emphasizing the generality of our results.

2.3.3 Diffusion

The existence of such distinct time-scales as a consequence of modular structure also appears in other dynamical processes, e.g., diffusion. Consider a discrete random walk on a network, where the walker moves from one node to a randomly chosen neighboring node at each time step. We analyze the time-evolution of the diffusion process by obtaining the distribution of first passage times for random

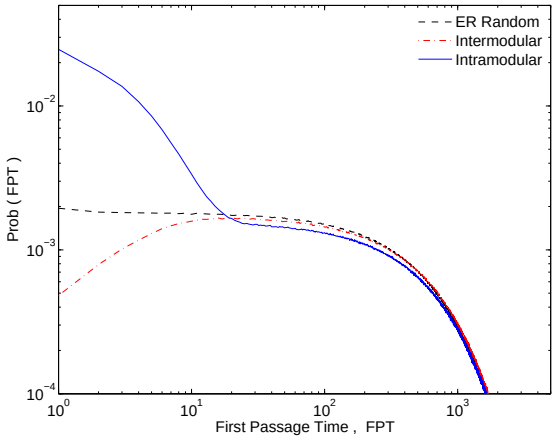


Figure 2.9: The distribution of first passage times (FPT) for diffusion process among the nodes in modular ($m = 16, r = 0.02$) and ER random networks. When the source and the target nodes belong to the same module, the FPT has a much higher probability of being small than when the nodes belong to different modules. The distribution of FPT for a homogeneous random network is also shown for comparison. This indicates the existence of two distinct time scales for random spreading in modular networks, the diffusive process within a module taking place much faster compared to diffusion between modules. For all networks, $N = 512$ and $\langle k \rangle = 14$.

walkers to reach a target node in the modular random network, starting from a source node [81]. Fig. 2.9 shows that this distribution differs quite significantly depending on whether the target node belongs to the same module as the source node or in a different module. This again suggests two distinct time-scales, with intra-modular diffusion occurring much faster than inter-modular diffusion. This is consistent with the results of Refs. [71, 82] where the degree of isolation of a module was assessed by comparing the participation of its nodes in different diffusion modes, using the internet as an example.

2.4 Linearized dynamics: Laplacian analysis

To understand the existence of two distinct time scales in a modular network, we consider the linearized dynamics around the synchronized state,

$$\frac{d\theta_i}{dt} = -\frac{\kappa}{k_i} \sum_j L_{ij}\theta_j, (\forall i = 1, \ldots, N) \tag{2.11}$$

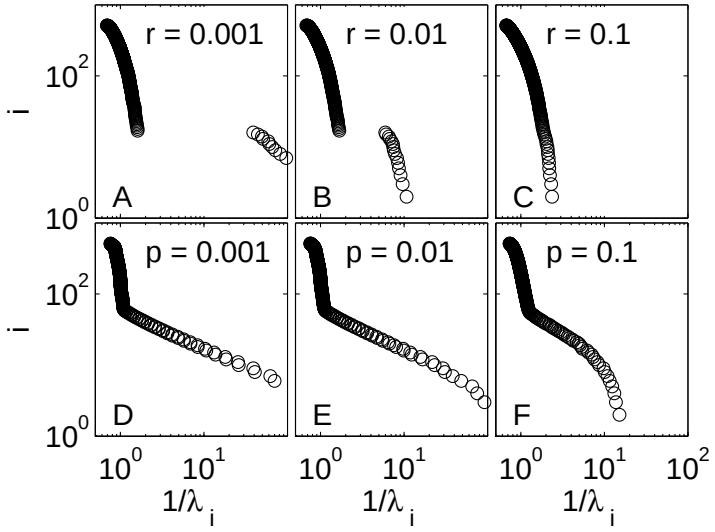


Figure 2.10: Rank index i plotted against the inverse of the corresponding eigenvalue of the Laplacian matrix $\mathscr{L}$ for modular random network ($m = 16$) at different r (A-C) compared with that of WS network at different p (D-F), indicating the existence of a distinct spectral gap in the former at low r ($N = 512$, $\langle k \rangle = 14$).

where $\mathbf{L}$ is the Laplacian matrix of the network, with $L_{ii} = k_i$ and $L_{ij} = -A_{ij}$ $(i \neq j)$. Solving in terms of the normal modes $\varphi_i(t)$, we get

$$\varphi_i(t) = \sum_j B_{ij}\theta_j = \varphi_i(0)\exp^{-\lambda_i t}, \tag{2.12}$$

where λ_i are the eigenvalues of $\mathbf{L}' = \mathbf{D}^{-1}\mathbf{L}$ ($\mathbf{D}$ being a diagonal matrix with $D_{ii} = k_i$), and $\mathbf{B}$ is the matrix of its eigenvectors. All the eigenvalues are real as $\mathbf{L}'$ is related to the symmetric normalized Laplacian $\mathscr{L} = \mathbf{D}^{\frac{1}{2}}\mathbf{L}'\mathbf{D}^{\frac{-1}{2}}$ through a similarity transformation. Any difference in the time scales of the different modes is manifested as gaps in the spectrum of $\mathscr{L}$, revealing different topological scales of the network. The mode corresponding to the smallest eigenvalue is associated with global synchronization, while other modes provide information about synchronization within different groups of oscillators. We observe a gap in the Laplacian spectrum for modular random networks that increases with decreasing value of r [Fig. 2.10 (A-C)] indicating that the very different time-scales for synchronization at the global and local levels originate from the modular organization

of the network structure. This is further supported by the absence of a similar gap in the Laplacian spectra for WS networks, shown at different values of p in Fig. 2.10 (D-F).

To relate this analysis with the diffusion process, we note that the transition probability from node i to j at each step of the random walk is $P_{ij} = A_{ij}/k_i$. This transition matrix $\mathbf{P}$ is related to the normalized Laplacian of the network as $\mathscr{L} = \mathbf{I} - \mathbf{D}^{\frac{1}{2}}\mathbf{P}\mathbf{D}^{\frac{-1}{2}}$, where $\mathbf{I}$ is the identity matrix [71]. The eigenvalues of $\mathbf{P}$ are all real, the largest being 1 while the others are related to the different diffusion timescales. As in the synchronization example, the spectrum of $\mathbf{P}$ for modular random network exhibits a gap reflecting the existence of distinct timescales in the system. Note that, although the above result strictly applies only when linear approximation is valid, we observe the property of time-scale separation predicted for modular networks to be a much more general phenomenon. In particular, the strong nonlinear interactions of the Ising model cannot be even approximately treated by the Laplacian analysis. Nevertheless, we see almost identical behavior for all three processes, indicating the universality of the dynamical signature of modular networks.

2.4.1 Laplacian analysis of cortical networks

In order to provide empirical evidence for the above distinction between dynamical behavior of the different small-world models, we have considered the connectivity data for cortical areas in the brains of the cat [83] and the macaque [84]. Such networks have been reported to have small-world structural properties [63]. As previously mentioned, local synchronization within a cluster has functional importance in the brain, whereas global coherence of activity may be undesirable. The theoretical arguments given above would, therefore, imply a modular structural organization for the connections between the cortical areas. This would be visibly manifested through the existence of gaps in the Laplacian spectra of the empirical networks, which is indeed what we observe [Fig. 2.11]. This strongly suggests that at least some of the empirically observed small-world networks that occur in nature may be organized in a modular fashion, and thus, have significantly different dynamical behavior from the WS or related models.

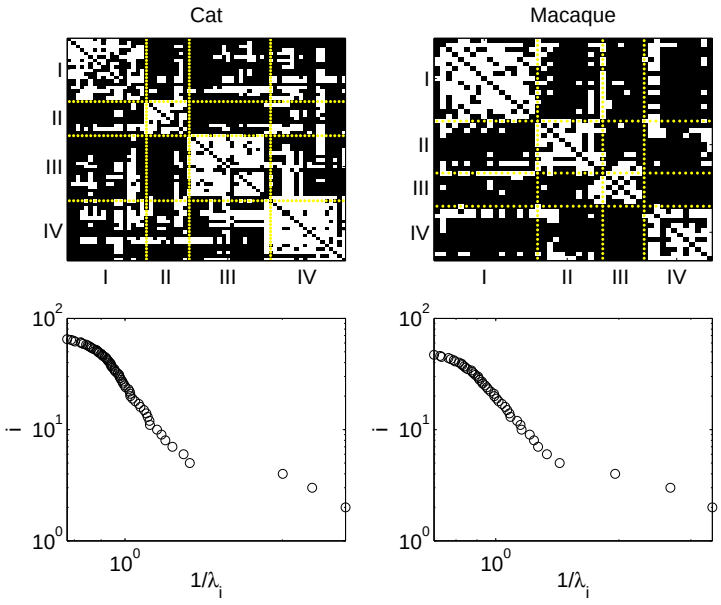


Figure 2.11: The adjacency matrix showing connections between different cortical areas in the cat (top left, $N = 65$) and macaque (top right, $N = 47$) cerebral cortex. The broken lines indicate clusters of cortical areas (labelled *I-IV*) that are densely connected within themselves. This structural division reflects, to some extent, the functional segregation among the different cortical areas (e.g., visual, somatosensory, etc.). The rank-ordered reciprocal eigenvalues of the corresponding Laplacian matrices (bottom) show well-defined spectral gaps, consistent with the existence of a modular structure for the cortico-cortical networks.

2.5 Discussion

In this chapter, we have shown that modular networks, where links within each module are much more numerous than those between different modules, can exhibit all the structural features associated with small-world networks even in the absence of a regular lattice substrate. By using a simple model, where nodes connect to each other at random within each community, we show that a modular organization can give rise to static properties (such as clustering or communication efficiency) almost identical to the widely-used WS model for small-world networks [56]. Note that, it is the modular organization which is crucial here, as the network structure within each module is irrelevant for our conclusions to be valid. Such modularity may arise in nature through multi-constraint optimization to which most networks occurring

in the real world are subjected [85]. This is discussed in detail in Chapter 3 of this thesis. The dynamical behavior of modular networks exhibits the striking feature of multiple, distinguishable time-scales corresponding to (a) fast intra-modular and (b) slow inter-modular processes, which is quite different from the behavior seen in WS model of small-world networks.

Empirical evidence for such behavior in cortico-cortical networks indicate that several systems for which small-world properties have been reported may indeed have modular organization with the associated dynamical signature. The increasing recognition that small-world networks underlie processes of vital importance to society, such as epidemics spreading through a few long-range links (e.g., the airline network that is instrumental in spreading a disease like SARS [86]), makes it of vital importance to understand the structural topology of a network that is responsible for the SW property. As different structures can result in distinct collective dynamical behavior, it is important to go beyond macroscopic measures (such as average path length) and focus on the underlying arrangement of interactions in such networks. This is essential for intelligent intervention to prevent a local problem from rapidly evolving into a global threat as a result of uncontrolled spreading through the network.

3

Evolution of modular networks

As modular structures are ubiquitous in complex networks, it is of immense interest to understand how such systems can evolve towards a modular configuration. In many of these networks, there is a significant presence of *hubs*, i.e., nodes with large degree or number of connections to other nodes. Hubs are crucial for linking the nodes in real networks, which have extremely sparse connectivity, with the probability C of connection between any pair of nodes varying between 10^{-1} and 10^{-8} [4]. By contrast, random networks with such small C are almost always disconnected. Hubs can also lead to the "small-world" effect [37] by reducing the average path length of the network. We note that most modular systems are subject to multiple structural and functional constraints. Examples of such constraints include the minimization of average path length, as well as, the total number of links, while maximizing robustness against perturbations in node activity. In this chapter, we show that the optimal networks which satisfy all these three constraints are characterized by the existence of multiple modules sparsely connected to each other. In addition, these modules have distinct hubs resulting in an overall heterogeneous degree distribution.

The majority of previous studies on modular networks have been concerned with methods to identify community structure [52]. There have been relatively few attempts to explain the potentially more interesting question of how and why modularity emerges in complex networks. Most such attempts are based on the notion of evolutionary pressure, where a system is driven by the need to adapt to a changing environment [87, 88]. However, such explanations involve complicated adaptive mechanisms, in which the environment itself is assumed to change in a

modular fashion. Further, adaptation might decrease connectivity through biased selection of sparse networks, which eventually results in *disruption of the network* with the modules becoming isolated nodes [87] or disconnected parts [89]. More recently, a social network model has shown the emergence of isolated communities through the rearrangement of links to form groups with homogeneous opinion [90].

A crucial limitation of these studies is that they almost always focus on a single performance parameter. However, in reality, most networks have to optimize between several, often conflicting, constraints. While structural constraints, such as path length, had been the focus of initial work by network researchers, there has been a growing realization that most networks have dynamics associated with their nodes [12]. The robustness of network behavior is often vital to the efficient functioning of many systems, and also imposes an important constraint on networks. Therefore, the role played by dynamical considerations in determining the topological properties of a network is a challenging and important question that opens up new possibilities for explaining observed features of complex networks [91].

In Sec. 3.1, we propose a simple model for the emergence of modularity in networks as an optimal solution for satisfying a minimal set of structural and functional constraints. We explicitly show this by performing a multi-constraint optimization with simulated annealing in Sec. 3.2. In Sec. 3.3, we show that while robustness is indeed necessary, it is not enough by itself to generate modularity, contrary to what is generally believed. We end the chapter with a discussion in Sec. 3.4, on how these modular networks are also structurally robust with respect to simultaneous targeted and random attacks.

3.1 Constraints on networks

Cost

Networks are subject to certain structural constraints. One of the structural constraint is the need to save resources, manifested in minimizing *link cost*, i.e., the cost involved in building and maintaining each link in a network [92]. This results in the network having a small total number of links, L.

Efficiency

However, such a procedure runs counter to another important consideration of reducing the average path length ℓ, which improves the network *efficiency* by increasing communication speed among the nodes [36].

The conflict between the above two criteria can be illustrated through the example of airline transportation networks. Although, fastest communication (i.e., small ℓ) will be achieved if every airport is connected to every other through direct flights, such a system is prohibitively expensive as every route involves some cost in maintaining it. In reality, therefore, one observes the existence of airline hubs, which act as transit points for passengers arriving from and going to other airports.

Stability

Another important constraint is to decrease the instability of dynamical states associated with the network. We investigate the dynamical stability of a network composed of N nodes, which are self regulating when isolated, by measuring the growth rate of a *small* perturbation $\mathbf{x}$ about an equilibrium state of the network dynamics. Although the system can be nonlinear in general, the dynamics of such perturbations are described by a *linear* system of coupled differential equations $\dot{x_i} = \sum_{j=1}^{N} J_{ij} x_j$. The stability of the equilibrium is then determined by the largest real part $\lambda_{\max}$ of the eigenvalues for the matrix $\mathbf{J}$ representing the interactions among the nodes. The perturbation decays if $\lambda_{\max} < 0$, and increases otherwise, at a rate proportional to $|\lambda_{\max}|$. Thus, minimizing λ_{max} makes the equilibrium less unstable, which is important for many systems including ecological networks [93].

Here $J_{ii} = -1\ \forall i$ such that we only consider instability induced through network interactions. The off-diagonal matrix elements $J_{ij} (\sim A_{ij} W_{ij})$ include information about both the topological structure of the network, given by the adjacency matrix $\mathbf{A}$ (A_{ij} is 1, if nodes i, j are connected, and 0, otherwise; $A_{ii} = 0\ \forall i$), as well as, the distribution of interaction strengths W_{ij} between nodes. In our simulations, W_{ij} has a Gaussian distribution with zero mean and variance σ^2; however, a nonzero mean does not qualitatively change our results. For an Erdos-Renyi (ER) random network, $\mathbf{J}$ is a sparse random matrix, with $\lambda_{\max} \sim \sqrt{NC\sigma^2} - 1$, according to the May-Wigner theorem [93]. Therefore, increasing the system size N, connectivity C or interaction strength σ, results in instability of the network. This result has been

shown to be remarkably robust with respect to various generalizations [94, 95, 96]. Further, for uniform coupling strength, λ_{max} is inversely related to the epidemic propagation threshold for the network [97], and hence, minimizing λ_{max} also makes the network more robust against spreading of infection.

3.2 Modularity through multi-constraint optimization

3.2.1 Minimum link-cost constraint $(L = N - 1)$

For ER random networks, although ℓ is low, L is high because of the requirement to ensure that the network is connected: $L > N\ln N$ [42]. Introducing the constraint of link cost (i.e., minimizing L) while requiring low average path length ℓ, leads to a starlike connection topology (Fig. 3.1C). A *star network* has a single hub to which all other nodes are connected, there being no other links. Its average degree $\langle k \rangle \approx 2$ is *non extensive* with system size, and is much smaller than a connected random network, where $\langle k \rangle \sim \ln N$. However, such starlike networks are extremely unstable with respect to dynamical perturbations in the activity of their nodes. The probability of dynamical instability in random networks increases only with average degree ($\lambda_{\max} \sim \sqrt{\langle k \rangle}$, since $\langle k \rangle = NC$), while for star networks it increases with the largest degree, and hence the size of the network itself ($\lambda_{\max} \sim \sqrt{N}$). To extend this for the case of weighted networks we look at the largest eigenvalue of $\mathbf{J}$, $\lambda_{max} = -1 + \sqrt{\sum_{i=2}^{N} J_{1i}J_{i1}}$, the hub being labeled as node 1. The stability of the weighted star network is governed by $\sum_{i=2}^{N} J_{1i}J_{i1}$, which is the displacement due to a 1-dimensional random walk of $N-1$ steps whose lengths are products of pairs of random numbers chosen from a $Normal\ (0, \sigma^2)$ distribution.

Simulated annealing

To obtain networks which satisfy the dynamical as well as the structural constraints we perform optimization using simulated annealing, with a network having N nodes and $N-1$ unweighted links (the smallest number that keeps the network

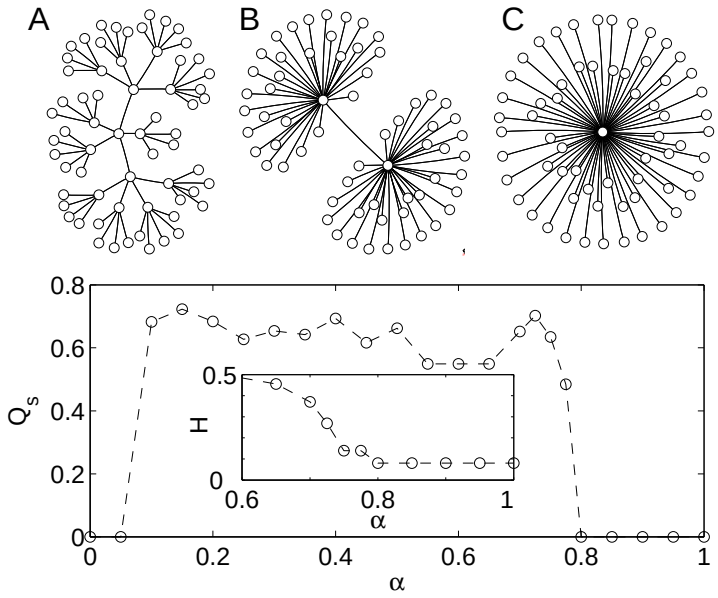


Figure 3.1: The optimized network structures for a system with $N = 64$ nodes and $L = N - 1$, at different values of α: (A) 0.4, (B) 0.775 and (C) 1. For $\alpha = 0$ the optimal network is a 1-dim chain. (Bottom) The modularity Q_s of the optimized network for different α, when each module is a community defined in the strong sense. The transition to star configuration occurs around $\alpha \simeq 0.8$, as observed in the variation of degree entropy H with α.

connected). Having fixed L, the energy function to be minimized is defined as

$$E(\alpha) = \alpha\ell + (1 - \alpha)\lambda_{\max},$$

where the parameter $\alpha \in [0, 1]$ denotes the relative importance of the path length constraint over the condition for reducing dynamical instability. Rewiring is attempted at each step and is (i) rejected if the updated network is disconnected, (ii) accepted if $\delta E = E_{final} - E_{initial} < 0$, and (iii) if $\delta E > 0$, then accepted with probability $p = \exp(-\delta E/T)$, where T is the "temperature". The initial temperature was chosen in such a way that energetically unfavorable moves had 80% chance of being accepted. After each monte carlo step (N updates) the temperature was reduced by 1% and iterated till there was no change in the energy for 20 successive monte carlo steps. For each value of α, the optimized network with lowest E was obtained from 100 realizations.

Emergence of modular configuration

As can be seen from Fig. 3.1, modularity emerges when the system tries to satisfy the twin constraints of minimizing ℓ as well as $\lambda_{\max}$. When α is very high (~ 0.8) such that the instability criterion becomes less important, the system shows a transition to a starlike configuration with a single hub. However, as α is decreased, the instability of the hub makes the star network less preferable and for intermediate values of α, the optimal network gets divided into modules, as seen from the measure of network modularity, Q [98]. To obtain a robust partitioning of the network, we consider modules to be communities defined in the *strong* sense, i.e., each node i belonging to a community has more connections with nodes within the community than with the rest of the network [77]. The resulting modularity measure Q_s is high for a modular network, whereas for homogeneous, as well as, for starlike networks, $Q_s = 0$. To determine the communities, we

1. Compute the betweenness measure for all edges and remove the one with highest score:

2. (a) if it results in splitting the network (or subnetwork) into communities in the strong sense, then the resulting Q_s is computed;

   (b) if not, we go back to step (1) and remove the edge with the next highest score.

The process is carried out iteratively until all edges of the network have been considered. Note that, in step (2a), checking whether the splitting results in communities in the strong sense is considered with respect to the full network. We verified these results by also calculating Q_s with the network modules determined through stochastic extremal optimization [99]. The transition between modular and star structures is further emphasised in the behavior of the degree entropy defined as

$$H = \sum_{k}^{N-1} p_k \log p_k, \tag{3.1}$$

where p_k is the probability of a node having degree k. The network entropy provides an average measure of the network's heterogeneity, since it measures the diversity of the link distribution [100]. Two extreme cases are the maximal value and the minimal one. The maximum value is $H_{\max} = \log(N-1)$ obtained for

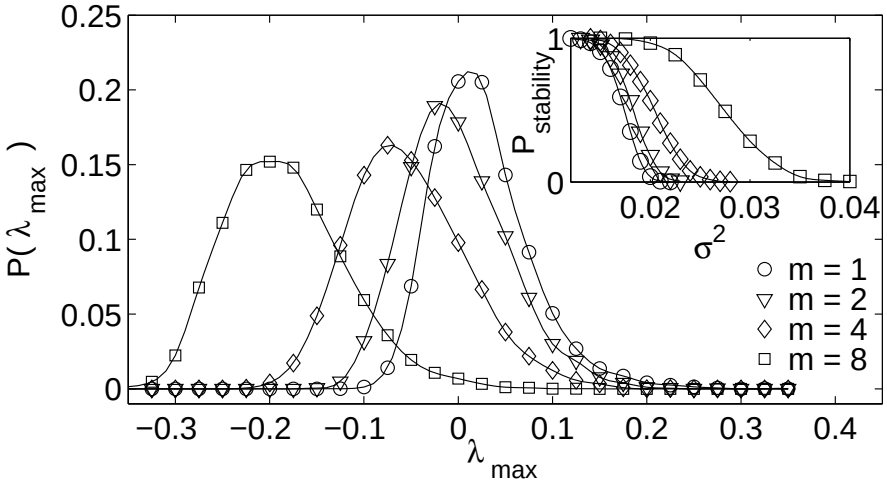


Figure 3.2: Probability distribution of $\lambda_{\max}$ for a clustered star network ($N = 256, L = 15N$) with different numbers of modules, m. Modules of equal size are connected by single link between respective hubs. Link weights W_{ij} follow a $Normal\ (0, \sigma^2)$ distribution with $\sigma^2 = 0.018$. (Inset) Probability of stability $[P(\lambda_{\max} < 0)]$ varying with σ^2. Increasing m results in the transition to instability occurring at higher σ^2, implying that network stability increases with modularity.

$p_k = \frac{1}{N-1}\ \forall k = 1, 2, \cdots, N-1$ and minimum value $H_{\min} = 0$ occurs when $p_k = 0, \cdots, 1, \cdots, 0$. The emergence of a dominant hub at a critical value of α is marked by H reducing to a low value.

Why modular configuration evolves?

To understand why modular networks emerge on simultaneous optimization of structural and functional constraints we look at the change in stability that occurs when a star network is split into m modules, the modules being connected through links between their hubs. The largest eigenvalue for the entire system of N nodes is the same as that for each isolated module, $\lambda_{\max} \sim \sqrt{N/m}$, as the additional effect of the few intermodular links is negligible. At the same time, the increase in the average path length ℓ with m is almost insignificant. Therefore, by dividing the network into a connected set of small modules, each of which is a star subnetwork, the instability of the entire network decreases significantly while still satisfying the structural constraints.

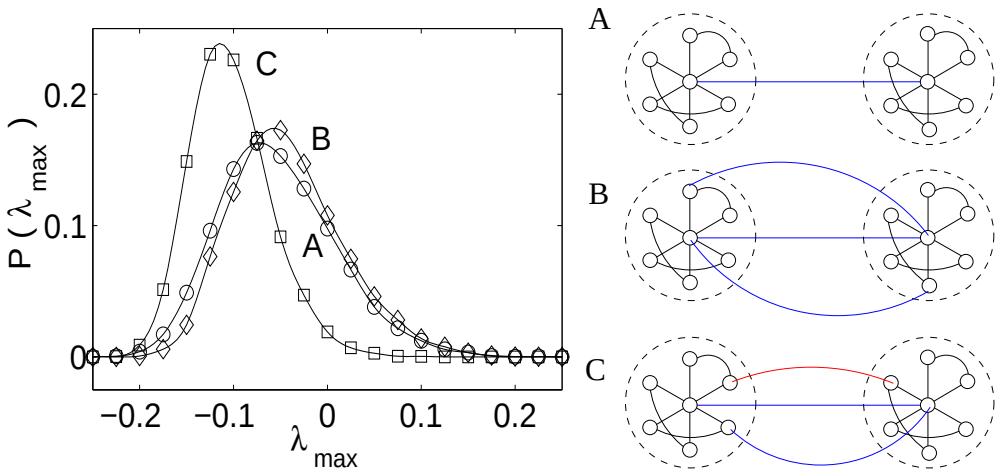


Figure 3.3: Probability distribution of $\lambda_{\max}$ for clustered star networks ($N = 256, L = 15N$) having four modules with different types of intermodular connectivities (A), (B) and (C), which are represented schematically here. Link weights W_{ij} have a $Normal\ (0, \sigma^2)$ distribution with $\sigma^2 = 0.018$.

3.2.2 Relaxing the link-cost constraint $(L > N - 1)$

The above results were obtained for a specific value of L $(= N - 1)$. We now relax the constraint on link cost and allow a larger number of links than that strictly necessary to keep the network connected. The larger L is manifested as random links between nonhub nodes, resulting in higher clustering within the network. Even for such *clustered star* networks, $\lambda_{\max}$ increases with size as $\sqrt{N}$, and therefore, their instability is reduced by imposing a modular structure (Fig. 3.2). The effect of increasing the number of modules, m, on the dynamical stability of a network can be observed from the stability-instability transition that occurs on increasing the network parameter σ keeping N, C fixed. The critical value at which the transition to instability occurs, σ_c, increases with m (Fig. 3.2, inset) while ℓ does not change significantly. This signifies that even for large L, networks satisfy the structural and functional constraints by adopting a modular configuration.

As L is increased, we observe that the additional links in the optimized network occur between modules, in preference to, between nodes in the same module. To see why the network prefers the former configuration, we consider three different types of intermodular connections: (A) only the hub nodes of different modules

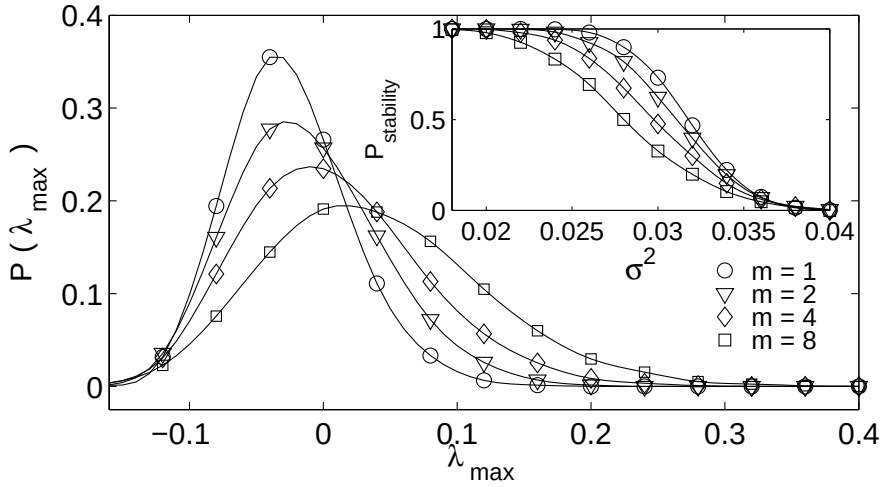


Figure 3.4: Probability distribution of $\lambda_{\max}$ for random networks ($N = 256, L = 15N$) as a function of the number of modules, m, which are connected to each other by single links. Link weights W_{ij} follow $Normal\ (0, \sigma^2)$ distribution with $\sigma^2 = 0.03$. The inset shows the probability of stability $[P(\lambda_{\max} < 0)]$ varying with σ^2. Increasing m results in transition to instability at lower σ^2, indicating that increasing modularity decreases stability for random networks.

are connected, (B) nonhub nodes of one module can connect to the hub of another module, and (C) nonhub nodes of *different* modules are connected. Arrangement (B) where intermodular connections that link to hubs of other modules actually increase the maximum degree in the modules, making this arrangement more unstable than (A). On the other hand, (C) connections between nonhub nodes of *different* modules not only decrease the instability (Fig. 3.3), but also reduce ℓ. As a result, the optimal network will always prefer this arrangement (C) of large number of random intermodular connections over other topologies for large L.

3.3 Robustness and modularity

Our observation that *both* structural and dynamical constraints are necessary for modularity to emerge runs counter to the general belief that modularity necessarily follows from the requirement of robustness *alone*, as modules are thought to limit the effects of local perturbations in a network. To further demonstrate that the three constraints are the minimal required for a network to adopt a modular

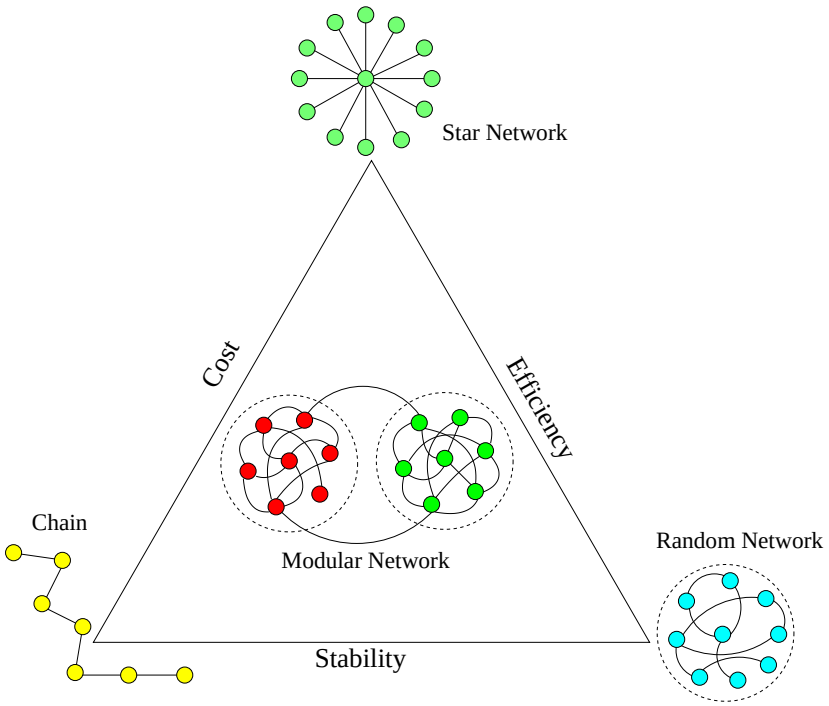


Figure 3.5: Schematic diagram indicating the types of optimal networks obtained by satisfying different constraints. Each vertex represents networks obtained by satisfying a pair of constraints indicated on the two sides of the triangle that meet at that vertex. Note that, modular networks emerge by optimizing all three constraints, viz., cost, efficiency and stability, indicated by the three arms of the triangle.

configuration, we remove the hub from a clustered star while ensuring that the network is still connected. This corresponds to the absence of the link cost constraint altogether and the optimal graph is now essentially a random network.

To see why modularity is no longer observed in this case, we consider the stability of an ER random network on which a modular structure has been imposed. A network of N nodes is divided into m modules, connected to each other with a few intermodular links. We then consider the stability-instability transition of networks for increasing m, with the average degree, $\langle k \rangle$ kept fixed. Although from the May-Wigner theorem, it may be naively expected that $\sigma_c \simeq 1/\sqrt{\langle k \rangle}$ is constant w.r.t. m, we actually observe that increasing m decreases stability (Fig. 3.4). This is because when a network of size N is split into m modules, the stability of the entire network is decided by that of the most unstable module, ignoring the small additional effect of intermodular connections. Thus, the stability of the

entire network is decided by randomly drawing m values from the distribution of $\lambda_{\max}$ for the modules. Therefore, for modular networks it is more likely that a positive $\lambda_{\max}$ will occur, than for the case of a homogeneous random network of size N [101]. The *decrease* of stability with modularity for random networks shows that, in general, it is not necessary that modularity is always stabilizing and results in a robust network, as has sometimes been claimed [87].

3.4 Discussion

In this chapter we have shown that modules of interconnected nodes can arise as a result of optimizing between multiple structural and functional constraints. In particular, we show that, by minimizing link cost as well as path length, while simultaneously increasing robustness to dynamical perturbations, a network will evolve to a configuration having multiple modules characterized by hubs, that are connected to each other (Fig. 3.5). At the limit of extremely small L (total number of links in the network), this results in bimodal degree distribution, that has been previously shown to be robust with respect to both targeted and random removal of nodes [102]. Therefore, not only are such modular networks dynamically less unstable, but they are also robust with respect to structural perturbations. In general, on allowing larger L, the optimized networks show heterogeneous degree distribution that has been observed in a large class of networks occurring in the natural and social world, including those termed as scale-free networks [3]. Thus, our results provide a glimpse into how the topological structure of complex networks can be related to functional and evolutionary considerations.

4

Hierarchical organization in complex networks

Complex networks exhibit many common organizational features at the mesoscopic level. Apart from modularity, which has been discussed in the previous chapters, many systems also show *hierarchical* ordering of their nodes. In other words, they are composed of successive layers of interconnected or nested communities. Such structural hierarchy not only describes how nodes link to form communities, but also, how communities join with each other to form the entire network which may exhibit multiple levels of larger meso-level structures, such as meta-modules. In the literature, often the terms hierarchy and modularity are used inter-changeably, although, as shown in Fig. 4.1, they represent distinct properties of the network. This confusion in usage could have stemmed from the fact that these two properties are found to coexist in many networks occurring in real life [55, 103, 69, 19], including technological networks such as the Internet [8, 71] and biological networks, like that of cortical areas in the mammalian brain [104].

As discussed in the previous chapters, most complex systems seen in real life also have associated dynamics [12]. The structural properties of such networks have been sought to be linked with their dynamical behavior [6, 105]. In this connection, one of the questions of obvious significance is whether there is a relation between the *stability* of the system dynamics (with respect to small perturbations in the variables describing the state of the nodes) and the specific topological arrangement of connections in the network. Such robustness is necessary if complex systems are to survive in the noisy environment that characterizes the real world.

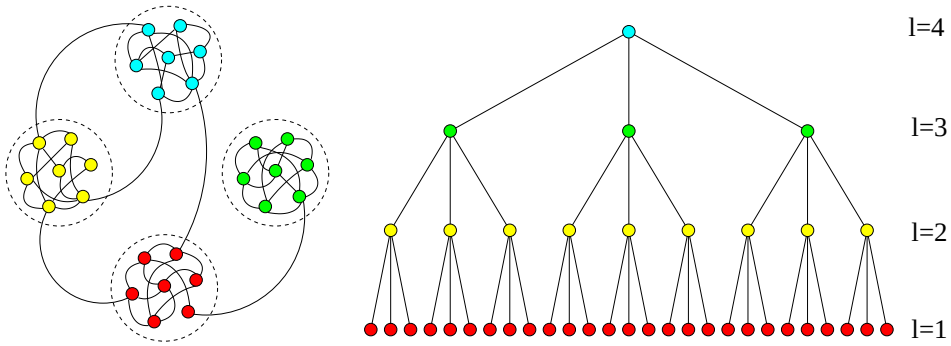
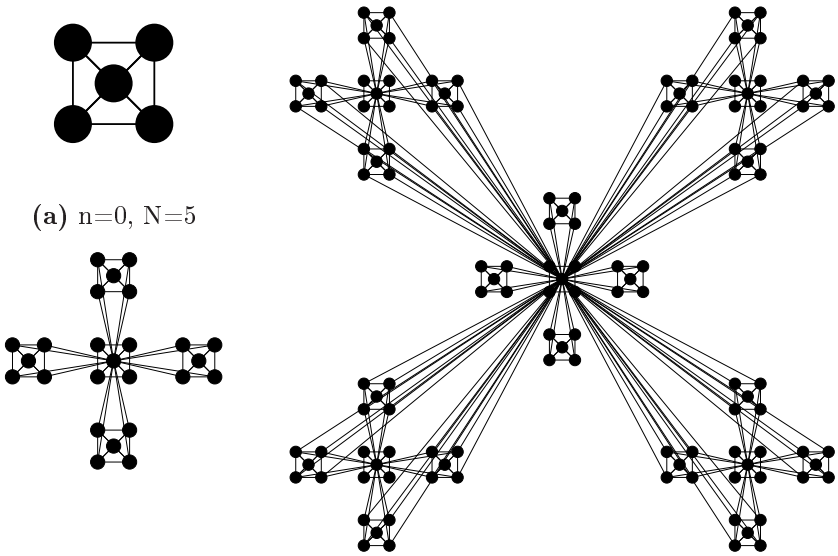


Figure 4.1: Schematic diagrams of (left) a modular network, with modules demarcated by broken circles, and (right) a hierarchical network with 4 levels, each indicated by a level number l.

It has sometimes been argued that, networks with larger number of nodes, links and stronger inter-connections are more stable [106]. On the other hand, theoretical results on the stability of model networks, e.g., the May-Wigner theorem, suggest the opposite [93]. However, as these results are based on the study of networks whose connection topology shows none of the structures that are seen in real life networks, in particular, modularity and hierarchy, it is of interest to see whether introducing hierarchical organization and modular structures can reveal limitations in the validity of May-Wigner theorem. We study this problem by proposing a network model that exhibits both these properties and observing the local stability of the system dynamics with respect to perturbations. We also consider the stability of synchronization over the network, as the issue of network synchronization has assumed importance in recent years, owing to its connection with, e.g., brain dynamics [104].

In Sec. 4.1, we describe earlier models that have been proposed to describe hierarchical organization in networks, in particular, the Ravasz-Barabasi deterministic model [107]. In Sec. 4.2, we propose an alternate model that allows a detailed study of the relation between dynamical stability and hierarchical modular organization of the network. We also show that the occurrence of a scaling relation between clustering and degree of the nodes cannot be considered as a signature for the existence of hierarchical modular structure. This is contrary to what has been claimed in Ref. [107] and tacitly assumed in many subsequent studies [108]. In Sec. 4.3 we observe that both hierarchy and modularity actually increase the



(a) n=0, N=5

(b) n=1, N=25

(c) n=2, N=125

Figure 4.2: The Ravaz-Barabasi model of hierarchical scale-free network showing the first three steps in an iterative construction procedure leading to hierarchical network structure. (a) A fully connected cluster consisting of $N = 5$ nodes, (b) a network with $N = 25$ nodes, and (c) a network with $N = 125$ nodes. (From Ref. [107])

instability compared to an equivalent random network. This may appear counter-intuitive as both these structural properties are observed in networks occurring in nature, which necessarily have to be robust to survive environmental fluctuations. However, as noted in the preceding chapter, the emergence of modular structures can be understood as a response to multiple (and often conflicting) constraints imposed on such networks [85]. In Sec. 4.4, we discuss how these observations can be extended to explain the emergence of hierarchical organization in networks. We conclude with a short discussion of the importance of constraints related to physical space on which networks are embedded in Sec. 4.5.

4.1 Hierarchical networks: Ravasz-Barabasi model

One of the most cited models for hierarchical modular networks is a deterministic model proposed by Ravasz and Barabasi (RB) [107]. This model generates a set

of inter-nested modules in a hierarchical fashion using a deterministic procedure that has both high clustering (because of the modular nature of the network at the most fundamental level) and a scale-free degree-distribution.

This model is constructed as follows: Initially, a fully connected cluster of five nodes is constructed (Fig. 4.2 a). Next, four replicas of this hypothetical module are generated and the four external nodes of the replicated clusters are connected to the central node of the old cluster, obtaining a large 25-node module (Fig. 4.2 b). Subsequently, four replicas of this 25-node module are generated, and the 16 peripheral nodes are connected to the central node of the old module (Fig. 4.2 c), obtaining a new module of 125 nodes. These replications and connection steps are repeated, increasing the number of nodes in the system by a factor of five at each iteration.

In the RB model, a scaling relation is observed between the clustering coefficient of a node C and its number of connections (i.e., degree) k:

$$C(k) \sim k^{-1}. \tag{4.1}$$

Similar relations were also observed in several real networks, such as the web of semantic connections between two English words which are synonyms [107]. This occurrence of the scaling relation between clustering and degree of the nodes in a network has often been taken as a signature for the existence of hierarchical modular structure in that network. However, recently, this scaling relation was shown to be actually an outcome of degree-correlation bias in the usual definition of clustering coefficient [109].

It can be easily seen that this scaling relation is not a necessary indicator for the existence of either modularity or hierarchy. For example, consider a modular network consisting of N nodes and m modules of equal size. Let each node have degree k, with the links initially occurring exclusively between nodes belonging to the same module (i.e., the modules are isolated from each other). To make the network connected we rewire a small fraction of the links keeping the degree of each node fixed. Plotting clustering as a function of degree for this network will only show vertical spread of points at a single node degree value. Let us consider another example, this time a hierarchical structure, viz., the Cayley tree with b branches at each vertex. Again, it is easy to see that the clustering versus degree

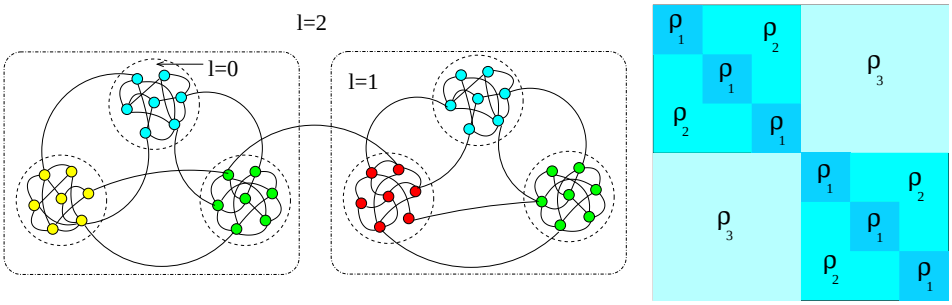


Figure 4.3: Schematic diagram of the hierarchical modular network model (left) with the modules occurring at the various hierarchical levels (l) indicated by broken lines, and the corresponding adjacency matrix (right) where ρ_1 indicates the density of connections within and, ρ_{l+1}, between the different modules at each level l.

curve will not show the characteristic scaling seen for the RB model. In fact, in the next section, we show that even for networks where both hierarchy and modularity are present, it is not necessary that this scaling relation between clustering and node degree will hold.

4.2 Hierarchical modular networks: A model

Here we propose a general model for networks having modular as well as hierarchical structure. Let us begin with a modular network consisting of m modules, each containing n nodes. The connectivity (i.e., the probability of a link between any pair of nodes) within each module is ρ_1, while the connectivity between modules is ρ_2 ($\leq \rho_1$). We now introduce hierarchy by adding another set of m modules (each having n nodes) with the same ρ_1 and ρ_2. The nodes belonging to these two different sets of modules are now connected, but with a probability ρ_3 ($\leq \rho_2$). The resulting network has $2nm$ nodes and $l = 2$ hierarchical levels (Fig. 4.3). To increase the number of hierarchical levels to $l = 3$, we add a similar network with $2nm$ nodes to the existing network and, as above, add links between these two networks with a probability ρ_4 ($\leq \rho_3$). Thus, to get a network with $l = h$ hierarchical levels, the above procedure is repeated $h - 1$ times. The final network contains $M = 2^{h-1}m$ number of modules. Note that, all connections between nodes are made randomly. To reduce the number of model parameters, we assume that the

connectivities $\rho_1, \ldots, \rho_{h+1}$ are related as:

$$\frac{\rho_2}{\rho_1} = \frac{\rho_3}{\rho_2} = \cdots = \frac{\rho_{h+1}}{\rho_h} = r, \tag{4.2}$$

where, $0 \leq r \leq 1$, the ratio of inter-modular connections between two successive hierarchical levels, is a control parameter. By varying r, one can switch between isolated modular ($r = 0$) and homogeneous random ($r = 1$) networks, with intermediate values of r giving hierarchical modular networks. We compare between networks having different number of hierarchical levels h, keeping the total number of modules M and average degree $\langle k \rangle$ fixed.

To consider the effect of hierarchy in isolation, while keeping modularity fixed (e.g., as measured by the Newman modularity measure Q [72]), we use a variant of the above model, where, $\rho_1 =$ constant, while other connectivities are still related by

$$\frac{\rho_3}{\rho_2} = \cdots = \frac{\rho_{h+1}}{\rho_h} = r. \tag{4.3}$$

This implies that the average number of intra-modular ($\langle k_{\text{intra}} \rangle$) and inter-modular ($\langle k_{\text{inter}} \rangle$) connections per node are also constant [1].

The stochastic construction procedure of this network, along with the ability to vary modularity (by changing r) independently of the number of hierarchical levels (h), makes it an extremely general model. In addition, as it is hierarchical by construction, we can show that the criterion suggested in Ref. [107], namely, the scaling relation between clustering and degree, is not a necessary condition for the existence of hierarchical modularity. As shown in Fig. 4.4 (left), when the modules are random networks, the scaling relation is clearly absent for our model network. To counter the possible argument that this failure of the relation is due to the non-scale-free degree distribution, we have also considered the case where each of the modules is a BA network. For the entire network, although the inter-modular connections are made randomly, the degree distribution is still scale-free. Even for this case, a clear scaling relation between clustering and degree is absent (Fig. 4.4 (right)).

[1] Note that, $\langle k_{\text{intra}} \rangle = \rho_1 \left(\frac{N}{M} - 1\right)$, and
$\langle k_{\text{inter}} \rangle = N\rho_2 \left[\frac{(m-1)}{M} + r\left(\frac{1}{2}\right)^{h-1} + \cdots + r^{h-2}\left(\frac{1}{2}\right)^2 + r^{h-1}\left(\frac{1}{2}\right)\right]$.

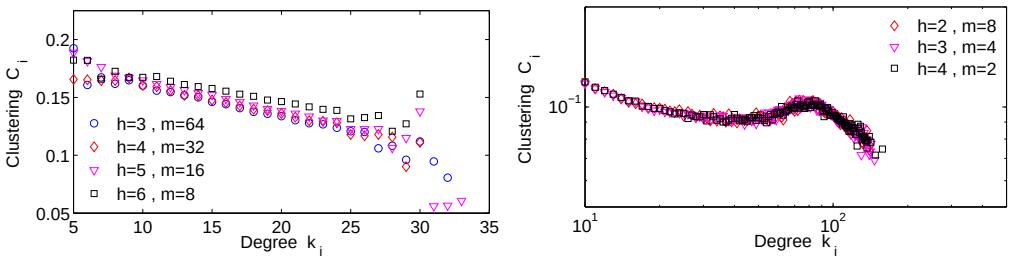


Figure 4.4: Clustering coefficient C_i of the i-th node as a function of its degree k_i for the hierarchical modular network model proposed here, where each module at $l = 1$ is (left) a random ER network and (right) a scale-free BA network. The different symbols indicate networks with differing total number of hierarchical levels, h. For both types of networks, the total number of nodes is $N = 8192$ with average intra-modular degree, $\langle k_{\rm intra} \rangle = 10$, inter-modular degree, $\langle k_{\rm inter} \rangle = 5$, and the ratio of inter-modular connections between two successive hierarchical levels, $r = 0.1$. Note that, in neither case is a scaling relation observed between C_i and k_i, although the modules are arranged in a hierarchical manner by construction.

4.3 Dynamics on hierarchical networks

4.3.1 Linear Stability of Equilibria

To look at the effect of hierarchy on network dynamics, we consider the linear stability of an arbitrarily chosen equilibrium state for a set of coupled differential equations defining the time-evolution of the system. For a network of N nodes, a dynamical variable x_i is associated with each node i. The state of the system, $\mathbf{x}$, can be characterized by $\dot{\mathbf{x}} = f(\mathbf{x})$, where f is a general nonlinear function. To investigate the stability around an arbitrary fixed point $\mathbf{x}^*$ (i.e., $f(\mathbf{x})|_{\mathbf{x}^*} = 0$), we check whether a small perturbation $\delta\mathbf{x}$ about $\mathbf{x}^*$ grows or decays with time. This perturbation evolves as

$$\dot{\delta\mathbf{x}} = \mathbf{J}\mathbf{x}, \tag{4.4}$$

where, $\mathbf{J}$ is the Jacobian matrix representing the interactions among the nodes: $J_{ij} = \partial f_i / \partial x_j|_{\mathbf{x}^*}$. As we are interested in the instability induced through the connections of the network, rather than the intrinsic instability of individual unconnected nodes, we can (without much loss of generality) set the diagonal element $J_{ii} = -1$. This implies that, in the absence of any connections, the nodes are self-

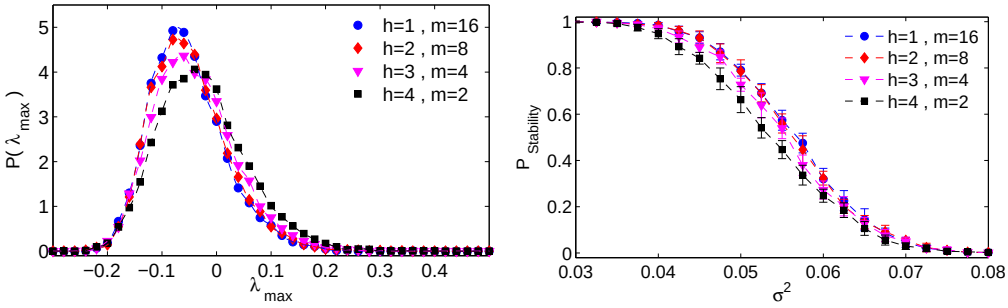


Figure 4.5: (Left) Probability distribution for the largest real part of the eigenvalues of the Jacobian J, as a function of total number of hierarchical levels, h (the interaction strength parameter, $\sigma^2 = 0.05$). (Right) Probability of stability for a hierarchical modular network as a function of σ^2, with different symbols corresponding to differing total number of hierarchical levels h. Link weights are chosen from a $normal(0, \sigma^2)$ distribution. For all cases, the network consists of $N = 256$ nodes with average intra-modular degree, $\langle k_{\text{intra}} \rangle = 10$, inter-modular degree, $\langle k_{\text{inter}} \rangle = 5$, and the ratio of inter-modular connections between two successive hierarchical levels, $r = 0.1$. At all hierarchical levels $l > 1$, the network is split into two sub-networks. At $l = 1$, each subnetwork is split into m modules ($l = 0$). Thus, $N = 256$ nodes are divided equally among $2^{h-1}m = 16$ modules, with the four curves corresponding to ($\square$) $h = 4$, $m = 2$, ($\bigtriangledown$) $h = 3$, $m = 4$, ($\diamond$) $h = 2$, $m = 8$, and ($\circ$) $h = 1$, $m = 16$. Note that, increasing h causes the transition to instability to occur at a smaller value of σ^2, implying that increasing hierarchy increases instability.

regulating, i.e., the fixed point $\mathbf{x}^*$ is stable. The behavior of the perturbation is determined by the largest real part, $\lambda_{\max}$, of the eigenvalues of $\mathbf{J}$. If $\lambda_{\max} > 0$, an initially small perturbation will grow exponentially with time, and the system will be rapidly dislodged from the equilibrium state $\mathbf{x}^*$.

The relation between the dynamical properties and the static structure of the network is provided by its adjacency matrix $\mathbf{A}$ (with $A_{ij} = 1$, if nodes i and j are connected, and 0 otherwise). There is a direct correspondence between the nature of the matrices $\mathbf{J}$ (specifying the dynamical behavior of perturbation) and $\mathbf{A}$ (which determines the structure of the underlying directed network), because $A_{ij} = 0$ implies $J_{ij} = 0$. In our model, we have generated J_{ij} by randomly choosing the non-zero elements from a Gaussian distribution with zero mean and variance σ^2. For Erdos-Renyi (ER) random networks, $\mathbf{J}$ is an unstructured random matrix and the largest real part of its eigenvalues, $\lambda_{\max} \sim \sqrt{N\rho\sigma^2} - 1$, where ρ is the connectivity of the network, and σ measures the dispersion of interaction strengths [93]. When

any of the parameters, N, ρ, or σ, is increased, there is a transition from stability to instability. The critical value at which the transition to instability occurs is $\sigma_c \sim 1/\sqrt{N\rho}$. This result, implying that complexity promotes instability, has been shown to be remarkably robust with respect to various generalizations [94, 95, 96, 110].

Here, using the above formalism, we examine the effect of hierarchy on the stability of equilibria when one of the network parameters (namely, σ) is varied. We study the critical value at which the transition to instability occurs, σ_c, as a function of the total number of hierarchical levels, h, keeping the total number of modules M fixed. We find that, with increasing h, the distribution of λ_{max} shifts towards more positive values (Fig. 4.5, left). As the system becomes unstable when $\lambda_{max} > 0$, it follows that the probability of stability for the network decreases with increasing number of hierarchical levels (Fig. 4.5, right).

4.3.2 Synchronization

It is of interest to look not only at the stability of equilibria for network dynamics, but also at the stability of synchronized activity in networks. Let us consider a network of N identical oscillators. The time-evolution of this coupled dynamical system is described by:

$$\dot{x}_i = \mathrm{F}(x_i) + \epsilon \sum_{j=1}^{n} L_{ij}\mathrm{H}(x_j). \tag{4.5}$$

Here, x_i is a variable associated with node i; F and H are evolution and output functions, respectively; ϵ is the strength of coupling; and $\mathbf{L}$ is the Laplacian matrix, defined as: $L_{ii} = k_i$, the degree of node i, $L_{ij} = -1$ if nodes i and j are connected, 0 otherwise. It has been shown that the linear stability of the synchronized state x_s ($=x_1 = \ldots = x_N$) can be determined by diagonalizing the variational equation (Eq. 4.5) into N blocks of the form, $\dot{y_i} = [DF(s) + \epsilon\lambda_i DH(s)]y_i$, where y_i represent different modes of perturbation from the synchronized state. This is also referred to as the *master stability equation* [105]. These equations have the same form but different effective couplings $\alpha_i = \epsilon\lambda_i$. The synchronized state is stable, i.e., the maximum Lyapunov exponent is in general negative, only within a bounded interval $[\alpha_A, \alpha_B]$ [111]. Let the eigenvalues of the Laplacian matrix be arranged

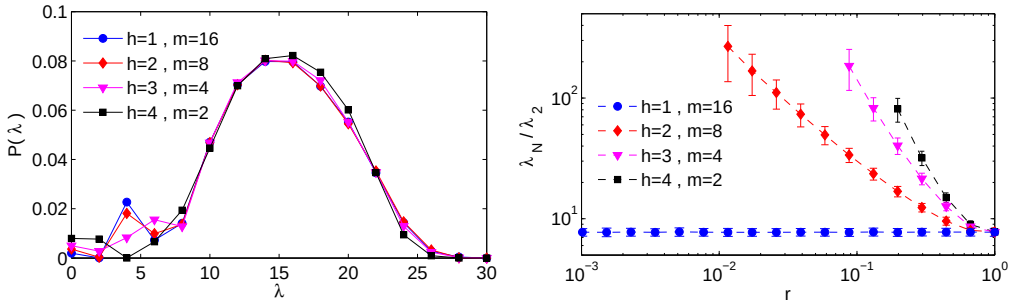


Figure 4.6: (Left) Probability distribution of eigenvalues of the Laplacian L, as a function of the total number of hierarchical levels, h ($r = 0.1$). (Right) The ratio of the largest eigenvalue (λ_N) to the second smallest eigenvalue (λ_2) as a function of r, the ratio of inter-modular connections between two successive hierarchical levels, with different symbols corresponding to differing total number of hierarchical levels h. For all cases, the network consists of $N = 256$ nodes with average intra-modular degree, $\langle k_{\text{intra}} \rangle = 10$ and inter-modular degree, $\langle k_{\text{inter}} \rangle = 5$. At all hierarchical levels $l > 1$, the network is split into two sub-networks. At $l = 1$, each subnetwork is split into m modules ($l = 0$). Thus, $N = 256$ nodes are divided equally among $2^{h-1}m = 16$ modules, with the four curves corresponding to ($\square$) $h = 4$, $m = 2$, ($\bigtriangledown$) $h = 3$, $m = 4$, ($\diamond$) $h = 2$, $m = 8$, and ($\circ$) $h = 1$, $m = 16$. Note that, increasing the number of hierarchical levels leads to divergence of the eigenratio, implying that synchronization becomes harder to achieve.

as $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$. Then, requiring all effective couplings to lie within the interval $\alpha_A < \epsilon\lambda_2 \leq \cdots \leq \epsilon\lambda_N < \alpha_B$, implies that a synchronized state is linearly stable, if and only if, $\lambda_N/\lambda_2 < \alpha_B/\alpha_A$. Thus, a network having a smaller eigenratio λ_N/λ_2, is more likely to show stable synchronized activity.

Here, we obtain the eigenvalues of the Laplacian $\mathbf{L}$ for a hierarchical modular network (Fig. 4.6, left) and observe the eigenratio λ_N/λ_2 as a function of ratio of the inter-modular connections between two successive hierarchical levels, r, and the total number of hierarchical levels, h. First, keeping the number of hierarchical levels fixed, we vary the parameter r. We find that with decreasing r, i.e., as the number of connections between two successive hierarchical levels decrease, the instability of the synchronized state increases. Next, keeping the total number of modules fixed we increase the number of hierarchical levels (h) in the network. Fig. 4.6 (right) shows that as the number of hierarchal levels of the network is increased, λ_2 decreases, resulting in an increasing eigenratio. Thus, arranging the modules of a network in a hierarchical fashion also makes a network difficult to

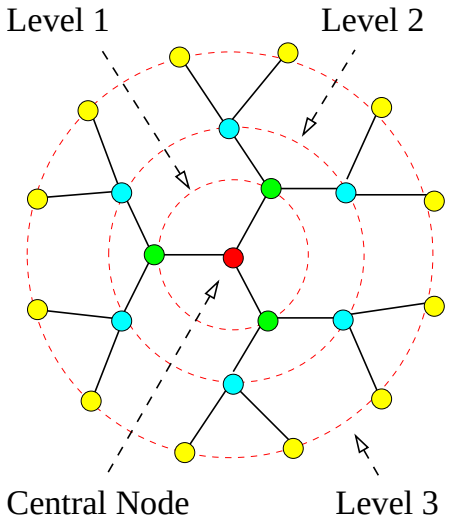


Figure 4.7: Schematic diagram of an optimal tree structure with minimum average path length, where the highest degree ($k_{max} = 3$) of the network is fixed. The central node and the three different levels ($l = 1, 2, 3$) indicate a hierarchical organization.

synchronize.

4.4 Evolution of hierarchy in complex networks

In the preceding chapter we have shown using a model, how modular network can emerge in the real world through multi-constraint optimization. In this section, we generalize this model to understand how optimizing for certain coexisting constraints on system performance can lead to hierarchical configuration of the network.

The network constraints consider here are similar to the ones in the previous chapter:

1. Minimizing the average path length: This is essential for rapid transportation of resources or propagation of information within a network. Further, in the presence of noisy information channels, signals are likely to be degraded during transit between nodes. In such a system, reducing the average path length of the network will also increase the robustness of the system.

2. Minimizing the total number of links in the system: Generally, each link in

a system has some associated resource cost. Thus, a network having a large number of links will incur a high cost in terms of overheads for maintaining the connections.

3. Minimizing the largest degree (k_{max}) of the network: This is associated with reducing congestion in any node, a criterion that is important in transportation and information networks. E.g., in information transfer networks like the Internet, unless the maximum degree is limited within a reasonable bound, the hub nodes are likely to get clogged with incoming packets. In social networks, this is related to the limited attention that each individual can give towards maintaining each additional social tie owing to time constraints. In addition, for general dynamical systems, decrease in k_{max} is associated with increasing linear stability for the network dynamics.

As simultaneously optimizing a network for all three of the above constraints is a difficult problem, we first consider a network having N nodes and $N-1$ links (the minimum number required to maintain connectivity) that automatically satisfies the constraint of minimum link cost. We simplify the problem further by considering the largest degree of the network, k_{max}, to be fixed, and seek to obtain the tree structure which has minimum average path length. Such a network can be constructed as follows: (i) choose a node to be the central node for the network and attach k_{max} nodes to it. Thus, there will be k_{max} nodes in the first layer which are located unit distance from the central node, (ii) add $k_{max}-1$ nodes to each of the nodes in the first layer, (iii) continue this procedure until the prescribed number of nodes in the network is achieved. In such a network all the nodes except those in the outer layer have degree k_{max}.

To prove that the above algorithm generates a network which has the shortest average path length for a given value of k_{max} we use the method of induction. For a network with $k_{max}+1$ nodes, the algorithm generates a star configuration where the central node is connected to the other k_{max} nodes. This has, by construction, the least average path length ($\simeq 2$) among all possible network configurations with the same number of nodes and links. Thus, for the set of k_{max} nodes belonging to the first layer ($l=1$) of the model network, the algorithm guarantees the shortest average path length. Let us now consider nodes belonging to layers beyond the first one. If two nodes, i and j, are at same level n, then the distance between

these two nodes is given by

$$d(i,j) = 2 + d(p(i), p(j)), \tag{4.6}$$

where $p(i)$ and $p(j)$ are the parents of nodes i and j respectively, i.e., the nodes in level $n-1$ to which i and j are connected. It is obvious that if $d(p(i), p(j))$ is minimum, so is $d(i,j)$, as the increment of 2 is the least possible length [$= d(i,p(i)) + d(j,p(j))$] that one needs to add to the distance between the parents of i and j to obtain $d(i,j)$.

Let us next consider the case when the nodes are not at same level, e.g., let node i be at level m and node j be at level $n < m$. Thus, the shortest distance between the jth node and a node in level m is $m-n$. Therefore, the distance between nodes i and j is given by

$$d(i,j) = (m-n) + d(p^{m-n}(i), j). \tag{4.7}$$

where $(p^{m-n}(i)$ is a $(m-n)$-th grandparent of node i, which occurs in the level $m-(m-n) = n$. As the nodes $p^{m-n}(i)$ and j occur at the same n-th level, for which $d(p^{m-n}(i), j)$ has been shown to be minimal by the argument in the previous paragraph, $d(i,j)$ is minimum even when they belong to different levels. Thus, the network constructed by the algorithm proposed above will have the minimum average path length for a fixed maximum degree among all possible network configuration with same number of nodes and links (Fig. 4.7).

We now calculate how the number of levels in the hierarchical tree is related to the maximum degree k_{max}. The total number of nodes in a model network with r levels, constructed according to the above algorithm, is given by,

$$\begin{aligned} f(k_{max}, r) &= 1 + k_{max} + k_{max}(k_{max}-1) + \cdots + k_{max}(k_{max}-1)^{r-1} \\ &= \frac{k_{max}(k_{max}-1)^r - 2}{k_{max} - 2}. \end{aligned} \tag{4.8}$$

Thus, the total number of nodes in the network, N, will be bounded by $f(k_{max}, r) \leq N \leq f(k_{max}, r-1)$. For large r and k_{max}, we can replace the above relation with an equality, giving $r = \frac{\log N}{\log(k_{max}-1)}$. This expression for the number of hierarchical levels, r, expressed in terms of network size and maximum degree can be considered

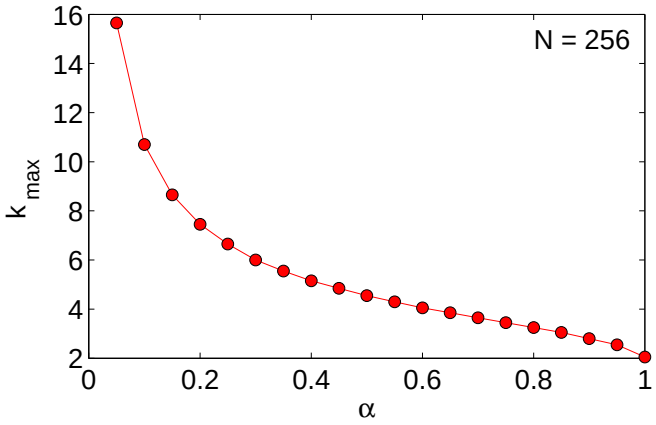


Figure 4.8: Variation of the largest degree, k_{max}, for the optimal network with the parameter α for a network with $N = 256$.

as a measure of the average path length for the network. Thus, the constraint of minimizing the path length is seen to be equivalent to a constraint dictating the minimization of the number of levels.

So far, we had been considering the simple problem of a network with fixed largest degree k_{max}. To construct networks that simultaneously satisfy the constraints on minimum number of levels and least maximum degree, we define a cost (or energy) function in terms of r and k_{max}:

$$E = \alpha k_{max} + (1 - \alpha)\frac{\log N}{\log(k_{max} - 1)}. \tag{4.9}$$

Here, the parameter $\alpha \in [0, 1]$ determines the relative importance of the constraint on largest degree with respect to that on the number of hierarchical levels. For any specified value of α, the optimal network is the one which has minimum E. Note that, for a network with fixed N, as r is expressed in terms of k_{max}, the cost is exclusively a function of the largest degree. Therefore, to obtain the minima of E, we can differentiate it with respect to k_{max} and by equating the result to 0, we obtain the following implicit relation for the optimal value of k_{max}:

$$\frac{(1 - \alpha)}{\alpha} \log N = (k_{max} - 1) \left[\log(k_{max} - 1)\right]^2. \tag{4.10}$$

Thus, for a given N and α, using the above relation we can obtain the k_{max} for a

network which is the optimal configuration that simultaneously satisfies the constrains specified above. Fig. 4.8 indicates that as $\alpha \to 0$, the optimal network has a star configuration, while for $\alpha = 1$, it is a 1-dimensional chain. At intermediate values of α, the network will resemble the hierarchical configuration schematically shown in Fig. 4.7.

4.5 Discussion

In Chapter 2, we have shown that modularity in random networks leads to higher probability of instability for the equilibria of the network dynamics. The work presented here is an extension and generalization of the above result, demonstrating that increased number of hierarchical levels also tend to destabilize these equilibria. Moreover, the same phenomena is observed for the stability of synchronized activity in a network with respect to both increasing modularity, as well as, hierarchy. This raises the question of how can systems with hierarchical modular structures exist in nature, where they have to be robust enough to survive constant environmental fluctuations. To answer this, we note that many constraints operate on networks occurring in real life, such as, the minimization of (a) resource cost for maintaining links and (b) the time required for communicating between nodes, in addition to (c) the linear instability of equilibria, which together can make modular networks the optimal configuration [85].

However, while this can explain the ubiquity of modularity, it does not answer the question of why hierarchical organization is so common in nature. The fact that tree-like networks with extensive ramifications occur so often in the context of resource transport (e.g., the circulatory system in plants and animals), suggests that additional constraints related to the functional goal of maintaining steady flow at high flux may be at work in this case. One such constraint is the need to reduce congestion at any point in the system, which is equivalent to minimizing the largest degree in the network. We have shown that introducing this constraint, can lead to networks with hierarchical organization, when operating in conjunction with the previously introduced constraints on resource cost for links, and communication time between nodes. Another possible candidate for a constraint that may give rise to non-trivial mesoscopic organization is the need to minimize wiring cost, the

total geographic length for all links in the network [92]. This is applicable when the network is embedded on physical (as opposed to topological) space, so that the wiring cost can been defined as the sum of the Euclidean distances between all connected pairs of nodes. As many of the networks that show hierarchical organization (such as the Internet and the network of cortical areas in the brain) indeed occur on a physical space, with the geographic link cost being related to the metric distance between nodes, this is a possibility that is worth pursuing.

5

Inferring network structure from dynamics

Having analyzed several network models for understanding the dynamical consequences of modularity and hierarchy, in this chapter we consider how their existence in real-world complex systems can be inferred from the knowledge of dynamics of the component nodes. We first demonstrate the possibility of reconstructing networks through an analysis of the temporal information on interactions between its components, by using the behavioral data of individuals belonging to a troop of macaque monkeys. This example shows that the knowledge of dynamics of a system can reveal its underlying modular network structure, which has been verified by later field observations. To reconstruct a much larger network from the time-series data of its components, we consider the example of financial markets. These complex systems have many interacting elements and exhibit large fluctuations in their associated observable properties, such as stock price or market index.

In Sec. 5.1 we analyze the structure of a bonnet macaque social organization. Using data on their grooming and approach behavior, we determine the network of interactions between the individuals comprising a troop. We first show that grooming frequency, grooming time and approach frequency between each pair, all have exponential distributions. We were able to determine the distinct groups in the social network of female macaques, whereas that for the males do not show any such unambiguous structure. Next, we consider the Seyfarth model, a theoretical model for reproducing the patterns of social behavior observed in a primate troop. We verify the efficacy of the model in explaining the observed group behavior and

look at possible causative factors behind female grooming interaction.

In Sec. 5.2, by analyzing the cross-correlation matrix of stock price fluctuations through spectral techniques, we reveal the underlying network of interactions between stocks in different markets. We find that emerging markets (e.g., NSE of India) exhibit stronger correlations compared to developed markets (e.g., NYSE of USA). In Sec. 5.3, we show through a simple multi-factor model, that most of the observed correlations among stocks in emerging markets are due to effects common to the entire market (e.g., external signals such are news breaks or intrinsic global signals such as market indices). Coversely, correlations arising through direct interactions between related stocks (e.g., between those belonging to the same business sector) are weak. Our results suggest that the emergence of an internal structure, comprising multiple groups of strongly coupled components, is a signature of market evolution. This work also has ramifications for other similar complex systems that develop over time, as our analysis provides tools for distinguishing dynamical correlations that arise as a result of mutual interactions between nodes, as opposed to those arising through a common response to a global signal.

5.1 Determining the social organization of Bonnet Macaques

5.1.1 Social structure in primates

Primates are among the most social of all mammalian species, bonding together for the purpose of survival. Such bonding between pairs of individuals is extremely important in terms of increasing the cohesiveness of the group. This has immediate relevance in making the group more effective in gathering food, protecting themselves from predators, and, other functions that are vital for survival. Thus, the pattern and quality of social interactions among the individuals have a direct impact on the functional properties of the system. Primates do not interact at random but rather has certain characteristic patterns of social behavior. These may be invariant with respect to group size, age, composition and habitat quality. Indeed, such patterns define the structure of social organization in the species.

A commonly observed behavior that is often used to infer such patterns is that

of grooming between individuals. As opposed to conflict or aggressive behavior, grooming is indicative of cooperative behavior. In addition to its benefit in removing ecto-parasites, grooming results in the formation and fostering of affiliative relationships between individuals that could help them in other spheres of activity, such as building a coalition against a third (possibly more powerful) individual.

Another common social interaction among primates that is indicative of the relative status of two individuals is their behavior when one animal approaches the other. It is seen that the aggressive approach of one individual is reciprocated by the retreat of the other individual, a pattern that is almost invariant over time for the pair involved. As the direction of such approach-retreat interactions is relatively stable over time, one can therefore define relative dominance between the two individuals. Moreover, these dominance relations are also transitive, i.e., if A is dominant to B and B is dominant to C, A is invariably dominant to C. Thus, the members of a primate troop can be arranged in a linear dominance hierarchy, with a rank associated with each individual. Usually, field studies concentrate on the social interactions between members of the same sex in a troop, as male-female interactions involve additional factors. Thus, the data for social interactions in a group of primates is collected for the two subsets: one of all female members and the other of all male members.

To understand grooming behavior, simple mathematical models of interactions between individuals in a group have been proposed. These interactions define the social network of the primate troop. R. M. Seyfarth has introduced one such theoretical model to understand certain features that are commonly observed in the grooming behavior of adult female primates across several species, viz., (i) higher ranked individuals receive more grooming than others, and, (ii) majority of grooming occurs between individuals of adjacent rank. Using this model, where every individual follows the same strategy in choosing grooming partners, Seyfarth has shown that relatively complex features of social behavior can be explained in terms of simple principles governing the actions of individuals.

5.1.2 Bonnet Macaques

In the work reported here, we shall be focusing on one particular primate species, the bonnet macaque (*Macaca radiata*), which is the most commonly observed pri-

mate in peninsular India. Members of this species usually live in large troops containing multiple males and females of 8 to 60 individuals, where the adult individuals develop strong affiliative relationships with each other [112]. Female bonnet macaques usually remain with the group in which they are born throughout their lives. As adults they form stable matrilineal dominance hierarchies, with daughters having dominance ranks just below those of their mothers. Among sisters, dominance is ordered in a reverse chronological order, with the youngest being the most dominant. The close affiliative relations between females is demonstrated through high levels of allogrooming [1] exchanged between both genetically related and unrelated individuals across the dominance hierarchy.

In contrast to females, adult (as well as juvenile) bonnet macaque males usually emigrate from the troops in which they were born. Another marked difference from the females is that adult males form unstable dominance hierarchies. By direct aggression and formation of coalitions, macaque males may move up from low ranks occupied while very young to relatively high positions when they are mature and in peak physical condition. Conversely, older macaque males may slip down the hierarchy to lower ranks. Although, just as their female counterparts, macaque males also demonstrate high levels of allogrooming and other affiliative behavior towards each other, in marked contrast to females, there is absence of any correlation between individual dominance ranks and the levels of affiliative behavior displayed or received.

5.1.3 Description of the dataset

The analysis presented here is based on data acquired in the field by the group of Prof. Anindya Sinha (NIAS, Bangalore) through demographic monitoring and behavioral observations on a troop of bonnet macaques inhabiting 1 square km of dry deciduous scrubland and mixed forests in the GKVK campus of the University of Agricultural Sciences in Bangalore, India. The original observations were carried out for over 1200 hours on two troops occupying adjacent overlapping home ranges during March 1993-September 1995. We have selected the larger of these two groups for our analysis, which consisted of 12 adult males and 11 adult females. Data collected include information about (i) allogrooming frequency GF (measured

[1] Grooming performed by one individual on another is called allogrooming

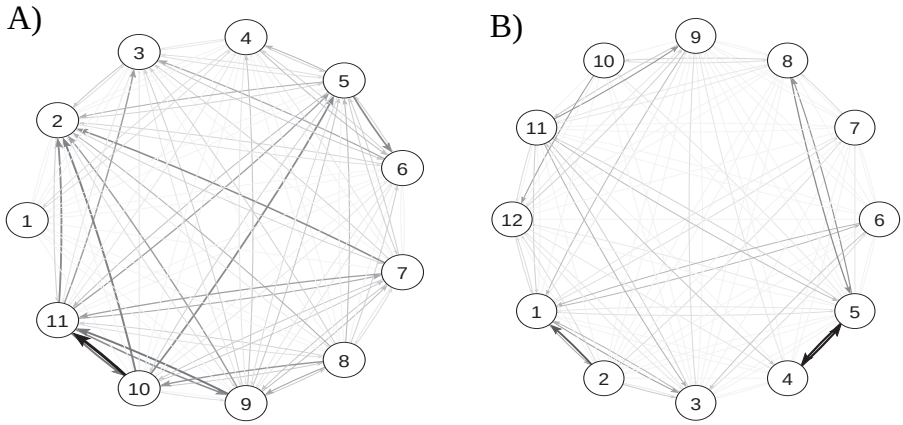


Figure 5.1: The social network of (a) female and (b) male members of the bonnet macaque troop, where the interaction strength between each pair of individuals is determined by their corresponding grooming frequency (GF).

for every pair of giving and receiving individuals as grooming bouts/hour), (ii) allogrooming time GT (measured for the pair of giving and receiving individuals in terms of hours) and (iii) approach frequency AF (measured as the number of approaches initiated by one individual towards another per hour).

5.1.4 Distribution of interaction strengths

We have constructed the network of social interactions in the macaque troop by using the data described above, where nodes represent the individual members, and links represent the relation between them in terms of GF, GT or AF (Fig. 5.1). Note that, the network is directed as, in general, the relation between a pair of individuals is not symmetric. For example, the time spent by individual i in grooming individual j may not equal the time j spends in grooming i. Each link of the network has an associated weight, w_{ij}, which is proportional to GF, GT or AF, depending on which relation is being used to construct the social network.

We first consider the distribution of weights in the links of the network. For many complex networks occurring in the real world, this distribution is seen to have either a power-law (e.g., in air-transportation network), or a log-normal (e.g., in the international trade network) nature. This is indicative of significant levels

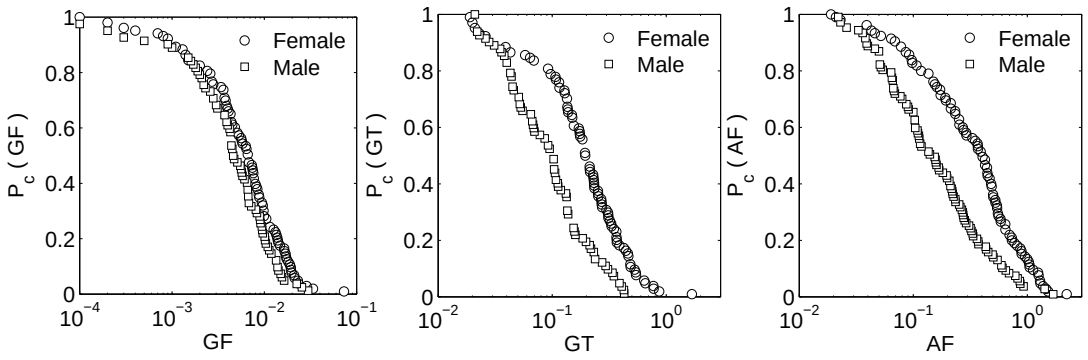


Figure 5.2: The distribution of interaction strengths, defined in terms of (a) grooming frequency (GF), (b) grooming time (GT), and (c) approach frequency (AF), for both the female and male macaque social networks. The data indicates an exponential nature for all three distribution.

of heterogeneity in the interaction strengths of the network. For social networks, quantifying the strength of interactions is often not possible, which makes the determination of the link weight distribution for such systems a very difficult problem. Fortunately, in the case of macaque social network, the interaction strengths can be quantified in terms of their grooming and approach behavior. Fig 5.2 indicates that for both the male and female networks, the strength distribution seems to decay exponentially. The implication of relative homogeneity in the link weights suggests that this social network is very different in this respect from the networks mentioned above. Moreover, the relatively fast decay in the distribution of link weights for the males indicates a weaker degree of social interactions among them as compared to the females.

5.1.5 Community structure in macaque social network

We next test the utility of a network description in allowing us to infer the existence of subsets whose members are strongly bound to each other. This is possible through the determination of community structure in the network and verifying it with empirically observed behavior. We have attempted partitioning the network into several closely knit communities (or modules) using GF, GT and AF. The community in a weighted network is defined as a group of nodes that are strongly interconnected, i.e., have links among themselves with higher weights, as compared

Table 5.1: Modular decomposition of the male and female bonnet macaque social networks, indicating the membership of individuals in different modules. Each individual is indicated by a number that corresponds to its rank in the linear dominance hierarchy, with '1' corresponding to the most dominant. The number of communities obtained is indicated by m, whereas the maximum modularity of the empirical network and the corresponding randomized network is represented by Q and Q_{rand} respectively.

Gender	Type	Q	m	Q_{rand}	Modular identity
Female	GF	0.121	2	0.081±0.017	(1 2 3 4 5 10) (6 7 8 9 11)
	GT	0.140	2	0.098±0.021	(1 2 3 4 5 10) (6 7 8 9 11)
	AF	0.110	2	0.073±0.020	(1 2 3 4 5 10) (6 7 8 9 11)
Male	GF	0.085	2	0.130±0.025	(1 2 3 4 9 12) (5 6 7 8 10 11)
	GT	0.165	4	0.137±0.024	(1 2 3) (4 5) (6 8 10) (7 9 11 12)
	AF	0.240	4	0.143±0.025	(1 2 3 6 7) (4 5 8) (9 11) (10 12)

to links with other nodes which belong outside the community. We determine the modules by obtaining the optimal partitioning of the network, that corresponds to the partition having the maximum value of the modularity measure Q [98]. Table 5.1 indicates the different community structures obtained for both the male and female macaque network by applying this method.

For the female macaques, the communities determined from the three different social networks (defined in terms of GF, GT and AF) are identical, indicating that this modular structure is significant and might be observed in other social behavior. This is verified by subsequent field observations carried out at a later period (end of 1995) when it was seen that the female macaques had split into two distinct troops, with the membership of each exactly matching the results of our network analysis. However, for male macaques, the modular decomposition yields different communities according to the type of social network used. To check the statistical significance of the determined community structure, we compare the partitioning of the empirical network with its randomized version. The randomized networks are obtained by shuffling the weights of the links. The average Q_{rand} for 100 such realizations is compared with the Q of the empirical network. The result again suggests that, while the modular decomposition of the female network is indeed significant, this is not so for the male network.

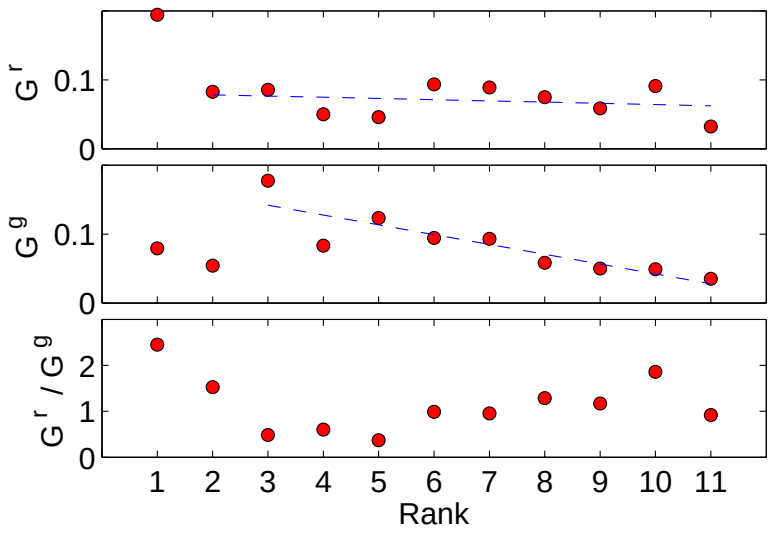


Figure 5.3: Relation between the rank of a macaque with (a) grooming received, G^r, (b) grooming given, G^g, and (c) the ratio G^r/G^g.

5.1.6 From individual strategy to social behavior

In this section we focus exclusively on the female members of the troop, as their dominance hierarchy is stable unlike that for the males. Here, we investigate how the observed interaction structure can arise from a common set of strategies or principles governing the selection of interaction partner, that are followed by each individual member of the troop. Using our data, we test the theoretical model proposed by Seyfarth for describing social behavior in monkeys.

Seyfarth model

In this model, a number of adult females are arranged in a linear rank order that defines priority of access to resources. Some of the individuals are related to each other in terms of common ancestry. Every individual follows the same behavioral strategy, viz., distribute half of their grooming among close relatives (kin), and, the other half among unrelated animals in direct proportion to their ranks (with the most dominant animal getting top preference). Using this strategy, every individual is allowed to interact with others, each pursuing her goal within the constraints imposed by competition. After a period of interactions, the pattern

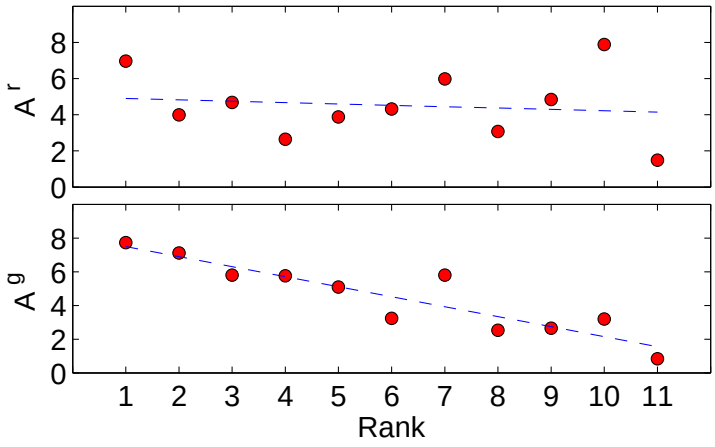


Figure 5.4: Relation between the rank of a macaque with (a) the frequency of approach received per hour, A^r and (b) the frequency of approach initiated per hour, A^g.

of grooming behavior is obtained which is then compared with data from actual groups. Simulations of the model show that the network structure constructed on the basis of the interactions are relatively unaffected by variation in the number of kin, group size, the amount of time available for grooming, or the relative strength of an individual's attraction to kin and her attraction to those of higher status.

Testing the Seyfarth model

We have tried to verify the basic principles of the Seyfarth model by testing it using the bonnet macaque dataset. We first calculate the grooming given and received by each macaque. For an individual i, the grooming received is the weighted in-degree, $G_i^r = \sum_j w_{ij}$, while the grooming given is the weighted out-degree $G_i^g = \sum_i w_{ij}$. We find that, apart from the highest ranked macaque, grooming received by all other individuals is approximately the same (Fig. 5.3). This is in significant contrast to the prediction of the Seyfarth model, according to which the grooming received should increase with rank. Next, we look at the variation of grooming given by the macaque with its rank. Except for the two highest ranked macaques, G^g decreases with rank, indicating the competition among individuals. By plotting the ratio of grooming received to the grooming given by a macaque, G^r/G^g, along with its rank, we show that: (i) for the upper part of the dominance hierarchy there is a positive correlation, but (ii) for lower ranking females, the correlation

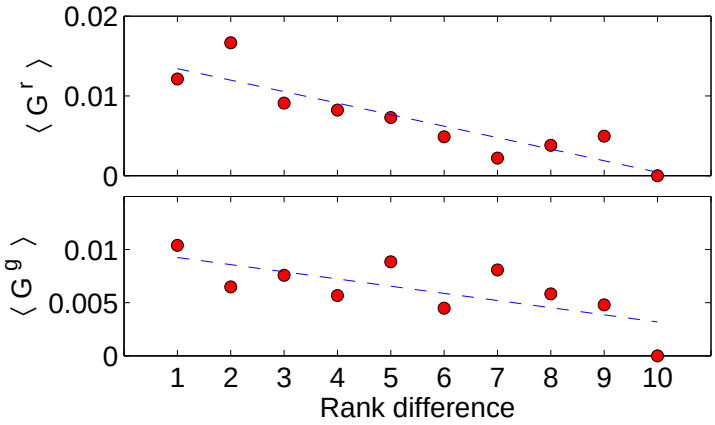


Figure 5.5: Relation between rank difference of macaque with (a) average grooming received, $\langle G^r \rangle$, from low rank individual and (b) average grooming given, $\langle G^g \rangle$, to the low rank individual. The curves are the best fit for the data and indicate negative correlation.

between rank and G^r/G^g is not obvious.

Next, we look at a different type of social interaction, namely the frequency of approach behavior between individuals. In Fig. 5.4, we plot the number of times an individual is approached per hour, A^r, against its rank and find them to be uncorrelated. However, when we plot the frequency of approaches initiated per hour, A^g, against the rank of the individual, we find significant positive correlation. This indicates that the higher ranking females approach most other members of the group, but the reverse does not happen.

Next, we test the reciprocity assumption of Seyfarth model in the grooming behavior of macaque, according to which each individual regardless of its own rank always tries to groom the highest ranked individual who is available. Here, we test it against an alternative hypothesis, that an individual may choose a grooming partner who is ranked higher according to the preceding strategy, but, chooses randomly when the available partners are all ranked lower than it. In Fig. 5.5, we plot the amount of grooming received by a higher ranked individual from a lower ranked partner, against their rank difference. We compare this to the grooming given by a higher ranked individual to a lower ranked partner. In both of these cases, we find that as the rank difference increases, the amount of grooming decreases. However, grooming received by a higher ranked individual from one at a

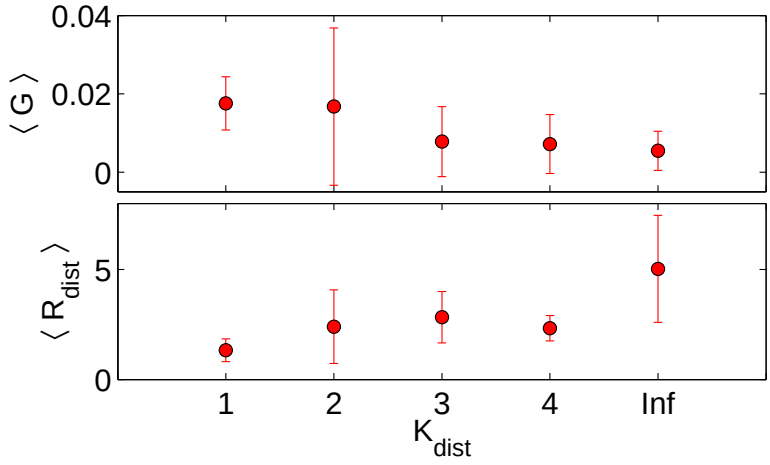


Figure 5.6: Relation between the kinship distance, K_{dist}, between macaque individuals with (a) average grooming, $\langle G \rangle$, and (b) average rank difference, R_{dist}.

lower rank shows a steeper decay compared to the corresponding data for grooming given. This indicates that during the quest for grooming partners, a macaque individual may be using the following guiding principles: (i) rank is more important while grooming a higher ranked individual, but (ii) it is less significant when grooming a lower ranked macaque. This indicates the non-reciprocal nature of social ties in this system.

Finally, we look at the role of kinship distance between individuals on the bonnet macaque grooming behavior. To define the kinship distance, we first construct the family tree, which is a chart representing relationships between individuals in terms of common ancestry. The path length between two individuals on this tree is defined as the kinship distance (K_{dist}) between them. Thus, the distance between mother and daughter is 1, whereas the distance between sisters is 2 and so on. We plot the average grooming between pairs of individuals against their kinship distance in Fig. 5.6. Our results suggest that, grooming decreases as the kinship distance increases. Thus distance between kin does matter in macaques when choosing a grooming partner. However, the increase in kinship distance between two individuals is also associated with a corresponding increase in rank difference, R_{dist}. Thus, whether the increase in average grooming is exclusively an outcome of the increase in K_{dist} needs to be investigated further.

5.2 Reconstructing the internal structure of a financial market

In order to reconstruct a much larger network from the time-series data of its components, we now consider the example of financial markets. By analyzing the cross-correlation matrix $\mathbf{C}$ of stock price fluctuations in the National Stock Exchange (NSE) of India, we uncover the structure of interactions between the stocks that are traded in that market.

5.2.1 Financial market: A complex system

Financial markets can be considered as complex systems having many interacting elements and exhibiting large fluctuations in their associated observable properties, such as stock price or market index [113, 114]. The state of the market is governed by interactions among its components, which can be either traders or stocks. In addition, market activity is also influenced significantly by the arrival of external information. Statistical properties of stock price fluctuations and correlations between price movements of different stocks have been analyzed by physicists in order to understand and model financial market dynamics [115, 116]. The fluctuation distribution of stock prices is found to follow a power law with exponent $\alpha \sim 3$, the so-called "inverse cubic law" [117, 118]. This property is quite robust, and has been seen in developed as well as emerging markets [85]. On the other hand, it is not yet known whether the cross-correlation behavior between stock price fluctuations has a similar universal nature. Although the existence of collective modes have been inferred from the study of market dynamics, such studies have exclusively focused on developed markets, in particular, the New York Stock Exchange (NYSE).

To uncover the structure of interactions among the elements in a financial market, physicists primarily focus on the spectral properties of the correlation matrix of stock price movements. Pioneering studies investigated whether the properties of the empirical correlation matrix differed from those of a random matrix that would have been obtained had the price movements been uncorrelated [119, 120]. Such deviations from the predictions of random matrix theory (RMT) can provide clues about the underlying interactions between various stocks. It was observed that,

while the bulk of the eigenvalue distribution for the correlation matrix of NYSE and Tokyo Stock Exchange follow the spectrum predicted by RMT [119, 120, 121, 122], the few largest eigenvalues deviate significantly from this. The largest eigenvalue has been identified as representing the influence of the entire market, common for all stocks, whereas, the remaining large eigenvalues are associated with the different business sectors, as indicated by the composition of their corresponding eigenvectors [123, 121]. The interaction structure of stocks in NYSE have been reconstructed using filtering techniques implementing matrix decomposition [124] or maximum likelihood clustering [125].

While it is generally believed that stock prices in emerging markets tend to be relatively more correlated than the developed ones [126], there have been very few studies of the former in terms of analyzing the spectral properties of correlation matrices [127, 110, 128, 129]. Here, we analyze the cross-correlations among stocks in the Indian financial market, one of the largest emerging markets in the world.

5.2.2 The financial market data

The National Stock Exchange (NSE) is the largest stock market in India. We have considered the daily closing price data of 201 stocks (see Table I) traded in NSE from Jan 1996 to May 2006, which corresponds to 2607 days. This data is obtained from the NSE web-site [130]. The selected stocks were traded over the entire period 1996-2006 and had the minimum number of missing data points (i.e., days for which no price data is available). For comparison we also consider the daily closing price of 434 stocks of NYSE belonging to the S&P 500 index over the same period as the Indian data. However, the total number of working days is slightly different, viz., 2622 days. This data was obtained from the Yahoo! Finance website [131]. In all our analysis, while comparing with the NSE data, we have used multiple random samples of 201 stocks each, from the set of 434 NYSE stocks.

5.2.3 The Return Cross-Correlation Matrix

To observe correlation between the price movements of different stocks, we first measure the price fluctuations such that the result is independent of the scale of measurement. If $P_i(t)$ is price of the stock $i = 1, \ldots, N$ at time t, then the

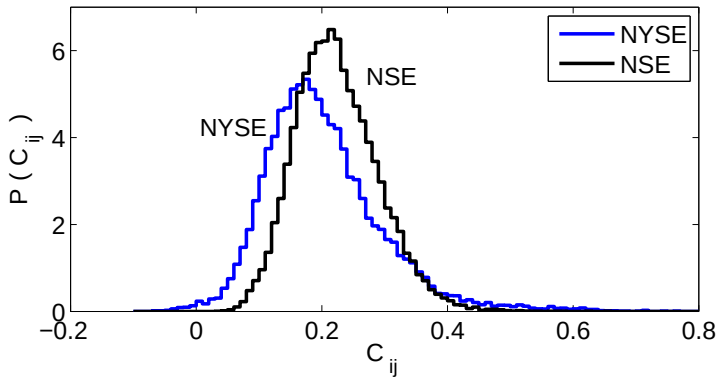


Figure 5.7: The probability density function of the elements of the correlation matrix **C** for 201 stocks in the NSE of India and NYSE for the period Jan 1996-May 2006. The mean value of elements of **C** for NSE and NYSE, $\langle C_{ij} \rangle$, are 0.22 and 0.20 respectively.

(logarithmic) price return of the ith stock over a time interval Δt is defined as

$$R_i(t, \Delta t) \equiv \ln P_i(t + \Delta t) - \ln P_i(t). \tag{5.1}$$

As different stocks have varying levels of volatility (measured by the standard deviation of its returns) we define the normalized return,

$$r_i(t, \Delta t) \equiv \frac{R_i - \langle R_i \rangle}{\sigma_i}, \tag{5.2}$$

where $\sigma_i \equiv \sqrt{\langle R_i^2 \rangle - \langle R_i \rangle^2}$, is the standard deviation of R_i and $\langle \ldots \rangle$ represents time average over the period of observation. We then compute the equal time cross-correlation matrix **C**, whose element

$$C_{ij} \equiv \langle r_i r_j \rangle, \tag{5.3}$$

represents the correlation between returns for stocks i and j. By construction, **C** is symmetric with $C_{ii} = 1$ and C_{ij} has a value in the domain $[-1, 1]$. Fig. 5.7 shows that, on the average correlation among stocks in NSE is larger compared to the stocks in NYSE. This supports the general belief that developing markets tend to be more correlated than developed ones. To understand the reason behind this excess correlation, we perform an eigenvalue analysis of the correlation matrix.

5.2.4 Spectral properties of correlation matrix

If the N return time series of length T are mutually uncorrelated, then the resulting random correlation matrix is called a Wishart matrix, whose statistical properties are well known [132]. In the limit $N \to \infty$, $T \to \infty$, such that $Q \equiv T/N \geq 1$, the eigenvalue distribution of this random correlation matrix is given by

$$P_{\mathrm{rm}}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}, \tag{5.4}$$

for $\lambda_{min} \leq \lambda \leq \lambda_{max}$ and, 0 otherwise. The bounds of the distribution are given by $\lambda_{max,min} = [1 \pm (1/\sqrt{Q})]^2$. We now compare this with the statistical properties of the empirical correlation matrix for the NSE. In the NSE data, there are $N = 201$ stocks each containing $T = 2606$ returns; as a result $Q = 12.97$. Therefore, it follows that, in the absence of any correlation among the stocks, the distribution should be bounded between $\lambda_{min} = 0.52$ and $\lambda_{max} = 1.63$. As observed in developed markets [119, 120, 121, 122], the bulk of the eigenvalue spectrum $P(\lambda)$ for the empirical correlation matrix is in agreement with the properties of a random correlation matrix spectrum $P_{\mathrm{rm}}(\lambda)$, but a few of the largest eigenvalues deviate significantly from the RMT bound (Fig. 5.8).

However, the number of these deviating eigenvalues are relatively few for NSE compared to NYSE. We verify that, these outliers are not an artifact of the finite length of the observation period, by randomly shuffling the return time series for each stock, and then re-calculating the resulting correlation matrix. The eigenvalue distribution for this surrogate matrix matches exactly with the random matrix spectrum $P_{\mathrm{rm}}(\lambda)$, indicating that the outliers are not due to "measurement noise" but are genuine indicators of correlated movement among the stocks. Therefore, by analyzing the deviating eigenvalues, we may be able to obtain an understanding of the structure of interactions between the stocks in the market.

Properties of the "deviating" eigenvalues

The largest eigenvalue λ_0 for the NSE cross-correlation matrix is more than 28 times greater than the maximum predicted by RMT. This is comparable to NYSE, where λ_0 is about 26 times greater than the random matrix upper bound. Upon testing with synthetic US data containing same number of missing data points as

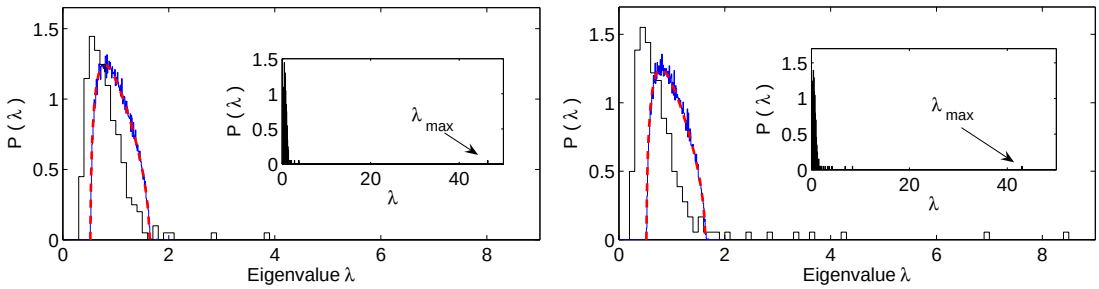


Figure 5.8: The probability density function of the eigenvalues of the correlation matrix **C** for NSE (left) and NYSE (right). For comparison, the theoretical distribution predicted by Eq. (5.4) is shown using broken curves, which overlaps with the distribution obtained from the surrogate correlation matrix generated by randomly shuffling each time series. In both figures, the inset shows the largest eigenvalue.

in the Indian market, we observed that λ_0 remains almost unchanged compared to the value obtained from the original US data. The corresponding eigenvector shows a relatively uniform composition, with all stocks contributing to it and all elements having the same sign (Fig. 5.9, a). As this is indicative of a common factor that affects all the stocks with the same bias, the largest eigenvalue is associated with the *market mode*, i.e., the collective response of the entire market to external information [119, 121].

Of more interest for understanding the market structure are the intermediate eigenvalues, i.e., those occurring between the largest eigenvalue and the bulk of the distribution predicted by RMT. For the NYSE, it was shown that corresponding eigenvectors of these eigenvalues are localized, i.e., only a small number of stocks, belonging to similar or related businesses, contribute significantly to each of these modes [123, 121]. However, for NSE, although the Technology and the IT & Telecom stocks are dominant contributors to the eigenvector corresponding to the third largest eigenvalue, a direct inspection of eigenvector composition does not yield a straightforward interpretation in terms of a related group of stocks corresponding to any particular eigenvalue (Fig. 5.9). This implies that distinct groups, whose members are mutually correlated in their price movement, do exist in NYSE, while their existence is far less clear in NSE.

To obtain a quantitative measure of the number of stocks contributing to a given eigenmode, we calculate the inverse participation ratio (IPR), defined for the kth eigenvector as $I_k \equiv \sum_{i=1}^{N} [u_{ki}]^4$, where u_{ki} are the components of eigenvector k. An

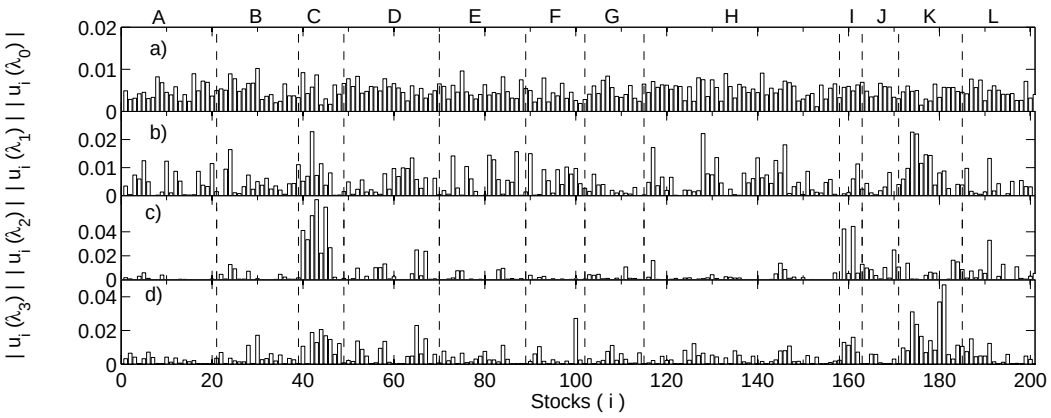


Figure 5.9: The absolute values of the eigenvector components $u_i(\lambda)$ of stock i corresponding to the four largest eigenvalues of $\mathbf{C}$ for NSE. The stocks i are arranged by business sectors separated by broken lines. A: Automobile & transport, B: Financial, C: Technology D: Energy, E: Basic materials, F: Consumer goods, G: Consumer discretionary, H: Industrial, I: IT & Telecom, J: Services, K: Healthcare & Pharmaceutical, L: Miscellaneous.

eigenvector having components with equal value, i.e., $u_{ki} = 1/\sqrt{N}$ for all i, has $I_k = 1/N$. We find this to be approximately true for the eigenvector corresponding to the largest eigenvalue, which represents the market mode. To see how different stocks contribute to the remaining eigenvectors, we note that if a single stock had a dominant contribution in any eigenvector, e.g., $u_{k1} = 1$ and $u_{ki} = 0$ for $i \neq 1$, then $I_k = 1$ for that eigenvector. Thus, IPR gives the reciprocal of the number of eigenvector components (and therefore, stocks) with significant contribution.

On the other hand, the average value of I_k, for eigenvectors of a random correlation matrix obtained by randomly shuffling the time series of each stock, is $\langle I \rangle = 3/N \approx 1.49 \times 10^{-2}$. Fig. 5.10 shows that the eigenvalues belonging to the bulk of the spectrum indeed have this value of IPR. But at the lower and higher end of eigenvalues, both the US and Indian markets show deviations, suggesting the existence of localized modes. However, these deviations are much less significant and fewer in number in the latter compared to the former. This implies that distinct groups, whose members are mutually correlated in their price movement, do exist in NYSE, while their existence is far less clear in NSE.

In order to graphically present the interaction structure of the stocks in NSE, we use a method suggested by Mantegna [113] to transform the correlation between

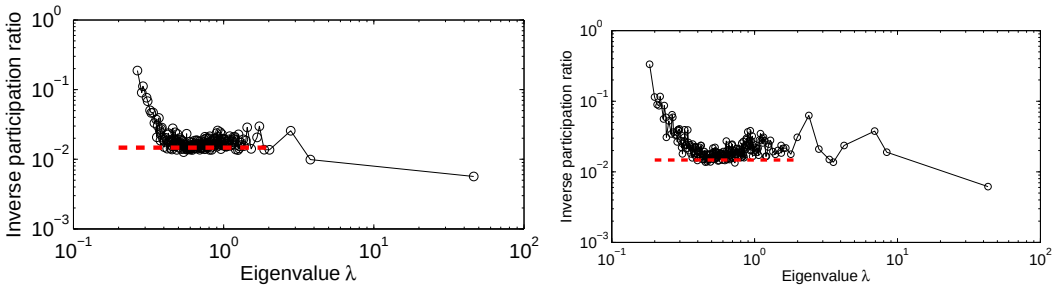


Figure 5.10: Inverse participation ratio as a function of eigenvalue for the correlation matrix **C** of NSE (left) and NYSE (right). The broken line indicates the average value of $\langle I \rangle = 1.49 \times 10^{-2}$ for the eigenvectors of a matrix constructed by randomly shuffling each of the N time series.

stocks into distances to produce a connected network in which co-moving stocks are clustered together. The distance d_{ij} between two stocks i and j are calculated from the cross-correlation matrix **C**, according to $d_{ij} = \sqrt{2(1 - C_{ij})}$. These are used to construct a minimum spanning tree, which connects all the N nodes of a network with $N - 1$ edges such that the total sum of the distance between every pair of nodes, $\sum_{i,j} d_{ij}$, is minimum. For the NYSE, such a construction has been shown to cluster together stocks belonging to the same business sector [32]. However, as seen in Fig. 5.11, for the NSE, such a method fails to clearly segregate any of the business sectors. Instead, stocks belonging to very different sectors are equally likely to be found within each cluster. This suggests that the market mode is dominating over all intra-sector interactions.

5.2.5 Filtering the data using spectral statistics

The above analysis suggests the existence of a market-induced correlation across all stocks, which makes it difficult to observe the correlations that might be due to interactions between stocks belonging to the same sector. Therefore, we now use a filtering method to remove market mode, as well as the random noise [124]. The correlation matrix is first decomposed as

$$\mathbf{C} = \sum_{i=0}^{N-1} \lambda_i \mathbf{u}_i \mathbf{u}_i^T, \tag{5.5}$$

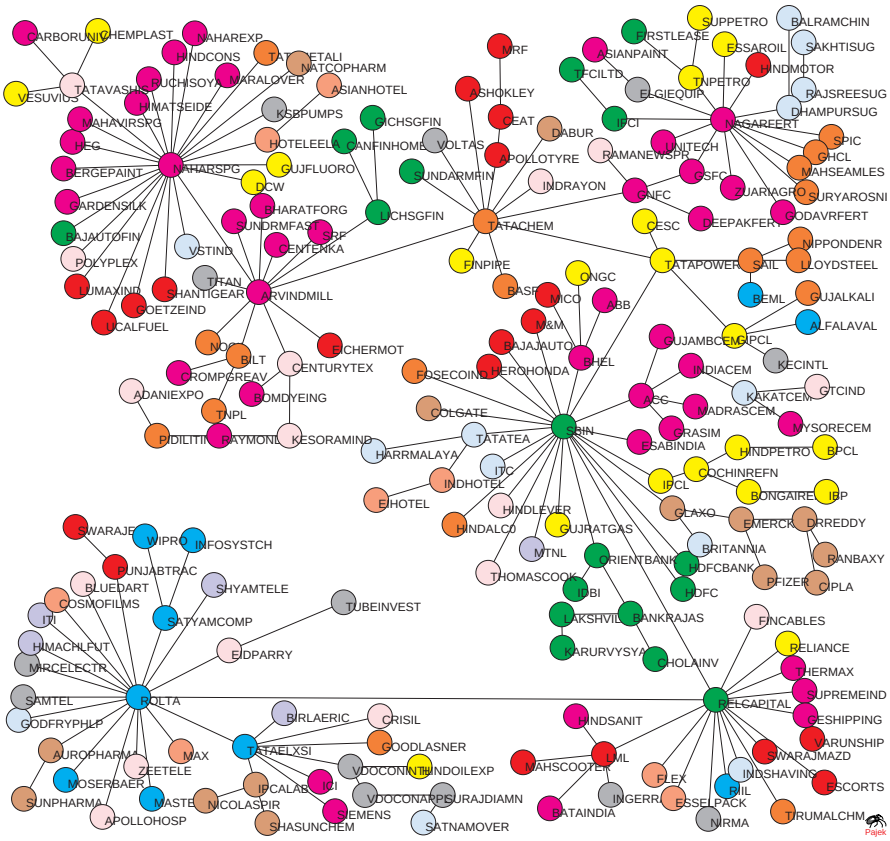


Figure 5.11: The minimum spanning tree connecting 201 stocks of NSE. The node colors indicate the business sector to which a stock belongs.

where λ_i are the eigenvalues of $\mathbf{C}$ sorted in descending order and $\mathbf{u}_i$ are corresponding eigenvectors. As only the eigenvectors corresponding to the few largest eigenvalues are believed to contain information on significantly correlated stock groups, the contribution of the intra-group correlations to the $\mathbf{C}$ matrix can be written as a partial sum of $\lambda_\alpha \mathbf{u}_\alpha \mathbf{u}_\alpha^T$, where α is the index of the corresponding eigenvalue. Thus, the correlation matrix can be decomposed into three parts, corresponding to the *market*, *group* and *random* components:

$$\begin{aligned} \mathbf{C} &= \mathbf{C}^{market} + \mathbf{C}^{group} + \mathbf{C}^{random} \\ &= \lambda_0 \mathbf{u}_0 \mathbf{u}_0^T + \sum_{i=1}^{N_g} \lambda_i \mathbf{u}_i \mathbf{u}_i^T + \sum_{i=N_g+1}^{N-1} \lambda_i \mathbf{u}_i \mathbf{u}_i^T, \end{aligned} \tag{5.6}$$

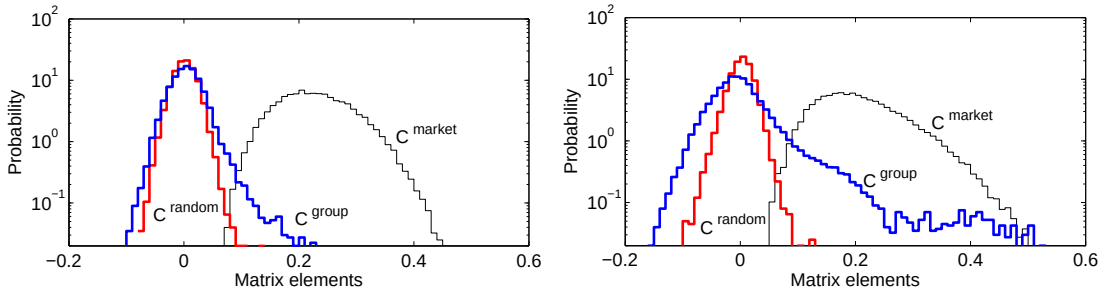


Figure 5.12: The distribution of elements of correlation matrix corresponding to the market, $\mathbf{C}^{market}$, the group, $\mathbf{C}^{group}$, and the random interaction, $\mathbf{C}^{random}$. For NSE (left) $N_g = 5$ whereas for NYSE (right) $N_g = 10$. The short tail for the distribution of the $\mathbf{C}^{group}$ elements in NSE indicates that the correlation generated by mutual interaction among stocks is relatively weak.

where, N_g is the number of eigenvalues (other than the largest one) which deviates from the bulk of the eigenvalue spectrum. For NSE we have chosen $N_g = 5$. However, the exact value of this choice is not crucial as small changes in N_g do not alter the results, the error involved being limited to the eigenvalues closest to the bulk that have the smallest contribution to $\mathbf{C}^{group}$.

Fig. 5.12 shows the result of decomposing the correlation matrix into the three components, for both the Indian and US markets. Compared to the latter, the distribution of matrix elements of $\mathbf{C}^{group}$ in the former shows a significantly truncated tail. This indicates that intra-group correlations are not prominent in NSE, whereas they are comparable with the overall market correlations in NYSE. It follows that the collective behavior in the Indian market is dominated by external information that affects all stocks. Correspondingly, correlations generated by interactions between stocks, as would be the case for stocks in a given business sector, are much weaker, and hence, such correlated sectors would be difficult to observe.

5.2.6 The network of stock interactions

We indeed find this to be true when we use the information in the group correlation matrix to construct the network of interacting stocks [124]. The adjacency matrix $\mathbf{A}$ of this network is generated from the group correlation matrix $\mathbf{C}^{group}$ by using a threshold c_{th} such that $A_{ij} = 1$ if $C_{ij}^{group} > c_{th}$, and $A_{ij} = 0$ otherwise. Thus,

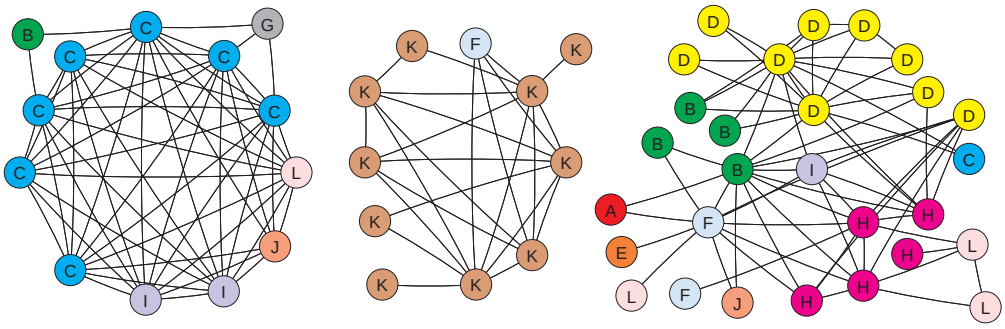


Figure 5.13: The structure of interaction network in the Indian financial market at threshold $c^* = 0.09$. The left cluster comprises of mostly Technology stocks, while the middle cluster is composed almost entirely of Healthcare & Pharmaceutical stocks. By contrast, the cluster on the right is not dominated by any particular sector. The node labels indicate the business sector to which a stock belongs and are as specified in the caption to Fig 5.9.

a pair of stocks are connected if the group correlation coefficient C_{ij}^{group} is larger than a preassigned threshold value, c_{th}. To determine an appropriate choice of $c_{th} = c^*$ we observe the number of isolated clusters (a cluster being defined as a group of connected nodes) in the network for a given c_{th}. We found this number to be much less in NSE compared to that observed in NYSE for any value of c_{th} [124]. Fig. 5.13 shows the resultant network for $c^* = 0.09$, for which the largest number of isolated clusters of stocks are obtained. The network has 52 nodes and 298 links partitioned into 3 isolated clusters. From these clusters, only two business sectors can be properly identified, namely the Technology and the Pharmaceutical sectors. The fact that the majority of the NSE stocks cannot be arranged into well-segregated groups reflecting business sectors illustrates our conclusion that intra-group interaction is much weaker than the market-wide correlation in the Indian market.

When the interaction networks between stocks are generated for the two periods, they show less distinction into clearly defined sectors than was obtained with the data for the entire period. This is possibly because the shorter data sets create larger fluctuations in the correlation values, thereby making it difficult to segregate the existing market sectors. However, we do observe that, using the same threshold value for generating networks in the two periods yield, for the later period, isolated clusters that are distinguishable into distinct sub-clusters connected to each

other via a few links only, whereas in the earlier period the clusters are much more homogeneous. This implies that as the Indian market is evolving, the interactions between stocks are tending to get arranged into clearly identifiable groups. We propose that such structural re-arrangement in the interactions is a hallmark of emerging markets as they evolve into developed ones.

5.3 Model of Market Dynamics

To understand the relation between the interaction structure among stocks and the eigenvalues of the correlation matrix, we perform a multivariate time series analysis using a simple two-factor model of market dynamics. We assume that the normalized return at time t of the ith stock from the kth business sector can be decomposed into (i) a market factor $r_m(t)$, that contains information or signal common to all stocks, (ii) a sector factor $r_g^k(t)$, representing effects exclusive to stocks in the kth sector, and (iii) an idiosyncratic term, $\eta_i(t)$, which corresponds to random variations unique for that stock. Thus,

$$r_i^k(t) = \beta_i r_m(t) + \gamma_i^k r_g^k(t) + \sigma_i \eta_i(t), \tag{5.7}$$

where β_i, γ_i^k and σ_i represent relative strengths of the three terms mentioned above, respectively. For simplicity, these strengths are assumed to be time independent. We choose $r_m(t)$, $r_g^k(t)$ and $\eta_i(t)$ from a zero mean and unit variance Gaussian distribution. We further assume that the normalized returns r_i, also follow Gaussian distribution with zero mean and unit variance. Although the empirically observed return distributions have power law tails, as these distributions are not Levy stable, they will converge to Gaussian if the returns are calculated over sufficiently long intervals. The assumption of unit variance for the returns ensures that the relative strengths of the three terms will follow the relation:

$$\beta_i{}^2 + (\gamma_i^k)^2 + \sigma_i{}^2 = 1. \tag{5.8}$$

As a result, for each stock we can assign σ_i and γ_i independently, and obtain β_i from Eq. (5.8). We choose σ_i and γ_i from a uniform distribution having width δ and centered about the mean values σ and γ, respectively.

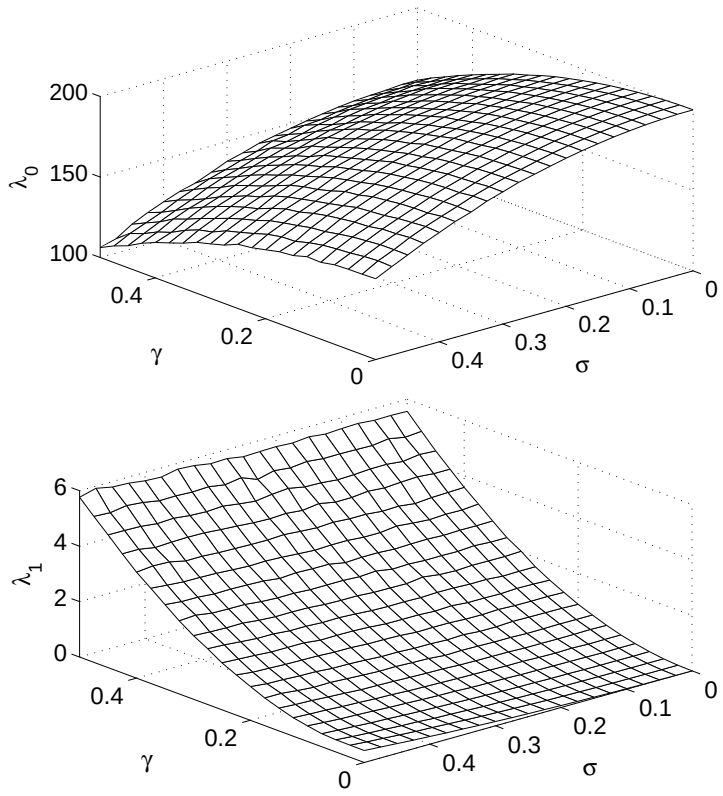


Figure 5.14: The variation of the largest (top) and second largest (bottom) eigenvalues of the correlation matrix of simulated return in the two-factor model (Eq. 5.7) with the model parameters γ and σ (corresponding to strength of the sector and idiosyncratic effects, respectively). The matrix is constructed for $N = 200$ stocks each with return time series of length $T = 2000$ days. We assume there to be 10 sectors, each having 20 stocks.

We now simulate an artificial market with N stocks belonging to K sectors by generating time series of length T for returns r_i^k from the above model. These K sectors are composed of $n_1, n_2, \ldots, n_K$ stocks such that $n_1 + n_2 + \cdots + n_K = N$. The collective behavior is then analysed by constructing the resultant correlation matrix $\mathbf{C}$ and obtaining its eigenvalues. Our aim is to relate the spectral properties of $\mathbf{C}$ with the underlying structure of the market given by the relative strength of the factors. We first consider the simple case, where the contribution due to market factor is neglected, i.e., $\beta_i = 0$ for all i, and the strength of sector factor is equal for all stocks within a sector, i.e., $\gamma_i^k = \gamma^k$, is independent of i. In this case, the spectrum of the correlation matrix is composed of K large eigenvalues, $1 + (n_j - 1)(\gamma^j)^2$, where $j = 1 \ldots K$, and $N - K$ small eigenvalues, $1 - (\gamma^j)^2$, each

with degeneracy $n_j - 1$, where $j = 1 \ldots K$ [133]. Now, we consider nonzero market factor which is equal for all stocks i.e., $\beta_i = \beta$ for all i, and the strength of sector factor is also same for all stocks, i.e., $\gamma_i^k = \gamma$ (independent of i and k). In this case too, there are K large eigenvalues and $N - K$ small eigenvalues. Our numerical simulations suggest that the largest and the second largest eigenvalues are

$$\begin{aligned} \lambda_0 &\sim N\beta^2, \\ \lambda_1 &\sim n_l(1-\beta^2), \end{aligned} \tag{5.9}$$

respectively, where n_l is the size of the largest sector, while the $N - K$ small degenerate eigenvalues are $1-\beta^2-\gamma^2$. We now choose the strength γ_i^k and σ_i from a uniform distribution with mean γ and σ respectively and with width $\delta = 0.05$. Fig. 5.14 shows the variation of the largest and second largest eigenvalues with σ and γ. The strength of the market factor is determined from Eq.5.8.

Note that, decreasing the strength of the sector factor relative to the market factor results in decreasing the second largest eigenvalue λ_1. As $Q = T/N$ is fixed, the RMT bounds for the bulk of the eigenvalue distribution, $[\lambda_{min}, \lambda_{max}]$, remain unchanged. Therefore, a decrease in λ_1 implies that the large intermediate eigenvalues occur closer to the bulk of the spectrum predicted by RMT, as is seen in the case of NSE. The analysis of the model supports our hypothesis that the spectral properties of the correlation matrix for the NSE are consistent with a market in which the effect of information common for all stocks (i.e., the market mode) is dominant, resulting in all stocks exhibiting a significant degree of correlation.

5.4 Discussion

In this chapter, we have reconstructed networks from two types of empirical data. First, we have analyzed the structure of bonnet macaque social organization as an example of a social network. We determine the interaction network by using data on grooming behavior, an affiliative interaction that is frequently observed between primate individuals. We show that weights of the links in this network (i.e., grooming strength) has an exponential distribution, indicating that the interactions in such a social network is very different from other real-world complex networks. This may be due to the limited time for performing specific acts that

constrains the interaction behavior of the agents in a social network. Further, the female macaque network shows a distinct and unambiguous community structure, with the different communities matching exactly with the splitting of the troop observed in a subsequent field study. In order to understand how relatively simple principles governing the strategy of individuals for selecting interaction partners may be used to explain the complex social structure of non-human primate groups, we consider the Seyfarth model. By using the bonnet macaque data, we tested the basic principles of the model, which gives a set of strategic rules that governs the interactions of each individual. Based on our results, we have indicated how these principles can be possibly modified so that the model can better represent the patterns of social behavior in the bonnet macaque.

Next, we analyze the market interaction structure and demonstrate that the stocks in an emerging market are much more correlated than in developed markets. Although, the bulk of the eigenvalue spectrum of the correlation matrix of stocks $\mathbf{C}$ in an emerging market is similar to that observed for developed markets, the number of eigenvalues deviating from the upper bound predicted by random matrix theory are smaller in number. Further, most of the observed correlations among stocks is found to be due to effects common to the entire market, whereas correlations due to interactions between stocks belonging to the same business sector are weak. This dominance of the market mode relative to modes arising through interactions between stocks makes an emerging market appear more correlated than developed markets. Using a simple two-factor model, we show that a market factor, that is dominant relative to the sector factor, results in spectral properties similar to that observed empirically for the Indian market. Our study helps in understanding the evolution of markets as complex systems, suggesting that strong interactions may emerge within groups of stocks as a market evolves over time. How such self-organization occurs and its relation to other changes that a market undergoes during its development, e.g., large increases in transaction volume, is a question worth pursuing in the future with the tools available to physicists.

This chapter also makes a significant point regarding the physical understanding of markets as complex dynamical systems. In recent times, the role of the interaction structure within a market in governing its overall dynamical properties has come under increasing scrutiny. However, such intra-market interactions affect

very weakly certain market properties, which is underlined by the observation of identical fluctuation behaviour in markets having very different interaction structures, viz., NYSE and NSE [85, 134]. For the purpose of explaining features such as the price fluctuation distribution, the system can be considered to be a single homogeneous entity responding only to external signals. This suggests that the earlier approach for studying financial markets that ignored their internal structure and considered prices to be essentially executing random walks in response to independent external shocks [135], may still be considered to be accurate for explaining market fluctuation phenomena. In other words, complex interacting systems like financial markets can have simple mean field-like description for some of their properties.

6

Role of network structure in system function

6.1 Introduction

The relatively simple nervous systems of invertebrate organisms provide vital insights into how nerve cells integrate sensory information from the environment, resulting in a coordinated response. Analysing the intermediate or mesoscopic level of organization in such systems is a crucial step in understanding how micro-level activity of single neurons and their interactions eventually result in macro-level behavior of the organism [136]. The nematode *Caenorhabditis elegans* is a model organism on which such an analysis can be performed, as its entire neuronal wiring layout has been completely mapped [18]. This information enables one to trace in full the course of activity along the neuronal network, from sensory stimulation to motor response [137]. We study its somatic nervous system, comprising 282 neurons that control all activity except the pharyngeal movements. This can lead to an understanding of the command and control processes occurring at the mesoscopic level that produce specific functional responses, including avoidance behavior and movement along a chemical gradient. The neuron locations as well as their connections being completely determined by the genetic program, are invariant across individual organisms. Further, unlike in higher organisms, the connections do not change with time in the adult nematode. In combination with the possibility of experimenting on the role of single neurons in different functional modalities, these invariances allow one to uniquely identify the important neurons in the system

having specific behavioral tasks.

The recent developments in the theory of complex graphs has made available many analytical tools for studying biological networks [4, 3]. The initial emphasis was on developing gross macroscopic descriptions of such systems using measures such as average path length between nodes of the network, the clustering among nodes and the degree sequences. However, such global characterizations of systems ignore significant local variations in the connection topology that are often functionally important. Therefore, investigating the network at a mesoscopic level which consider the broad patterns in the inhomogeneous distribution of connections, may reveal vital clues about the working of an organism that could be hidden in a global analysis. Further, these large-scale features help in understanding how coordination and integration occurs across different parts of the system, in contrast to a study of microscopic patterns comprising only a few neurons, e.g., motifs [138].

The existence of *modules*, marked by the occurrence of groups of densely connected nodes with relatively fewer connections between these groups [52], provides a natural meso-level description of many complex systems [67]. Modular organization in the brains of different species have been observed, both in *functional* networks derived from EEG/MEG and fMRI experiments and in *structural* networks obtained from tracing anatomical connections [139]. The functionally defined networks, where different brain areas, each of which comprise a large number of localized groups of neurons, are considered to be linked if they are simultaneously active, have been shown to be modular for both human [140] and non-human [141] subjects. Tract-tracing studies in the brains of cat [83] and macaque [142] have also revealed a modular layout in the structural inter-connections between different brain areas. However, as neurons are the essential building blocks of the nervous system, ideally one would like to explore the network of interconnections between these most basic elements. In the extremely complicated mammalian brains, it is so far only possible to analyze such networks for extremely limited regions that do not give a picture of how the system behaves as a whole [143]. The relative simplicity of the nervous system of C. elegans allows a detailed analysis of the network, defined in terms of both electrical (gap junctional) and chemical (synaptic) connections between the neurons (Fig. 6.1).

The ubiquity of modularity in brain networks leads to the obvious question about how to explain the evolution of such a structural organization [58]. One

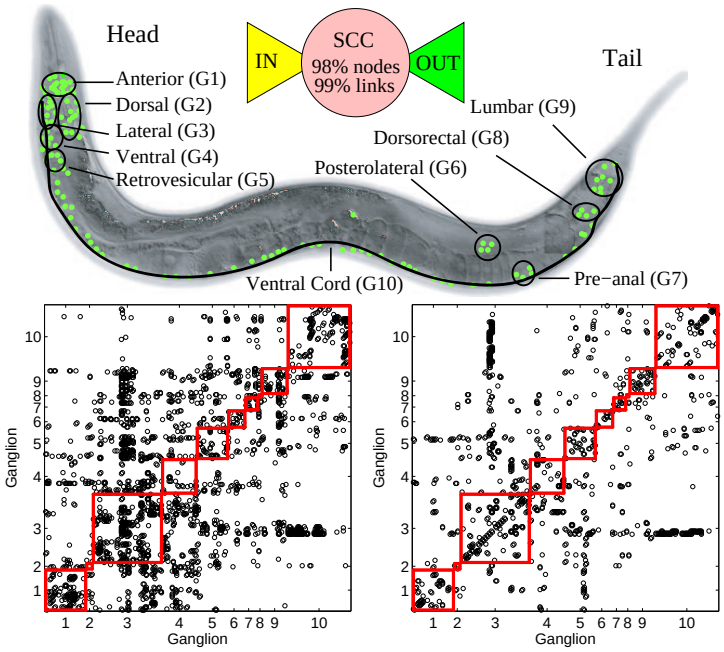


Figure 6.1: (a) Schematic diagram of C. elegans, indicating the different ganglia. (Inset) Schematic representation of connectivity between the neurons, partitioned into a *strongly connected component* (SCC), an *in-component* (IN), and an *out-component* (OUT). A directed path exists from any neuron in IN to any neuron in OUT through neurons in SCC, all of whose members can be reached from each other. The large SCC suggests that it is possible to transfer signals between almost all neurons of the network. The IN and OUT components have only 1.5% and 0.5%, respectively, of the 279 connected neurons in the somatic nervous system. (b, c) The connectivity matrix corresponding to the (b) Synaptic and (c) Gap-junctional connections between the somatic system neurons. In all figures, the partition symbols correspond to (G1) Anterior, (G2) Dorsal, (G3) Lateral, (G4) Ventral, (G5) Retrovesicular, (G6) Posterolateral, (G7) Preanal, (G8) Dorsorectal and (G9) Lumbar ganglion, and (G10) the Ventral cord.

possible reason for the existence of modular architecture is that they may result in low average path length (which is associated with high efficiency of signal communication) and high clustering (that allows local segregation of information processing) in networks [56]. An alternative possibility is that segregation of neurons into spatially localized communities minimizes the total cost associated with the wiring length (the physical distance spanned by connections between neurons). This cost arises from resources associated with factors such as wiring volume as well as metabolism required for maintenance and propagation of signals across long

distances [144]. Developmental constraints, such as the lineage relations between different neurons may also play an important role in determining the network topology [145]. In addition, the existence of empirically determined circuits responsible for specific functions (such as, movement associated with exploratory behavior, egg laying, etc.) in the C. elegans nervous system, raises the intriguing possibility that structurally defined modules are associated with definite functional roles [146]. The invariant neuronal connectivity profile of C. elegans allows us to explore the contributions of the above mentioned structural, developmental and functional constraints in governing the mesoscopic organization of the nervous system.

In Sec. 6.2, we begin our analysis of the organization of the C. elegans nervous system by identifying structurally defined modules in the network of neurons linked by synapses and gap-junctions. Next, we investigate whether the observed modular structure can be explained by using arguments based on universal principles. Such criteria, which include minimizing the cost associated with neuronal connections [144, 147] and their genetic encoding [148], or, decreasing the signal propagation path [149, 150], have recently been proposed to explain observed patterns of neuronal position and connectivity. We determine the role of physical proximity between a pair of neurons in deciding the connection structure, by investigating the correlation between their spatial positions and their modular membership. We also compare these modules with the existing classification of the nematode nervous system into several ganglia, as the latter have been differentiated in terms of anatomical localization of their constituent neurons. Results from the above analysis suggests that resource constraints such as wiring cost cannot be the sole deciding factor governing the observed meso-level organization. We also show that the modules cannot be only a result of the common lineage of their member nodes.

It is natural to expect that the structure of the nervous system is optimized to rapidly process signals from the environment so that the organism can take appropriate action for its survival [151]. By looking at the deviation between the actual network and a system optimized for maximal communication efficiency in conjunction with minimum wiring cost, we infer the existence of additional functional constraints related to processing of information (i.e., other than simple signal transmission). This is further supported by the observation of relatively high clustering in C. elegans neuronal network as compared to other information networks (e.g., electronic logic circuits [152]). As clustering increases the wiring

cost, while not assisting efficient communication, its presence in a system that has evolved under intense competition for survival may imply a key role for clustering in processing information. By looking at the correlation between local as well as global connectivity profiles with individual node characteristics, we observe that the nematode nervous system is significantly different from designed systems, including other information networks occurring in the technological domain, e.g., the internet. Further, in contrast to previous observations on the similarity between biological signalling networks having different origins [153, 154], we find that the C. elegans neuronal network has properties distinct from at least one other biological network that is involved in signalling tasks, namely the protein interaction network [155, 156].

Thus, the analysis of the network at the mesoscopic level provides an appropriate framework for identifying the roles that different classes of constraints (developmental, structural and functional) play in determining the organization of a nervous system. It also allows us to infer the existence of criteria related to processing of information governing the observed modular architecture in C. elegans neuronal inter-connections. Our results provide the means for identifying neurons having key roles in the behavioral performance of the organism exclusively from anatomical information about their structural connectivity. Our results can help experimentalists in focusing their attention to a select group of neurons which may play a vital part in some, as yet undetermined, function.

In Sec. 6.3, we investigate the hierarchical organization of the nematode neuronal network. We focus on the possible functional advantages of hierarchical structure and investigate its role in information processing. We show that the core-periphery structure of the network exhibits evidence of hierarchical ordering. By using a simple model for such networks, we see that such structures leads to significant reduction in the spread of local activity throughout the network while maintaining the communication efficiency almost unchanged.

6.2 Modularity in the C.elegans nervous system

In order to investigate the modular organization of a network, we need to first identify the modules. We perform this on the C. elegans neuronal network by carrying

out an optimal partitioning of the network, which corresponds to the maximum value of modularity parameter, Q. We use the generalization of a method introduced in Refs. [157, 158]. For a directed and weighted network, the modularity can be defined as

$$Q^W \equiv \frac{1}{L^W} \sum_{i,j} \left[W_{ij} - \frac{s_i^{\text{in}} s_j^{\text{out}}}{L^W} \right] \delta_{c_i c_j}, \tag{6.1}$$

where, $L^W = \sum_{i,j} W_{ij}$ is the sum of weights of all links in the network (W_{ij} is the weight of the link from neuron j to neuron i), and the weighted in-degree and out-degree of node i are given by $s_i^{\text{in}} = \sum_j W_{ij}$ and $s_i^{\text{out}} = \sum_j W_{ji}$, respectively. The optimal partitioning of the network is the one which maximizes the modularity measure Q (or Q^W). We obtain this by first defining a modularity matrix B,

$$B_{ij} = W_{ij} - \frac{s_i^{\text{in}} s_j^{\text{out}}}{L^W}. \tag{6.2}$$

To split the network, the eigenvectors corresponding to the largest positive eigenvalue of the symmetric matrix $(\mathbf{B} + \mathbf{B^T})$ is calculated and the communities are assigned based on the signs for the elements of the eigenvector. A repeated bisection method is then used to successively divide the obtained groups. The process terminates when further division does not increase the modularity of the network.

We have considered different cases corresponding to the different types of neuronal connections (viz., gap junction, synaptic and their combination) and the nature of such connections (viz., weighted or unweighted by the number of each type of connection). While the gap junctional network is undirected, both the synaptic as well as the combined network is directed. For each type of network, the maximum modularity value Q_M and the number of partitions for which this value is obtained, are given in Table 6.1. While using only the gap junctions fragment the network into as many as 15 modules, when we consider either synapses alone or the combination of both types of connections, the number of modules obtained is much less. In the remainder of this chapter, we have considered the combined network of synapses and gap junctions, unless otherwise stated. This is a weighted network where the link weights correspond to the total number of synaptic and gap junction connections from one neuron to another.

The high value of Q_M and dense inter-connectivity within modules (Fig. 6.2 (left)) suggest that the network has a modular organization. We further validate

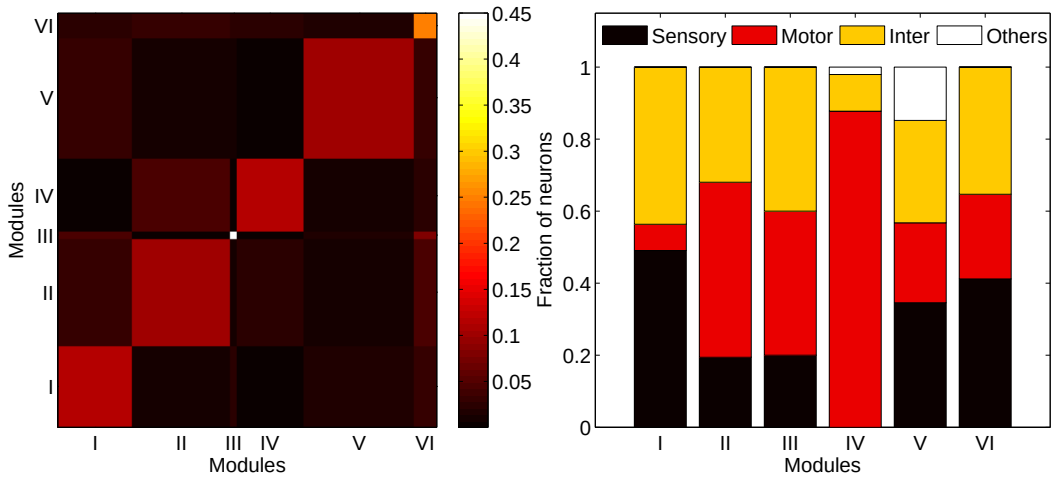


Figure 6.2: (Left) Matrix representing the average connection density between neurons occurring within modules and those in different modules. The figure indicates that neurons within a module are densely interconnected compared to the overall connectivity in the network. (Right) The modules are decomposed according to the different neuron types comprising them. The figure shows that the modules are not simply composed of a single type of neuron.

our results by considering the modularity of randomized versions of the network (keeping degree of each node fixed). The average modularity of these randomized networks is considerably lower then the empirical network. We have also analyzed the composition of the different modules in terms of distinct neuron types (viz. sensory, motor, interneuron and other). Fig. 6.2 (right) shows that none of the modules are exclusively composed of a single type, although motor neurons do tend to dominate one module.

6.2.1 Modules and spatial localization

To understand why modular structures have evolved in the neuronal network, we consider the relation between the optimal partition and the spatial localization of neurons in each module. This will help us to understand whether constraints related to the physical distance between neurons, such as wiring length optimization, dictate the topological organization of the network. Note that, wiring cost has already been shown to be the decisive factor for neuron positions in the body of C. elegans [144, 147]. Thus, a plausible hypothesis is that, if most neuronal

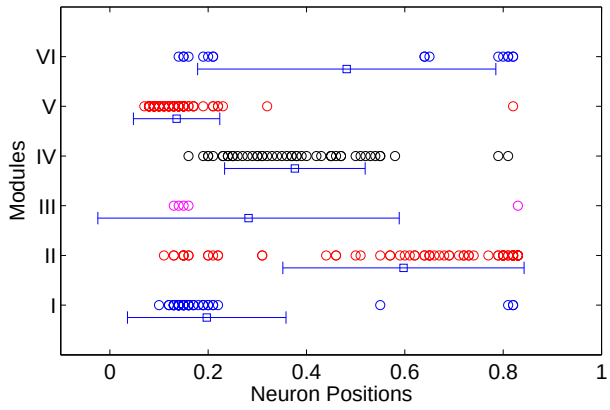


Figure 6.3: The position of neuronal cell bodies along the longitudinal axis of the C. elegans body plan is shown, with the vertical offset and color indicating the module to which a neuron belongs. The mean and standard deviation of neuronal positions for each module is also indicated, suggesting relative absence of spatial localization in the modules.

connections occur within a groups of neurons, all of whom are physically adjacent, then this will significantly decrease the wiring cost. This will be manifested, in terms of connectivity, as a module of the network. Therefore, it suggests that each module will mostly comprise of neurons which are in close physical proximity.

Fig. 6.3 indicates the spatial position of the cell body of each neuron on the nematode body (along the longitudinal axis), that belong to the different modules. It shows that, while a large fraction of the neurons belonging to the same module do indeed have their cell bodies close to each other, there is absence of any correlation between the modules and a specific physical location on the nematode body axis. This is brought out by the large standard deviations for the distribution of positions of the module components. Had wiring length minimization been the dominant factor governing the connectivity, we would have expected the modules to be spatially segregated, as in that case, most of the connections need to span only short distances.

Our conclusion that wiring cost cannot be the principal criterion determining network topology is further supported by an analysis of the connectivity in the different ganglia. The nine different ganglia of the nematode nervous system, in addition to the ventral cord, are defined in terms of physical proximity of their component neurons. Thus, a tendency to minimize the total connection length

Table 6.1: The modularity of the network is measured using the parameter Q, which requires a knowledge of the partitions or communities which divide the network. We obtain the modularity measure, Q_g, on assuming the communities to correspond to the ganglia. Its positive values indicate that neurons in the same ganglion have high density of inter-connections. We have also obtained Q by determining the modules of the network using a spectral method, the corresponding values being indicated by Q_M. The relatively high values of Q_M compared to Q_g, indicates that the ganglia do not match with this optimal partitioning of the network. Both measures, Q_g and Q_M, as well as the number of modules, n_M, have been obtained for both unweighted and weighted networks consisting of either gap junctions or synapses or both. We calculate the overlap between the ganglionic and the optimal partition of the network using the normalized mutual information index, I. For the case of perfect match between the two, the index, $I = 1$, whereas if they are independent of each other, $I = 0$. The measured values of I indicate that the overlap between the different modules and the anatomically defined ganglia is not significant. The modular nature of the somatic nervous system is emphasised by comparing the empirical network with networks obtained by randomizing the connections, keeping the degree of each neuron fixed. The mean and standard deviation of the modularity Q_M^{rand} and the corresponding number of partitions m_M^{rand} are shown for both weighted and unweighted networks, and for the different types of connections. For all cases, the randomized networks show a significantly lower modularity than the empirical network.

Network	Unweighted				Weighted			
	Q_g	Q_M	m_M	NMI	Q_g	Q_M	m_M	NMI
Gap Jn	0.2069	0.6297	11	0.3264	0.1700	0.6566	15	0.3473
Synaptic	0.1487	0.3491	2	0.2572	0.2106	0.4720	4	0.3137
Combined	0.1687	0.3776	3	0.3057	0.2031	0.4910	6	0.3763

between neurons would imply that such ganglia would be distinguished by having a significantly high density of connections between their constituent neurons. We, therefore, verify whether the connectivity inside each ganglion is high compared to the corresponding randomized networks. This is done by measuring the modularity value Q_g, with each ganglion assumed to be a true network module. We find that, although Q_g is nonzero which indicates that the neurons inside a ganglia do indeed have a higher connection density than the overall network, it is not as high as the maximum Q_M possible, obtained for the optimal partition (Table 6.1). We have also calculated the normalized mutual information index, NMI, to measure the overlap between the optimal partitioning of the network into modules and the different ganglia. Note that, in the case of perfect match, $NMI = 1$, while it is 0, if there is no match. The low values for NMI given in Table 6.1 suggest that

the composition of the different ganglia is quite distinct from that of the modules for the optimal partitioning of the neuronal network. This is shown explicitly in Fig. 6.4 (a), indicating that most ganglia are composed of neurons belonging to many different modules.

We have used this *modular decomposition* spectrum of each ganglion (i.e., the distribution of the neurons of the ganglion into the m different modules of the optimal partition) to define a metric for inter-ganglionic distance in an abstract m-dimensional "modular" space. Thus, if two ganglia have a similar modular decomposition profile, then they are close to each other in the "modular" space and have low modular distance (Fig. 6.4, b). This is then compared with the physical distance between ganglia, measured as the average separation between the cell bodies of all pairs of neurons i and j, where i, j belong to different ganglia (Fig. 6.4, c). The comparison of the two matrices shows that there are indeed certain similarities between these two different concepts of distance. For example, the five ganglia located in the head (G1-G5) cluster together, as do the three located towards the tail (G7-G9). This observation is in accord with previous reports which use the notion of wiring cost for explaining (to a certain extent) the observed relative positions of the ganglia. However, when we consider the corresponding dendrograms that indicate the relative closeness of the different ganglia in physical space and in "modular" space, we observe significant differences between the two: ganglia which are close to each other in physical space may not be neighbors in terms of their modular spectra. This reiterates our previous conclusion that wiring cost minimization, which is related to the physical distance between neurons, is not a dominant factor governing the organization of C. elegans somatic nervous system.

6.2.2 Modules and cell lineage

As developmental processes are believed to play a significant role in determining the structure of the nervous system, we also consider the alternative hypothesis that the structural modules reflect a clustering of neurons that are related in terms of their lineage. Lineage of a cell is the pattern of successive cellular divisions that occur during its development. This is invariant in C. elegans, allowing one to trace the individual developmental history of each cell in order to identify the cell-autonomous mechanisms and cell-cell interactions. We measure the average

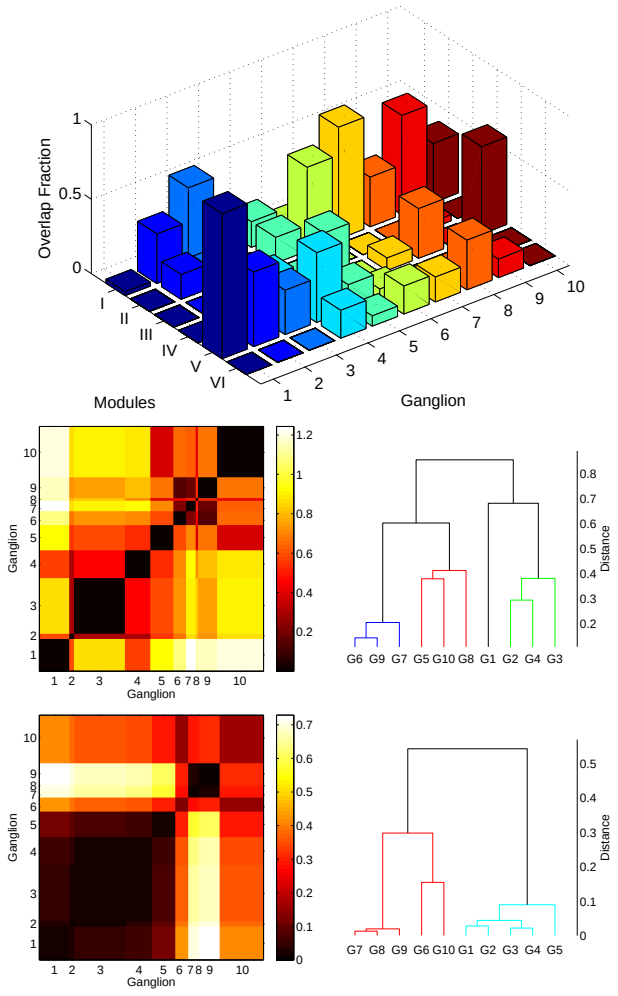


Figure 6.4: (a) Neurons belonging to different ganglia are decomposed according to their modular membership. The height of each bar in the histogram corresponds to the overlap between the ganglia and the modules, calculated as the fraction of neurons that are common to a particular ganglion and a specific module. (b) The matrix representing the average modular distance between the different ganglia, as calculated from the modular decomposition spectrum of each ganglion. The corresponding dendrogram indicates the closeness between different ganglia in the abstract 6-dimensional "modular" space. (c) The matrix of physical distances between the ganglia is shown for comparison with (b), calculated as the average distance between neurons belonging to the different ganglia. The corresponding dendrogram indicates the closeness between ganglia according to the geographical nearness of their constituent neurons in the nematode body. The difference indicates that the ganglia which are geographically close may not be neighbors in terms of their modular spectra.

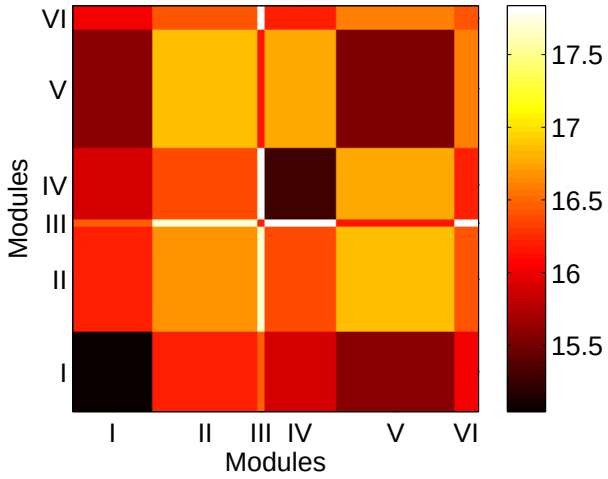


Figure 6.5: The matrix representing the average lineage distance between neurons occurring within the same module and those belonging to different modules. The figure indicates that neurons occurring in the same module have only a slightly lower lineage distance as compared to that between neurons occurring in different modules.

relatedness measure of neurons within and between modules. Fig 6.5 suggests that there is no significant segregation between the modules in terms of the lineage of their constituent neurons. Indeed, even coarse distinctions such as AB and non-AB lineage neurons are not apparent from the modular division. A detailed view of the relatedness between each pair of neurons (figure not shown) indicates that, while in each module there are subgroups of closely related neurons, different subgroups within the same module may be very far from each other in the lineage tree. Conversely, neurons occurring in different modules can have small distance in terms of lineage. This suggests that developmental constraints do not completely dictate the connection structure of the neuronal network. The fact that C. elegans neurons are largely non-clonally derived from many different parent cells [159] may partly explain this lack of correlation between lineage defined in terms of common ancestry and modules defined in terms of density of inter-connections.

6.2.3 Modules and functional circuits

Next, we look at the overlap of seven previously identified functional circuits of C. elegans with the structural modules. Note that, if a significant correlation is

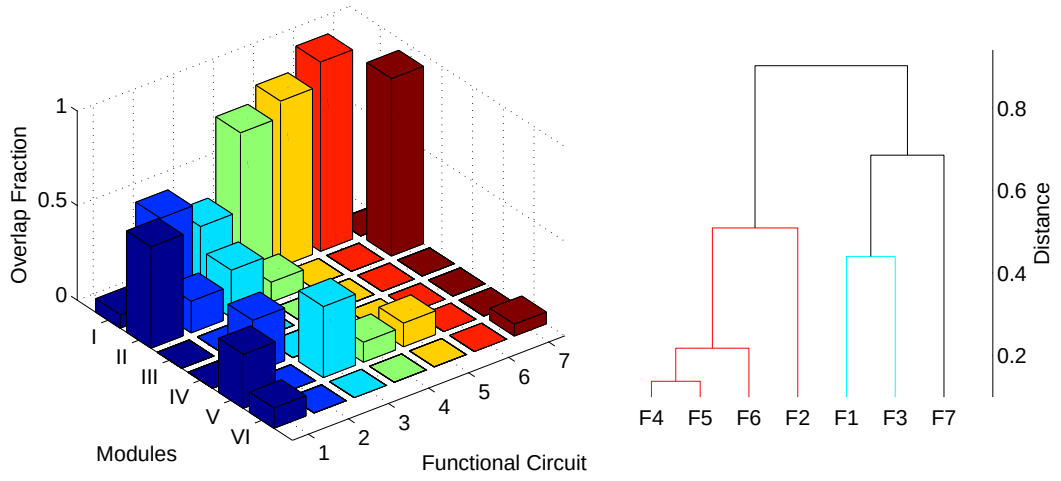


Figure 6.6: Neurons belonging to different functional circuits are decomposed according to their modular membership. The height of each bar in the histogram corresponds to the overlap between the modules and functional circuits (F1) mechanosensation, (F2) egg laying, (F3) thermotaxis, (F4) chemosensation, (F5) feeding, (F6) exploration and (F7) tap withdrawal. The overlap is measured in terms of the fraction of neurons common to a particular functional circuit and a specific module. The corresponding dendrogram represents the closeness between different functional circuits in the abstract 6-dimensional "modular" space.

observed between the two, then it may suggest means of identifying neurons responsible for certain functions from the information about the connection topology of the network. Functional circuits are subsets of neurons which are believed to play a vital role in performing a specific task, and are obtained by observing abnormal behavior of the organism when the neurons are individually removed from the nematode nervous system (e.g., by laser ablation). We consider the functional circuits for (F1) mechanosensation [160, 161, 162], (F2) egg laying [163, 164], (F3) thermotaxis [165], (F4) chemosensation [166], (F5) feeding [167, 160, 18], (F6) exploration [167, 160, 18] and (F7) tap withdrawal [168, 161].

Fig. 6.6 shows the modular decomposition of each functional circuit. The corresponding dendrogram clusters the circuits, in terms of the similarity in their modular spectra. We note that the circuits for chemosensation, feeding and exploration are clustered together. This is consistent with the fact that most of the neurons belonging to the feeding and exploration circuits are involved in chemosensation. However, none of the neurons in the feeding circuit are common to those

in the exploration circuit, although F6 is closer to F5 in terms of their modular spectra (distance = 0.18) than it is to F4 (distance =0.26) of which F6 is a subset. This implies that these two circuits F5 and F6 are strongly connected, thereby indicating the inter-relation of the corresponding functions. Indeed the feeding behavior of C.elegans is known to be regulated in a context-dependent manner by its nervous system which integrates external signals, such as, the availability of food and the nutritional status of the animal, to direct an appropriate response [169]. The mode of locomotion is also determined by the quality of food [170]. It is apparent that, integration of multiple behaviours is essential to feeding regulation [171], such as avoidance of high CO_2 concentrations by satiated animals [172]. Furthermore, F2 is seen to be closer to these circuits, which is significant in light of previous experimental observation that presence of food (as detected through chemosensory neurons) modulates the egg-laying rate in C. elegans [162, 173].

6.2.4 Functional role of different neurons

In this subsection, we establish the functional importance of specific neurons by investigating the role played by them within their module, and compare it with their role in the entire network. To parametrise this, we use (i) the participation coefficient, P, which is a measure of how dispersed the connections of a node are among the different modules, and, (ii) the within-module degree, z, that indicates the number of connections a node has to other members of its module. For each node i, we can define its participation coefficient P_i as

$$P_i = 1 - \sum_{s=1}^{m} \left(\frac{\kappa_s^i}{k_i} \right)^2, \tag{6.3}$$

where κ_s^i is the number of links to nodes in module s $k_i = \sum_s \kappa_s^i$ is its degree. Thus, the participation coefficient of a node is ~ 1 if its links are uniformly distributed among all modules and $= 0$ if it only links to nodes within its own module. The within-module degree, z, distinguishes nodes that play the role of hubs in their own module from non-hub and peripheral nodes. For the i-th node, it is defined as

$$z_i = \frac{\kappa_{s_i}^i - \langle \kappa_{s_i}^j \rangle_{j \in s_i}}{\sqrt{\langle (\kappa_{s_i}^j)^2 \rangle_{j \in s_i} - \langle \kappa_{s_i}^j \rangle_{j \in s_i}^2}}, \tag{6.4}$$

where κ_s^i is as defined above and the average $\langle \cdots \rangle_{j \in s}$ is performed over all nodes in the module s.

A node having a low within-module degree is called a non-hub node ($z < 0.7$). Such nodes can be further classified according to their fraction of connections with other modules, which following Ref. [156] are classified as: (R1) ultra-peripheral nodes ($P \leq 0.05$), having connections only within their module, (R2) peripheral nodes ($0.05 < P \leq 0.62$), which have a majority of their links within their module, (R3) satellite connectors ($0.62 < P \leq 0.8$), with many links connecting nodes outside their modules and (R4) kinless hubs ($P > 0.8$), which form links uniformly across the network. Hubs, i.e., nodes having a large number of connections to nodes within their module ($z \geq 0.7$), are further sub-divided according to their participation coefficient into (R5) provincial hubs ($P \leq 0.3$), with most connections within their module, (R6) connector hubs ($0.3 < P \leq 0.75$), with a significant fraction of links distributed among many modules and (R7) global hubs ($P > 0.75$), which connect homogeneously to all modules. This classification allows us to distinguish nodes according to their different roles as brought out by their intra-modular and inter-modular connectivity patterns.

We will now use the above methodology on the C. elegans network in order to identify neurons that play a vital role in coordinating activity through sharing information (either locally within their community or globally over the entire network). Fig. 6.7 shows the comparison between the empirical network and a corresponding randomized network (obtained by keeping the degree of each node fixed). We immediately notice that the randomized network does not have any nodes having the roles R1 and R5, indicating that the modular nature of the original network has been lost. In fact, in the randomized system, most nodes have higher participation coefficient, with a large majority being satellite connectors (R3). More interesting is the fact that, the empirical neural network does not posses any neuron having the global roles played by R4 and R7, whereas the randomized network does. This implies that modular identity in the C. elegans neuronal network is very pronounced.

It is possible to relate the intra- and inter-modular connectivity patterns of a neuron with its role in the functioning of the worm nervous system. For example, neurons having the role of provincial hubs may be involved in local coordination of neural activity, while, the connector hubs may be responsible for integration of

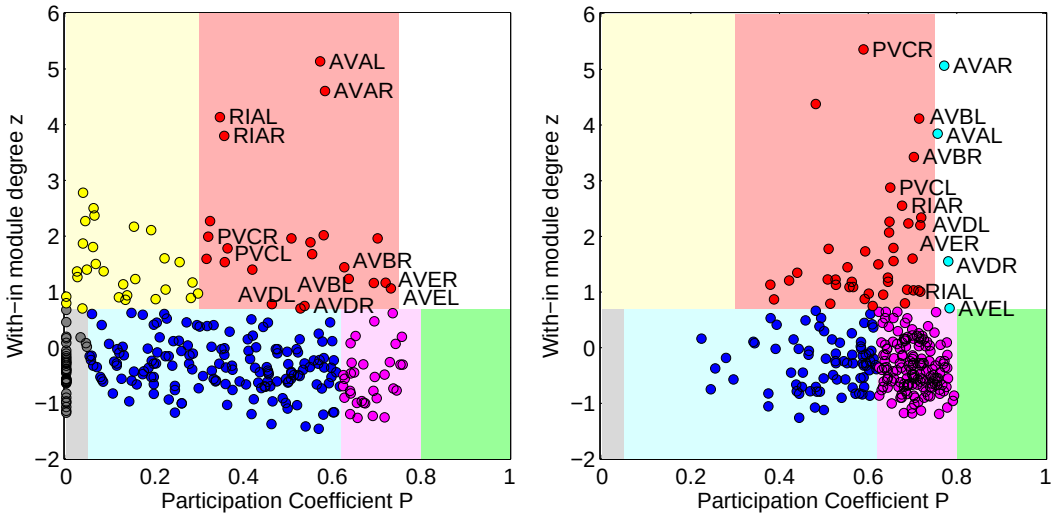


Figure 6.7: (Left) The within module degree z-score of each neuron in the empirical neuronal network is shown against the corresponding participation coefficient P. The within module degree measures the connectivity of a node to other nodes within its own module, while the participation coefficient measures its connectivity with neurons in the entire network. (Right) The corresponding result for a randomized version of the C. elegans network where the degree of each neuron is kept unchanged is also shown. Neurons belonging to the different regions in the $P - z$ space are categorised as: (gray) "ultra-peripheral nodes," i.e., nodes with all their links within their module, (blue) "peripheral nodes," i.e., nodes with most links within their module, (pink) "nonhub connector nodes," i.e., nodes with many links to other modules, (green) "nonhub kinless nodes," i.e., nodes with links homogeneously distributed among all modules, (yellow) "provincial hubs," i.e., hub nodes with the vast majority of links within their module, (red) "connector hubs," i.e., hubs with many links to most of the other modules, and (white) "global hubs," i.e., hubs with links homogeneously distributed among all modules. The neurons occurring as connector hubs are identified in the figure. Most of these neurons occur in different functional circuits indicating the close relation between functional importance and connectivity pattern of individual neurons.

local activities to produce a coherent response of the entire system. This hypothesis is supported by the observation that, all command interneurons (of the class AVA, AVB, AVD, AVE, PVC), which control forward and backward locomotion of the worm by regulating motor output, play the role of connector hubs. In fact, out of the 23 neurons in the class R6, 20 are known to belong to different functional circuits. Among the rest, although DVA is not part of any known circuit, it has recently been identified as being involved in mechanosensory response. In its

absence, the frequency and magnitude of the tap-induced reversal as well as the acceleration magnitude is diminished [168]. The two remaining neurons, AVKL and SMBVL, have not been implicated so far in any known functional circuit. However, their occurrence in this class suggests that they may be important for some, as yet unknown, function. This is a potentially interesting prediction that may be verified in the laboratory.

The significance of these results is underlined by a comparison with the randomized network. For instance, in the random realization shown in Fig. 6.7 (b), of the 49 neurons playing the role of connector or global hubs, less than half (viz., 23) actually belong to any of the known functional circuits. The appearance of most of the command interneurons in the high-z region of both the empirical and randomized networks indicates that their high overall degree is responsible for their observed role of "connecting hubs".

We now turn to the 28 neurons which play the role of provincial hubs. Half of all the inhibitory D-class motorneurons (viz., DD1-DD3 and VD1-VD6) are found to belong to this class. This is significant as these neurons have already been implicated in the ability of the worm to initiate backward motion. While they also contribute to forward locomotion, previous experiments have shown that they are not crucial for it [174]. This fits with our hypothesis that, R5 neurons are important for local coordination but may not be crucial for the global integration of activity. A pair of excitatory B-class motorneurons that sustain coordinated forward locomotion in the worm also appear as provincial hubs. Of the remaining R5 neurons, 9 have been previously identified to belong to various functional circuits. It will be interesting to verify the functional relevance of the remaining 8 neurons (OLLL/R, RMDVL/R, SMDVR, RIH, RMDDL/R) in the laboratory. Thus, overall, we find a very good correlation between the connectivity pattern and the functional importance of different neurons.

An analysis of neurons having different roles in terms of their membership in the different ganglia indicates that the lateral ganglion provides the majority of neurons acting as connector hubs (R6). This is consistent with an earlier study [175] where this ganglion was found to act as the link between the neuronal groups responsible for sensory processing and motor response. The corresponding randomized network, while also showing many neurons from the lateral ganglion, have significant representation from other ganglia too (e.g., the retrovesicular ganglion).

We have also carried out an analysis of the probability of connections between neurons having different roles, relative to the randomized network, in an attempt to compare with other networks which are involved in transportation and information propagation. However, we do not find any significant overlap of the nematode nervous system with networks in either of these two classes, suggesting that the C. elegans neuronal network does not belong exclusively to either class of networks. This assumes significance in light of recent work distinguishing information (or signalling) networks, such as the Internet and protein interactome, on the one hand, and transportation networks, such as metabolic and airport networks, on the other, into two classes [156].

6.2.5 Wiring cost vs communication efficiency

In this subsection, we investigate certain global properties of the C. elegans neural network in order to determine its differences with other classes of networks. We also seek to ascertain the possible constraints which might have given rise to the observed network topology.

We have already shown that wiring cost minimization, is at best, only a partial determining factor for network structure in this system. A recent study [149] has claimed that neuronal networks minimize the length of processing paths (i.e., the average number of links connecting any given pair of neurons) rather than minimizing the total wiring length or the average physical distance between connected neurons. Thus, we now look at the communication efficiency of the network, as measured by the harmonic average network path length between all pairs of neurons. As increasing the efficiency inevitably results in increasing the wiring cost, we analyze how the performance of the network as an information propagation system competes with the resource cost involved in setting up the required number of connections. This cost is measured as the Euclidean length between the cell bodies of all connected pairs of neurons (corresponding to the dedicated-wire model of Ref. [144]). It has been shown that the positions of the subset of sensory and motor neurons directly connected to sensory organs and muscles, respectively, can be determined quite accurately by minimizing their total wiring cost [147]. As our focus is on the connection structure of the neuronal network, we keep the neuron positions fixed and, thus, our study does not consider the wiring cost involved in

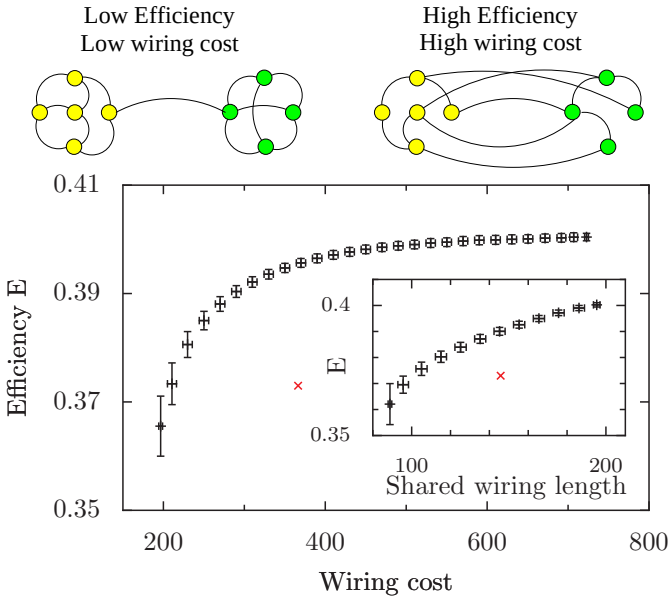


Figure 6.8: The variation of communication efficiency, E, as a function of the wiring cost (defined using the "dedicated-wire" model) in the ensemble of random networks with degree sequence identical to the C. elegans neuronal network. The trend indicates a trade-off between increasing communication efficiency and decreasing wiring cost. The corresponding values for the empirical network is indicated by a cross. The schematic figures shown above the main panel indicate the type of networks obtained in the limiting cases when only one of the two constraints are satisfied. (Inset) The communication efficiency of randomized networks as a function of the wiring cost calculated according to the "common-wire" model. The values for the empirical network are indicated as before. In both figures, error bars indicate the standard deviations calculated for 10^3 random realizations. We observe that the empirical network is suboptimal in terms of wiring cost and communication efficiency, suggesting the presence of other constraints governing the network organization.

connecting neurons to sensory organs and muscles. By randomizing the network keeping the degree of each node fixed, we can construct systems having a specified wiring cost, and then measure its communication efficiency. This analysis reproduces the expected result that decreasing the wiring cost of the network causes a decline in its performance in terms of information propagation. However, we note that the empirical network has a wiring cost much higher than one would expect for the corresponding communication efficiency.

As most of the neurons in C. elegans have at most one or two processes, on which all the synapses and gap junctions with other neurons are made, the earlier

definition of wiring cost as the sum of all Euclidean distances between connected cell bodies may be a gross over-estimate of the actual wiring cost. Thus, we consider an alternative measure, where, the wiring cost for connecting to a specific neuron is taken to be the Euclidean length between the neuron's cell body and those of the farthest neurons it is connected to. The simple one-dimensional simplification of the C. elegans body that we have used here ignores distance along the transverse plane. Thus, this measure is actually an under-estimate of the actual wiring cost, and should provide an insightful comparison with the above measure obtained from the dedicated-wire model. Fig.6.8 shows that wiring cost increases with communication efficiency for the randomized networks, which is qualitatively similar to the relation obtained using the preceding definition for wiring cost. In this case also, we find that the empirical C. elegans network has a much lower efficiency than would be expected from its wiring cost alone. This observation suggests the presence of other constraints, possibly related to information processing, that are responsible for the observed global properties of the network.

6.2.6 Possible existence of information processing constraint

In this subsection, we explore further the possibility that the additional constraints governing the topological structure of C. elegans nervous system may be related to information processing, rather than constraints arising from information (or signal) propagation, which are quantified by measuring communication efficiency. The property of information processing, i.e., the active transformation of signals into responses, differentiates the neuronal network from other well-studied networks where efficiency is of paramount importance, such as the internet and airport transportation network. While fast communication of information between different neurons is certainly an important functional criterion, we explore the possibility that the neuronal network may be different from networks which are only optimized for maximum communication efficiency.

First, we look at the overall network structure, by decomposing the network into a strongly connected component (SCC), within which it is possible to visit any node from any other node using directed links, an inward component (IN) and outward component (OUT), consisting of nodes from which the SCC can be visited or which can be visited for the SCC, respectively, but not vice versa. In addition, there are

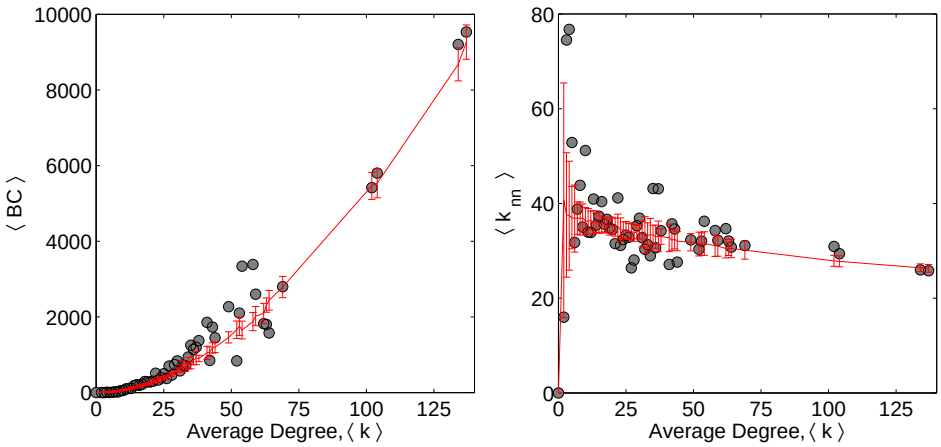


Figure 6.9: (Left) The average betweenness centrality, $\langle BC \rangle$, and (right) the average nearest neighbor degree, $\langle k_{nn} \rangle$ of each node as a function of its total degree, $\langle k \rangle = \langle k_{\text{in}} + k_{\text{out}} \rangle$. Betweenness centrality is a measure of how frequently a particular node is used when a signal is being sent between any pair of nodes in the network using the shortest path. In case of the internet, BC of nodes increases with its degree which is sought to be linked with its information transport property. In C. elegans, although BC increases with degree, this increase is not significant when compared to the randomized version of the network. In the case of the relation between the average connectivity of nearest neighbors of a node with its total degree k, we note that for both the internet and protein interaction network, k_{nn} decreases with k as a power law. This means that low connectivity nodes have high degree nodes as their neighbors and vice-versa. However, in the case of C. elegans, this relation is not very apparent and insignificant in comparison with the randomized version of the network. In both figures, error bars indicate the standard deviations calculated for 10^3 random realizations. These results suggest that the C.elegans network forms a class distinct from the class of networks optimized only for signal propagation.

components disconnected from the SCC, i.e., nodes which cannot be visited from SCC nor can any visits be made to SCC from there (Fig. 6.1). A comparison of the C. elegans neuronal network with a similar decomposition of the WWW [176] reveals that while in the latter the different components are approximately of equal size, the SCC of the nervous system comprises almost the entire network. Thus, any node can, in principle, affect any other node in the nervous system, suggesting the importance of feedback control for information processing.

Next, we consider the relation between two fundamental properties of the network: the degree of nodes and their Betweenness Centrality (BC), which char-

acterizes the importance of a node in information propagation over the network. The Betweenness Centrality of a node i is defined as the fraction of shortest paths between all pairs of nodes in the network that pass through i [177]. We observe that for both the C. elegans neuronal network and its randomized versions, the degree of a node and its BC are strongly correlated, i.e., highly connected nodes are also the most central (Fig. 6.9, left). This is similar to what has been observed in the internet [178], where the highest degree nodes are also those with the highest betweenness [179], but in sharp contrast to the airport transportation network, where non-hub nodes (low degree) may have very large BC [180].

However, the C. elegans neuronal network differs from both the internet and the protein interaction network (PIN), whose primary function is to allow signal propagation between nodes, in terms of the variation of the degree of a node i with the average degree of its neighboring nodes, $\langle k_{nn} \rangle$. While in the internet and PIN, $\langle k_{nn} \rangle$ decays as a power law with node degree, in the neuronal network, this dependence is very weak (Fig 6.9, right). This implies that unlike the internet and PIN, the C. elegans nervous system does not have multiple star-like subnetworks. Further, it is significantly different from the airport transportation network, where the high degree nodes are closely connected among themselves showing an assortative behavior [30]. Thus, we believe that these results strengthen our claim that there are additional constraints governing the nervous system connection topology in C. elegans, which are unrelated to wiring cost, lineage or communication efficiency. As the principal function of the system is to process information, this leads us to conjecture that it is this that provides the additional constraints leading to the observed organization of the nematode neuronal network. We further explore this possibility using a simple model of hierarchical networks in the next section.

6.3 Role of hierarchical organization in neuronal networks

As seen from the preceding analysis, it seems that the primary functional role of nervous system is to process information about the sensory environment so as to allow the organism to act appropriately for survival. Apart from the speed of signal propagation between various elements of the nervous system, the survival success

Table 6.2: The average directed path length within and between the sub-populations of sensory, motor and inter-neurons (ℓ) and the corresponding fraction of neuronal pairs for which a directed path does not exist (F_∞) are compared with the randomized versions of the network. The columns and rows indicate the neuronal types for the pre- and post-synaptic ends of a pair, respectively. The empirical network shows a relatively higher connectivity from sensory to both inter- and motor neurons, and from inter- to motor neurons, as compared to the reverse directions. For random networks, the average value is shown with the standard deviation given in parenthesis.

Neuron	Empirical network					
	Sensory		Motor		Inter	
	ℓ	F_∞	ℓ	F_∞	ℓ	F_∞
Sensory	3.46	10.24	4.75	27.96	3.37	23.04
Motor	3.00	1.87	3.79	20.88	2.74	15.72
Inter	2.86	5.88	4.05	24.31	2.72	18.96
Neuron	Randomized network					
	Sensory		Motor		Inter	
	ℓ	F_∞	ℓ	F_∞	ℓ	F_∞
Sensory	3.05 ± 0.03	9.37 ± 0.69	3.38 ± 0.04	24.43 ± 0.90	2.90 ± 0.03	22.25 ± 0.58
Motor	2.83 ± 0.02	2.12 ± 0.52	3.16 ± 0.03	18.10 ± 0.83	2.68 ± 0.02	15.91 ± 0.42
Inter	2.64 ± 0.02	6.03 ± 0.40	2.97 ± 0.03	21.53 ± 0.77	2.50 ± 0.02	19.06 ± 0.33

of an organism in the wild is crucially dependent on the specificity of response to the relayed stimulus. This latter criterion is necessary for the robustness of interpretation and processing of sensory information. It requires a high degree of control and coordination of activity in the network, in order to channel the signals through a limited part of the system, and preventing a local stimulation from spreading into an overall excitation of the entire system. A simple schema for a nervous system of an organism that has to execute a number of behavioral tasks would be one having a parallel set of neurons arranged into several levels. This constitutes a hierarchical system, each element of which is unambiguously assigned to a distinct layer of a given rank [181]. The original notion of hierarchical information processing, e.g., that proposed by Hubel and Wiesel to explain the progressive increase in complexity of the receptive field properties in the visual cortex [182, 183], had suggested a simple feed-forward scheme.

To compare this simplified system with the actual neuronal network, we first focus on the directionality of connections between the different neuron types, viz., sensory, motor and inter-neurons. As gap junctional connections are undirected and communication through them is much slower than that via chemical synapse [184], we have initially considered only the synaptic network for our anal-

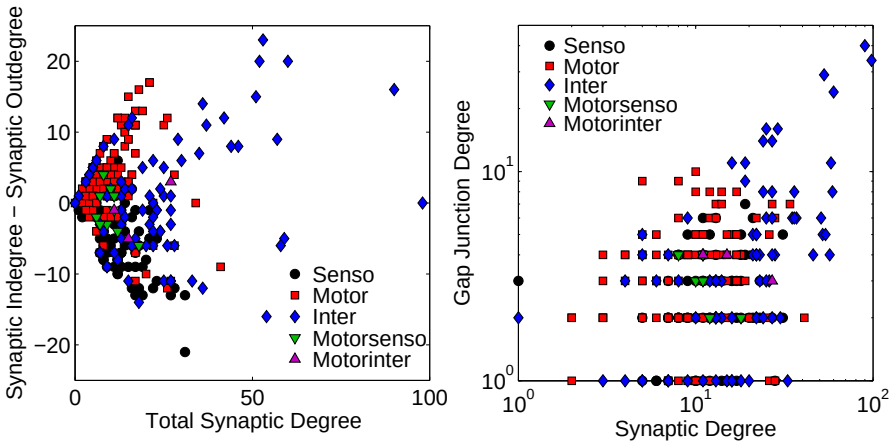


Figure 6.10: (Left) Scatter plot of the net difference between incoming and outgoing synaptic connections of a neuron as a function of its total synaptic degree. Different neuron types, viz., sensory, motor, inter, motor-sensory and motor-inter, are indicated by different symbols. The sensory and motor neurons have relatively higher out- and in-degrees respectively, whereas inter-neurons do not show any consistent pattern. This is consistent with the overall directed nature of information flow from sensory to motor layers as expected in a neuronal network. (Right) Gap-junctional degree as a function of the total synaptic degree, indicating that neurons with large number of synapses also have many gap junctional connections.

ysis. Table 6.3 shows the average length, ℓ, of *directed* paths within and between the different categories of neurons, as well, as the fraction of neurons (F_∞) of a given type that are unreachable from other neurons in the same or different category. We observe that there is indeed an asymmetry in these properties of directed paths between the different neuronal types, indicating a preferred direction of information flow from sensory to motor neurons via inter-neurons. However, there is far more connectivity in the reverse direction than would be expected for a simple hierarchical system having a feed-forward architecture.

These reverse (or feedback) connections are presumably responsible for the control of motor response, as well as, for the various kinds of associative and non-associative learning reported for C. elegans [185]. To explore in detail the feedback connectivity, we re-calculate F_∞ from motor to sensory neurons after removing the inter-connections among the motor neurons. The increase in the unreachable fraction from $\sim 28\%$ to $\sim 66\%$ indicates that, the feedback from motor to sensory neurons occur pre-dominantly through a few key motor neurons which send con-

nections to various inter- and sensory neurons, and to which other motor neurons are connected. In addition, if we also remove the direct feedback connections from the motor to sensory neurons, as well as, the inter-connections among the sensory neurons, the corresponding F_∞ increases to 71.6%. This suggests that these shortcuts from motor to sensory neurons are not essential for feedback communication, because the dominant contribution to the reverse flow of information from the output to input layers is through inter-neurons.

Another feature of the empirical network that suggests it does not follow a strictly hierarchical scheme having a sequence of several clearly defined levels between the input and output layers, is the existence of direct links from sensory to motor neurons. Sensory neurons have short directed paths to almost all motor and inter-neurons, with only the neurons AINL, PVDR, SDQR and DVB being unreachable from any sensory neuron. The shortest directed path length to a motor or inter-neuron from at least one sensory neuron (if a path exists) does not exceed 3, with 111 paths of length 1, 66 having length 2 and 8 having length 3. This is consistent with the fact that C. elegans neurons do not, in general, use regenerative action potentials to communicate with each other, which implies that signals will dissipate in transit unless the neurons are connected via short paths.

The directionality of the empirical network can be explained to an extent as an outcome of the degree sequence of the different categories of neurons. This is done by comparing with the degree-conserved random networks, which show similar asymmetry in the directed connectivity between the various neuronal types. This suggests that the differences in the in-degree and out-degree between sensory, motor and inter-neurons account for most of the variation in ℓ and F_∞ seen in the actual data. Fig 6.10 (left) explicitly shows that the sensory and motor neurons differ from each other in having relatively higher number of outgoing and incoming connections, respectively. In contrast, inter-neurons do not show any specific bias in terms of in-degree and out-degree. Moreover, we find that, neurons having high synaptic degree also tend to have high gap-junctional degree (Fig. 6.10, right).

6.3.1 Core-periphery organization of the nervous system

The above analysis suggests that the organization of the C. elegans neuronal network is much more complex than that of a simple feed-forward information process-

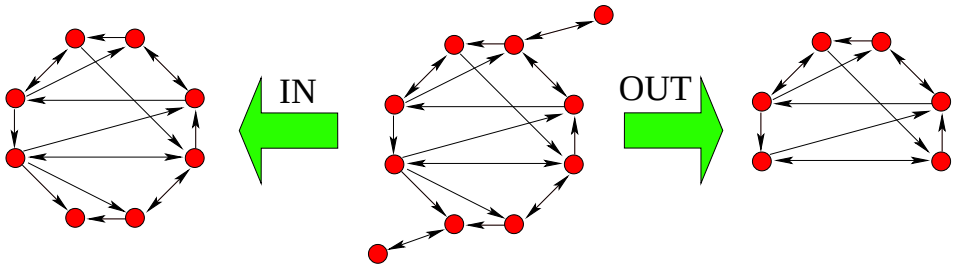


Figure 6.11: Schematic diagram showing the k-core decomposition of a directed network. The k-core of a graph is defined as the largest subgraph where every node has at least k links. For each choice of k, we determine the k-cores by iteratively pruning all nodes with degree lower than k and the links associated with them. For directed networks, the k-core obtained is different depending on whether one is considering the in-degree or the out-degree of the nodes.

ing network. The relatively high density of recurrent connections among the inter-neurons ($C_{inter-inter} = 0.068$, compared to $C_{sensory-sensory} \approx C_{motor-motor} = 0.026$) suggests that the empirical network may have a hierarchical arrangement consisting of a densely connected core comprising mostly inter-neurons and a sparsely connected periphery which is populated predominantly by sensory and motor neurons. To establish this, we analyze the network using k-core decomposition. The k-core of a graph is defined as the largest subgraph where every node has at least k links. For each choice of k, we determine the k-core by iteratively pruning all nodes with degree lower than k and their associated links. For a directed network, the k-core obtained depends on whether one is considering the in-degree or the out-degree of the nodes. Thus, each node of the in-degree k-core of a graph has at least k incoming links and each node of the out-degree k-core of a graph has at least k outgoing links (Fig. 6.11).

Our results show that there is indeed a set of inter-nested cores in the C. elegans nervous system, going upto 7 orders for cores defined in terms of out-degree. The composition of the cores according to different neuron types is even more illuminating (Fig. 6.12 A,B). While the innermost core defined in terms of in-degree consist almost entirely of motor and inter-neurons, those defined in terms of out-degree are dominated by sensory and inter-neurons. The intersection of the innermost cores for in- and out-degree has 25 neurons, comprising of one sensory (AQR), two motor (RIML/R) and 22 inter-neurons. While RIML/R belong to

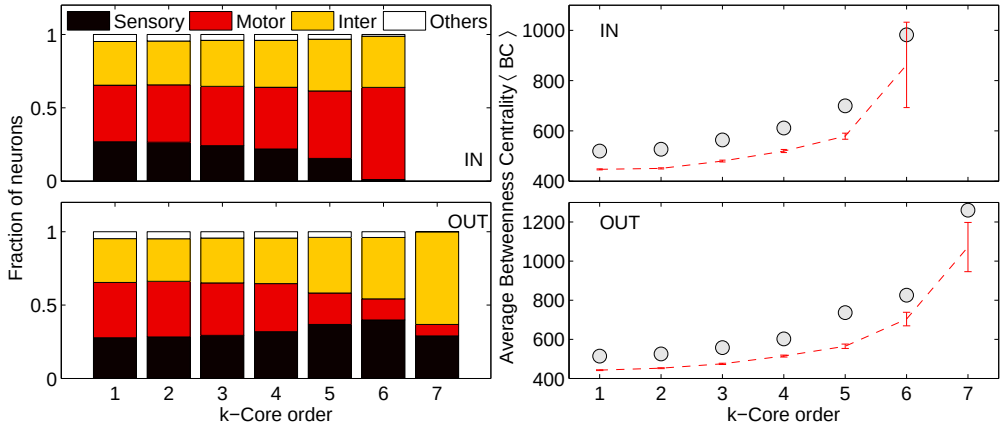


Figure 6.12: (A, B) The distribution of different neuronal types among the different cores for the incoming (A) and outgoing (B) synaptic connections. The fraction of motor neurons increases while that of sensory neurons decreases with increasing in-degree core order. Conversely, the out-degree inner cores exhibit high representation of sensory neurons and a corresponding low fraction of motor neurons. Inter-neurons are consistently well-represented in both in- and out-degree cores of higher order. (C, D) The average betweenness centrality of all neurons (filled circles) in the different cores of incoming (C) and outgoing (D) synaptic connections. For both types of cores, the increase of BC with core order k is significantly higher than that for the degree-conserved randomized version of the empirical network (shown using broken curve).

the functional circuit responsible for thermotaxis in the worm, the role of AQR is less clear, although it has been implicated in social versus solitary feeding in the worm [186]. Among the inter-neurons, 12 belong to the different functional circuits. Of the remaining neurons, DVA and AVKL has already been found to be connector hubs from our preceding analysis. Indeed, there is a significant overlap of the neurons occurring in the intersection of the innermost cores for in- and out-degree with those that play the role of connectors in the network (both hub and non-hub), with 84% of these neurons belonging to either the R3 or the R6 classes as defined previously. Thus, the different modules of the nematode nervous system, which have relatively low inter-connection between them, communicate mostly via neurons belonging to the innermost core.

The functional significance of the inner core neurons is also shown by the increasing betweenness centrality (BC) of neurons with core order (Fig. 6.12 C,D). This indicates that most of the shortest paths between pairs of neurons pass

through the neurons belonging to the innermost cores. As BC is related to degree, which is used for performing core-decomposition, this relation is not entirely unexpected. However, the random networks generated by preserving degree sequence show a consistently lower BC at any given core order compared to the empirical network, suggesting that the presence of neurons with high centrality is significant. This may have functional importance in terms of information processing, with all signals being channelled through a small group of "core" neurons instead of propagating through several distinct pathways between the sensory organs and muscles.

6.3.2 Hierarchical structure and noise filtering

It is important to consider why a nervous system as simple as that of C. elegans requires centralized processing, when, in principle, a set of semi-independent parallel pathways (e.g., reflex arcs) connecting dedicated sensory and motor neurons via specific interneurons, could have also been viable. To understand the advantages of the former model of information processing, we propose the hypothesis that it reduces the uncontrolled spreading of excitation globally through the network (as would have been the case for a non-hierarchical network) while having high efficiency for signal propagation between any pair of neurons. As C. elegans uses graded potential neurons, the transfer of activity is well described by a diffusive process. We therefore use diffusion to measure the extent of passive spreading of excitation in the network by considering the eigenvalues of the corresponding Laplacian matrix.

For a diffusion process, the mixing rate is defined as $\nu = \ln \mu^{-1}$, where

$$\mu = \max\{1 - \lambda_2, \lambda_N - 1\}, \tag{6.5}$$

and $\lambda_1 \ldots \lambda_N$ are the eigenvalues of the normalized laplacian matrix $\mathscr{L}$. The mixing rate is faster in networks where the eigenvalues of $\mathscr{L}$ are concentrated close to 1. A comparison of the spectral distribution for the Laplacian of the empirical network with that of the degree-conserved randomized ensemble (Fig. 6.13) shows that the smallest non-zero eigenvalue is three times larger in the latter. As this eigenvalue is related to the inverse of the dominant time-scale for diffusion, it sug-

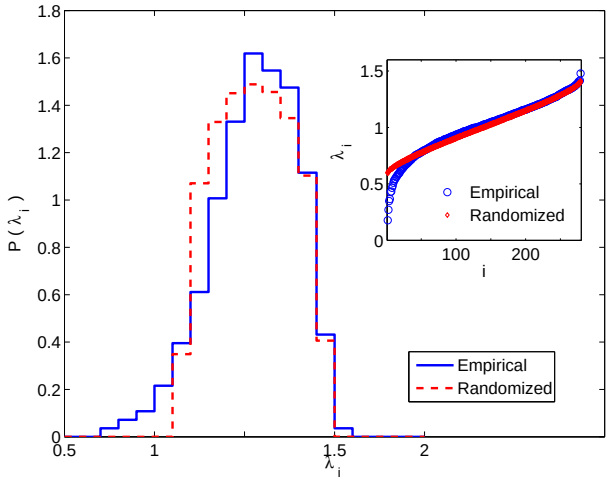


Figure 6.13: The distribution of eigenvalues for the normalized Laplacian $\mathcal{L}$, of the C. elegans neuronal network (solid) compared with that of degree-conserved randomized versions of the empirical network (broken), obtained by averaging over 10^3 realizations. (Inset) Rank-ordered eigenvalues of the empirical (circle) and degree-conserved randomized network (diamond). While, the eigenvalues in the bulk of the distribution are comparable, the smallest eigenvalues that are related to the dominant time-scale for diffusion over the system, are significantly lower for the empirical network, suggesting a slower rate of diffusion in the latter.

gests that excitation spreads in the empirical network much more slowly compared to the randomized network, although the communication efficiency of the two are comparable, the difference in the corresponding values of E being less than 10%. It is possible that the hierarchical core-periphery organization of network enables rapid motor response to a sensory stimulus, without resulting in global excitation of the network. The latter criterion is important because processing of information (i.e., "computation") in neuronal networks is believed to involve a highly controlled sequence of changes in the activation state of individual neurons [187].

To establish the above hypothesis, we propose a model network having a core-periphery hierarchical organization. Using this model we study the role of a centralized structure in reducing uncontrolled spreading of activation in the network without reducing the overall efficiency in responding to stimulus. In our model, N nodes are divided into l levels of equal size N/l. The nodes in the innermost (1st) level are densely connected among themselves with a connection density ρ_1. The successive $l-1$ levels are less densely connected, with the density of connections

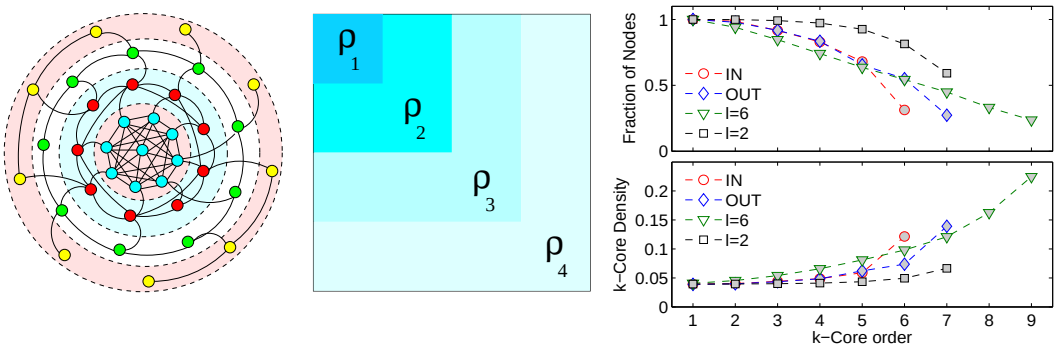


Figure 6.14: (Left) Schematic representation of the connectivity structure of the model, with (Right) the corresponding adjacency matrix indicating the non-uniform connection densities for the different levels ($l = 1, \ldots, 4$). Darkest shade represents the highest connection density. (C, D) Core-decomposition of the model reveals its core-periphery organization with (C) fraction of nodes in each core of order k and the (D) connection density at each core, increasing with k. The network models having $l = 2$, 5 and 6 levels show qualitatively similar variation with k as compared to the synaptic in-degree and out-degree cores of the empirical network. The curves shown for the model are averages over 100 realizations for each value of l with $N = 240$ and $q = 1.9$. The $\langle k \rangle = 9.36$ of the model is chosen so that the connectivity of the model and the empirical network are same.Schematic diagram of a network model with dense inner core structure. (Right) The adjacency matrix of the model network, with the color of each region representing the connection density there. Darker shades represent higher connection densities.

between nodes in the level i, $\rho_l < \cdots < \rho_i \cdots < \rho_1$ (Fig. 6.14 A,B). For simplicity, we consider the ratio of densities between subsequent levels to be constant: $\frac{\rho_1}{\rho_2} = \frac{\rho_2}{\rho_3} = \cdots = \frac{\rho_{l-1}}{\rho_l} = q$. Thus, by varying q, the density of connections in the different levels can be changed. When $q = 1$, the system is a homogeneous random network, while for $q > 1$ it has a core-periphery structure with an inter-nested arrangement of densely connected layers.

By construction, the inter-connection density is highest among the nodes in the "central kernel" of the network, and gradually decreases outward. The model network has a core-periphery organization similar to that of the nematode network, as observed by performing a k-core decomposition and comparing the connection densities between the two systems for each core order. Fig. 6.14 (C,D) shows the model network for two different values of l (number of levels) at a given value of q (the ratio of connection densities between subsequent levels). For small l, the network is relatively undifferentiated. With increasing l, the hierarchical structure

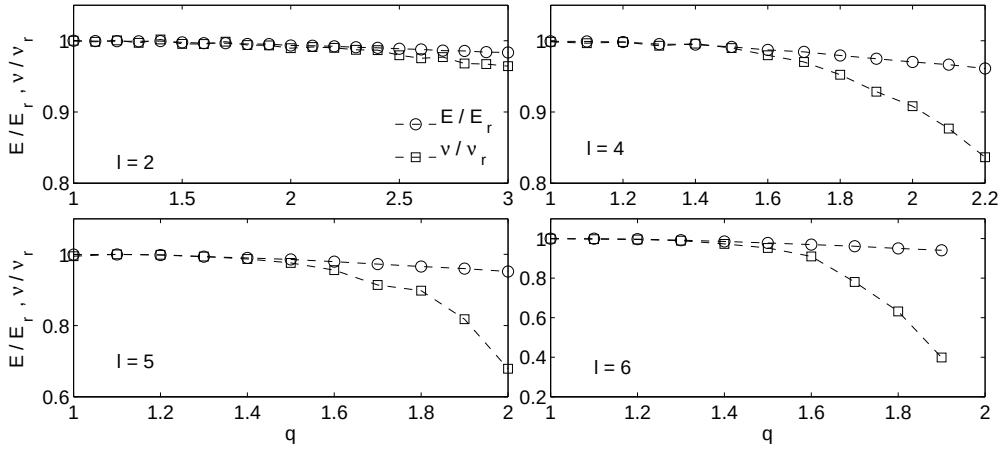


Figure 6.15: The relative efficiency E and mixing rate ν of the core-periphery model having different number of levels (A: $l = 2$, B: $l = 4$, C: $l = 5$ and D: $l = 6$) shown as a function of the relative density between successive levels, q. For larger number of levels, the mixing rate decreases sharply with q, while the relative efficiency remains almost unchanged. In all cases $N = 240$, $\langle k \rangle = 20$. Note the very different value of y-axis used in figures (A-D). It indicates that as the core structure of the network model is increase the relative change in ν is more than the change in E.

becomes prominent, as seen in the increasing connection density for successively higher order cores.

We now consider how core-periphery hierarchical organization affects diffusive spreading and communication efficiency in the system. Fig. 6.15 shows the variation of the scaled efficiency E/E_r, as well as, the scaled mixing rate ν/ν_r, both measurables being scaled with respect to randomized networks, as a function of q for model networks with different number of levels l. For large l, there is a rapid increase in the time required to diffuse across the entire network with increasing q, although the communication efficiency remains almost constant. This implies that, a network having a prominent core (i.e., large q) and a hierarchical structure with many levels (large l) is effective in reducing uncontrolled spreading of excitation through the system without sacrificing the rapid communication speed between nodes involved in information processing. It supports our hypothesis that the hierarchical core-periphery organization of the C. elegans nervous system is responsible for the significant difference in the diffusive time-scales for the empirical and randomized networks.

6.4 Discussion

In this chapter, we have carried out a detailed analysis of the mesoscopic structure in the connection topology of the C. elegans neuronal network. Inferring the organizing principles underlying the network may give us an understanding of the way in which an organism makes sense of the external world. We have focused primarily on the existence of modules, i.e., groups of neurons having higher connection density among themselves than with neurons in other groups. Presence of such mesoscopic organization naturally prompts us to ask the reasons behind the evolution of these features in the network.

There have been recent attempts at explaining neuronal position and structural layout of the network by using static constraints, such as wiring economy and communication path minimization. Although we find that membership of neurons in specific modules are correlated with their physical nearness, the empirical network is sub-optimal in terms of both the above-mentioned constraints. By comparing the system with other complex networks that have been either designed or have evolved for rapid transportation while being subject to wiring economy, we find that the C. elegans nervous system stands apart as a distinct class. This suggests that the principal function of neuronal networks, viz., the processing of information, distinguishes it from the other networks considered, and plays a vital role in governing its arrangement. Considering the importance of this constraint in ensuring the survival of an organism, it is natural that this should be key to the organizing principles underlying the design of the network. The intimate relation between function and structure of the nervous system is further brought out by our use of structural analysis to distinguish neurons that are critical for the survival of the organism. In addition to identifying neurons that have been already empirically implicated in different functions (which serve as a verification of our method), we also predict several neurons which can be potentially crucial for certain, as yet unidentified, functions.

When we compare the nervous system of C. elegans with the brains of higher organisms, we observe the modular organization of the latter to be more prominent [188]. For example, the network of cortical areas in the cat and macaque brains exhibit distinct modules [104, 56], with each module being identified with specific functions [189, 63]. A possible reason for the relatively weak modular struc-

ture in the nematode could be due to the existence of extended processes for the neurons of C. elegans. Many of these span almost the entire body length, an effect that is enhanced by the approximately linear nature of the nematode body plan. As a result, connections are not constrained by the physical distance between soma of the neurons, as would be the case in mammalian brains. It is apparent that such constraints on the geographical distance spanned by links between nodes (viz., cost of wiring length) can give rise to clustering of connections among physically adjacent elements. In addition, the small nervous system of C. elegans, comprising only 302 neurons, lacks redundancy. Therefore, individual neurons may often have to perform a set of tasks which in higher organisms are performed by several different neurons. Thus, functional modularity is less prominent in the nematode, as some neurons belong to multiple behavioral circuits.

Another principal distinction between the C. elegans nervous system and the brains of higher organisms such as human beings, is the relative high connectivity in the former (the connection density being $C \sim 0.1$). By contrast, the connectance for human brain is around 10^{-6} [190], which leads us to the question of how communication efficiency can remain high in such a sparsely connected network. It is possible that the more intricate hierarchical and modular structures seen in the brains of higher organisms is a response to the above problem. The fact that the rate at which the number of neurons N increase across species, is not matched by a corresponding increase in the number of links (which increases slower than N^2) implies the existence of constraints on the latter, which is a resource cost in addition to the earlier mentioned cost of wiring length.

The function of information processing implies an underlying hiererachy, that imposes a direction to the flow of signals in the system, from the input to the output. In real networks, this hierarchical scheme can be obscured by the presence of feedback and connections that span several levels. In particular, for C. elegans, we find that it is difficult to distinguish its somatic nervous system from an unstructured network by analysing the properties of directed paths between the sensory, inter and motor neurons. However, the inherent structure of the network is manifested as a sequence of nested cores of successively higher inter-connection densities. Our results showing increasing centrality for higher order cores, as well as, the varying composition of the in- and out-degree cores in terms of different neuronal types, reveal the hierarchical nature of the system which is implicit in

the core-periphery organization.

Many complex networks occurring in nature are often referred to as having hierarchical structure, although, there are no generally accepted measure of the degree of hierarchy present in a system. This is partly because hierarchy can have different connotations. E.g., while some studies consider the presence of inter-nested modules to be the signature of hierarchical organization [107, 191, 192, 188, 59], other papers have considered the presence of distinct layers in a network to be essential for the existence of hierarchy [193, 54, 194]. In the present paper, in the spirit of neuronal information processing, we adopt the usage of hierarchy in the sense of multiple levels of processing, with information flowing from the sensory (input) to the motor (output) layer through inter-neurons.

We propose that a hierarchical network, which has a dense core and an overall sparse structure, possesses high communication efficiency, and at the same time, relatively low diffusion rate, implying that neuronal networks have evolved subject to the constraint that a stimulation should not result in non-specific global activation of the network. This requires that while communication between a given pair of neurons should be fast, enabling a rapid response to sensory stimulus, it should also be robust to environmental and internal noise, so that perturbations can remain localized in the system, preventing indiscriminate activation of the entire network from a specific stimulus. The above analysis suggests that information does not diffuse passively throughout the network, but is guided towards a central group of densely interconnected neurons, where it is processed and appropriate commands are sent to motor-neurons for initiating muscular action.

7 Conclusions

Physicists look for universal principles that are valid across many different systems, often spanning several length or time scales. While the domain of physical systems has often offered examples of such widely applicable 'laws', social, economic and biological phenomena tended to be, until quite recently, less fertile in terms of generating similar universalities. However, this situation has changed after the study of complex networks emerged into prominence. While the existence of complex networks in various domains has been known for some time, the recent excitement among physicists has to do with the discovery of certain key universal principles governing the behavior of systems which had previously been considered very different from each other.

7.1 Structure of networks

Complex networks exhibit a variety of structural features. One of the most intriguing properties of many networks observed in nature and society is the occurrence of modular structure. Such structural modularity may also be connected to functional modules, which are independent subsystems responsible for different tasks, as seen in many biological systems. Another ubiquitous property of complex networks is the occurrence of "small-world" behavior along with high clustering between the nodes. Both modularity and clustered small-world properties have been observed in a wide range of networks, from those involved in metabolism and signalling in biological cells, to the set of cortico-cortical connections in the brain, interactions in social groups, the internet and food webs. In this thesis, we have shown that

the above properties, previously considered to be independent, are in fact related to each other. This shows how the mesoscopic organization of a network can have unexpected consequences for its global features.

7.2 Dynamics on networks

Networks often have associated dynamics, with variables associated with each node evolving over time. Examples include, the variation in the populations of different species in an ecological network, the changing metabolite concentrations in cellular networks, etc. Thus, focus of recent work in the area of complex networks has shifted from purely structural aspects of the connection topology to investigating their role in determining the dynamical behavior of the network. It has been found that specific structural properties of networks can have novel functional consequences, e.g., the absence of threshold for the propagation of epidemics in scale-free networks. In this thesis, we have not only looked at how network structure affects dynamics, and hence its function, but also the reverse problem of how functional criteria can constrain the topological properties of a network. In particular, we have investigated (i) how mesoscopic features, such as modular and hierarchical organization, affect the nature of dynamics on the network, and (ii) how dynamical considerations constrain the network structure, such that it evolves towards a modular or a hierarchically ordered configuration.

7.3 From structure to dynamics

We have shown that the dynamical behavior of modular networks are strikingly different from previously proposed small world network (SWN) models, such as, the Watts-Strogatz model. Modularity results in time-scale separation between fast intra-modular and slow inter-modular processes. As dynamics at the local and global levels have different consequences in most natural systems, the temporal separation between processes occurring at different scales, through modular organization, highlights the importance of such network structures. We have demonstrated the universality of this dynamical signature of modular networks with three very different processes: (i) spin ordering, (ii) synchronization among non-linear

oscillators, and (iii) diffusion. We have also shown this dynamical signature of modular networks in the cortico-cortical networks for cat and macaque. This suggests that many of the reported SWNs in nature are possibly better represented by a modular network model. We have also investigated the connection between structural features and functional objectives of the somatic nervous system for the nematode *C. elegans*. According to our results, the anatomical structure of the neuronal network can be only partly explained by static criteria, such as, wiring cost minimization and maximum communication efficiency. This indicates the existence of other important constraints, possibly related to the functional task of information processing, that determine the wiring diagram of the nervous system. We have shown that the network can be decomposed into modules, which can be partially correlated with functional circuits. Further, the nervous system has a hierarchical core-periphery organization, with inner cores having higher density of connections. We have shown that such an architecture reduces the uncontrolled spreading of activity in the network, thus acting as a noise filter, while retaining high communication efficiency.

7.4 From dynamics to structure

We have also considered how the existence of various structural features in real-world complex systems can be uncovered from the knowledge of collective dynamics of the nodes. We have reconstructed the network of social interactions in a troop of bonnet macaques by analyzing their allogrooming behavior. This allows us to obtain an understanding of social organization in primates using observational data, the equivalent of which would be very difficult to obtain for human groups. We have also examined whether the complex features of social behavior can be explained in terms of basic principles governing individual interactions, using a simple theoretical model proposed by Seyfarth. Next, we have investigated the network structure of interactions between stocks in financial markets. By analyzing the cross-correlation matrix of price fluctuations among stocks in the National Stock Exchange (NSE) of India, we show that this *emerging* market exhibits strong correlations compared to *developed* markets, such as the New York Stock Exchange (NYSE). We have shown this to be due to the dominant influence of a common

market mode on the different stock prices. In comparison, interactions between related stocks, e.g., those belonging to the same business sector, are much weaker. This lack of distinct sector identity in emerging markets is explicitly shown by reconstructing the network of mutually interacting stocks. We have shown this to be a result of the relative weakness of intra-sector interactions, compared to the response to signals common to the entire market, by modeling stock price dynamics with a two-factor model. Our results suggest that the emergence of a complex internal structure, comprising multiple groups of strongly coupled components, is a signature of market development.

7.5 Evolution of robust networks

Most networks around us did not originate in the form we see them today, but have emerged through a process of gradual evolution. We have shown using models the evolution of some of the commonly observed structural features in naturally occurring networks. For this, we have taken into account the fact that, most such systems have to optimize between several, often conflicting, constraints, which may be static, as well as, dynamical in nature. In particular, most networks need to have high communication efficiency and low connectivity, while being stable with respect to dynamical perturbations in the nodes. Our results show that, the simultaneous optimization of all three constraints can result in networks with modular structure, where each module possesses a prominent hub. As these evolved systems also exhibit heterogeneous degree distribution, our findings have implications for a wide range of systems in the biological and technological domains where similar features have been observed.

Bibliography

[1] Y. Bar-Yam, *Dynamics of Complex Systems.* Addison-Wesley, Reading, MA (1997).

[2] N. Boccara, *Modeling complex systems.* Springer, Heidelberg (2004).

[3] R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).

[4] M. E. J. Newman, *SIAM Review* **45**, 167 (2003).

[5] S. N. Dorogovtsev, J. Mendes, *Evolution of networks: From biological nets to the Internet and WWW.* Oxford University Press, Oxford (2003).

[6] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, *Phys. Rep.* **424**(4-5), 175 (2006).

[7] M. Boots, A. Sasaki, *Proc. R. Soc. B* **266**, 1933 (1999).

[8] R. Pastor-Satorras, A. Vespignani, *Phys. Rev. Lett.* **86**(14), 3200 (2001).

[9] R. Albert, H. Jeong, A.-L. Barabási, *Nature* **406**, 378 (2000).

[10] T. Araújo, R. Vilela Mendes, *Complex Systems* **12**, 357 (2000).

[11] J. P. Onnela, et al., *Proc. Natl. Acad. Sci. U.S.A.* **104**(18), 7332 (2007).

[12] S. H. Strogatz, *Nature* **410**, 268 (2001).

[13] R. Mendes, *International Journal of Bifurcation and Chaos* **15**(4), 1185 (2005).

[14] N. Dokholyan, L. Li, F. Ding, E. Shakhnovich, *Proc. Natl. Acad. Sci. U.S.A.* **99**(13), 8637 (2002).

[15] M. Marinissen, J. Gutkind, *Trends in pharmacological sciences* **22**(7), 368 (2001).

[16] B. Schwikowski, P. Uetz, S. Fields, *Nature Biotechnology* **18**, 1257 (2000).

[17] A. W. Rives, T. Galitski, *Proc. Natl. Acad. Sci. U.S.A.* **100**(3), 1128 (2003).

[18] J. White, E. Southgate, J. Thomson, S. Brenner, *Phil. Trans. R. Soc. B* **314**, 1 (1986).

[19] A. E. Krause, K. A. Frank, D. M. Mason, R. U. Ulanowicz, W. W. Taylor, *Nature* **426**, 282 (2003).

[20] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, E. Shir, *Proc. Natl. Acad. Sci. U.S.A.* **104**(27), 11150 (2007).

[21] I. Yoon, et al., *Proceedings of the IS&T/SPIE Symposium on Electronic Imaging, Visualization and Data Analysis* **5295**, 124 (2004).

[22] E. Almaas, B. Kovacs, T. Vicsek, Z. N. Oltvai, A.-L. Barabasi, *Nature* **427**, 839 (2004).

[23] S. Wasserman, K. Faust, *Social Network Analysis.* Cambridge University Press, Cambridge (1994).

[24] C. Lampe, N. Ellison, C. Steinfield, in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 435–444. ACM New York, NY, USA (2007).

[25] M. E. J. Newman, *Phys. Rev. E* **64**(1), 016131 (2001).

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

[27] R. Albert, I. Albert, G. Nakarado, *Phys. Rev. E* **69**(2), 025103 (2004).

[28] M. T. Gastner, M. E. J. Newman, *Eur. Phys. J. B* **49**(2), 247 (2006).

[29] P. Sen, et al., *Phys. Rev. E* **67**(3), 036106 (2003).

[30] A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani, *Proc. Natl. Acad. Sci. U.S.A.* **101**(11), 3747 (2004).

[31] G. D. Masi, G. Iori, G. Caldarelli, *Phys. Rev. E* **74**(6), 066112 (2006).

[32] J. P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, *Eur. Phys. J. B* **30**(3), 285 (2002).

[33] R. K. Pan, S. Sinha, *Phys. Rev. E* **76**(4), 046116 (2007).

[34] A.-L. Barabási, R. Albert, *Science* **286**, 509 (1999).

[35] L. A. N. Amaral, A. Scala, M. Barthelemy, H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **97**(21), 11149 (2000).

[36] V. Latora, M. Marchiori, *Phys. Rev. Lett.* **87**(19), 198701 (2001).

[37] D. J. Watts, S. H. Strogatz, *Nature* **393**(6684), 440 (1998).

[38] S. Dorogovtsev, J. Mendes, *cond-mat/0404593* (2005).

[39] A. Arenas, A. Diaz-Guilera, J. Kurths, Y. Moreno, C. Zhou, *Phys. Rep.* **469**, 93 (2008).

[40] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, *Rev. Mod. Phys.* **80**(4), 1275 (2008).

[41] P. Erdös, A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci., Ser. Au* **5**, 17 (1960).

[42] B. Bollobas, *Random Graphs*. Cambridge University Press, Cambridge (2001).

[43] W. Aiello, F. Chung, L. Lu, in *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pp. 171–180, New York, NY, USA (2000). ACM Press.

[44] A. Barrat, M. Weigt, *Eur. Phys. J. B* **13**(3), 547 (2000).

[45] G. Caldarelli, A. Capocci, P. De Los Rios, M. A. Muñoz, *Phys. Rev. Lett.* **89**(25), 258702 (2002).

[46] D. J. S. Price, *Journal of the American Society for Information Science* **27**, 292 (1976).

[47] P. L. Krapivsky, S. Redner, F. Leyvraz, *Phys. Rev. Lett.* **85**(21), 4629 (2000).

[48] R. Cohen, S. Havlin, *Phys. Rev. Lett.* **90**(5), 058701 (2003).

[49] B. Bollobas, O. Riordan, *Combinatorica* **24**(1), 5 (2004).

[50] A.-L. Barabási, Z. N. Oltvai, *Nature Reviews Genetics* **5**(2), 101 (2004).

[51] R. Milo, et al., *Science* **298**(5594), 824 (2002).

[52] M. Girvan, M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821 (2002).

[53] A. Arenas, A. Diaz-Guilera, C. J. Perez-Vicente, *Phys. Rev. Lett.* **96**(11), 114102 (2006).

[54] A. Trusina, S. Maslov, P. Minnhagen, K. Sneppen, *Phys. Rev. Lett.* **92**(17), 178702 (2004).

[55] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Barabási, *Science* **297**, 1551 (2002).

[56] R. K. Pan, S. Sinha, *Europhys. Lett.* **85**, 68006 (2009).

[57] S. Dasgupta, R. K. Pan, S. Sinha, *Phys. Rev. E* (2009), arXiv:0903.4529.

[58] R. K. Pan, S. Sinha, *Phys. Rev. E* **76**(4), 045103 (2007).

[59] R. K. Pan, S. Sinha, *Pramana - journal of physics* **31**, 331 (2008).

[60] R. K. Pan, N. Chatterjee, S. Sinha, *ArXiv:0905.3887* (2009).

[61] C. M. Gray, P. Konig, A. K. Engel, W. Singer, *Nature* **338**, 334 (1989).

[62] E. R. Kandel, J. H. Schwartz, T. M. Jessell, *Principles of Neural Science*. McGraw-Hill, New York, 4 edn. (2000).

[63] D. S. Bassett, E. Bullmore, *Neuroscientist* **12**(6), 512 (2006).

[64] M. E. J. Newman, D. J. Watts, S. H. Strogatz, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2566 (2002).

[65] D. A. Fell, A. Wagner, *Nature Biotechnology* **18**(11), 1121 (2000).

[66] M. E. J. Newman, *J. Stat. Phys.* **101**(3-4), 819 (2000).

[67] L. H. Hartwell, J. J. Hopfield, S. Leibler, A. W. Murray, *Nature* **402**, C47 (1999).

[68] R. Guimera, L. A. N. Amaral, *Nature* **433**, 895 (2005).

[69] P. Holme, M. Huss, H. Jeong, *Bioinformatics* **19**, 532 (2003).

[70] A. Arenas, L. Danon, A. Diaz-Guilera, P. M. Gleiser, R. Guimera, *Eur. Phys. J. B* **38**(2), 373 (2004).

[71] K. A. Eriksen, I. Simonsen, S. Maslov, K. Sneppen, *Phys. Rev. Lett.* **90**, 148701 (2003).

[72] M. E. J. Newman, M. Girvan, *Phys. Rev. E* **69**(2), 026113 (2004).

[73] C. P. Massen, J. P. K. Doye, *arXiv:cond-mat/0610077v1* (2006).

[74] S. Fortunato, M. Barthelemy, *Proc. Natl. Acad. Sci. U.S.A.* **104**(1), 36 (2007).

[75] I. Derenyi, G. Palla, T. Vicsek, *Phys. Rev. Lett.* **94**(16), 160202 (2005).

[76] G. Palla, I. Derenyi, I. Farkas, T. Vicsek, *Nature* **435**, 814 (2005).

[77] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, D. Parisi, *Proc. Natl. Acad. Sci. U.S.A.* **101**(9), 2658 (2004).

[78] T. W. Valente, *Network Models of the Diffusion of Innovations*. Hampton Press, Cresskill, NJ (1995).

[79] X. Castello, et al., *Europhys. Lett.* **79**(6), 66006 (2007).

[80] M. Granovetter, *American Journal of Sociology* **78**, 1360 (1973).

[81] A. Baronchelli, V. Loreto, *Phys. Rev. E* **73**(2), 026103 (2006).

[82] S. Maslov, K. Sneppen, A. Zaliznyak, *Physica A* **333**, 529 (2003).

[83] J. W. Scannell, C. Blakemore, M. P. Young, *Journal of Neuroscience* **15**(2), 1463 (1995).

[84] C. J. Honey, R. Kotter, M. Breakspear, O. Sporns, *Proc. Natl. Acad. Sci. U.S.A.* **104**(24), 10240 (2007).

[85] R. K. Pan, S. Sinha, *Europhys. Lett.* **77**, 58004 (2007).

[86] V. Colizza, A. Barrat, M. Barthelemy, A. Vespignani, *BMC Medicine* **5**, 34 (2007).

[87] E. A. Variano, J. H. McCoy, H. Lipson, *Phys. Rev. Lett.* **92**, 188701 (2004).

[88] N. Kashtan, U. Alon, *Proc. Natl. Acad. Sci. U.S.A.* **27**(39), 13773 (2005).

[89] R. V. Sole, P. Fernandez, *q-bio.GN/0312032* (2003).

[90] P. Holme, M. E. J. Newman, *Phys. Rev. E* **74**(5), 056108 (2006).

[91] R. Guimerà, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, A. Arenas, *Phys. Rev. Lett.* **89**(24), 248701 (2002).

[92] N. Mathias, V. Gopal, *Phys. Rev. E* **63**, 021117 (2001).

[93] R. M. May, *Stability and complexity in model ecosystems*. Princeton University Press, Princeton, NJ (1973).

[94] V. K. Jirsa, M. Ding, *Phys. Rev. Lett.* **93**, 070602 (2004).

[95] S. Sinha, *Physica A* **346**, 147 (2005).

[96] S. Sinha, S. Sinha, *Phys. Rev. E* **71**, 020902(R) (2005).

[97] Y. Wang, D. Chakrabarti, C. Wang, C. Faloutsos, *Proceedings of the symposium on reliable distributed systems* pp. 25–34 (2003).

[98] M. E. J. Newman, *Eur. Phys. J. B* **38**(2), 321 (2004).

[99] J. Duch, A. Arenas, *Phys. Rev. E* **72**(2), 027104 (2005).

[100] B. Wang, H. Tang, C. Guo, Z. Xiu, *Physica A* **363**(2), 591 (2006).

[101] H. M. Hastings, F. Juhasz, M. A. Schreiber, *Proc. R. Soc. B* **249**(1326), 223 (1992).

[102] T. Tanizawa, G. Paul, R. Cohen, S. Havlin, H. E. Stanley, *Phys. Rev. E* **71**, 047101 (2005).

[103] R. V. Solé, A. Munteanu, *Europhys. Lett.* **68**, 170 (2004).

[104] C. Zhou, L. Zemanová, G. Zamora, C. C. Hilgetag, J. Kurths, *Phys. Rev. Lett.* **97**(23), 238103 (2006).

[105] M. Barahona, L. M. Pecora, *Phys. Rev. Lett.* **89**(5), 054101 (2002).

[106] C. S. Elton, *The Ecology of Invasions by Animals and Plants*. Methuen, London (1958).

[107] E. Ravasz, A.-L. Barabási, *Phys. Rev. E* **67**, 026112 (2003).

[108] S. Yook, Z. Oltvai, A. Barabási, *Proteomics* **4**(4), 928 (2004).

[109] S. N. Soffer, A. Vázquez, *Phys. Rev. E* **71**(5), 057101 (2005).

[110] S. Sinha, S. Sinha, *Phys. Rev. E* **74**(6), 066117 (2006).

[111] L. M. Pecora, T. L. Carroll, *Phys. Rev. Lett.* **80**(10), 2109 (1998).

[112] A. Sinha, K. Mukhopadhyay, A. Datta-Roy, S. Ram, *Current Science* **89**(7), 1166 (2005).

[113] R. N. Mantegna, H. E. Stanley, *Introduction to Econophysics*. Cambridge University Press, Cambridge, UK (1999).

[114] J. P. Bouchaud, M. Potters, *Theory of Financial Risk and Derivative Pricing*. Cambridge University Press, Cambridge, UK, 2 edn. (2003).

[115] I. Kondor, J. Kertesz (eds.), *Econophysics: An Emerging Science*. Kluwer, Dordrecht (1999).

[116] A. Chatterjee, B. K. Chakrabarti (eds.), *Econophysics of Stock and other Markets*. Springer, Milan (2006).

[117] T. Lux, *Applied Financial Economics* **6**(6), 463 (1996).

[118] V. Plerou, P. Gopikrishnan, L. A. Nunes Amaral, M. Meyer, H. E. Stanley, *Phys. Rev. E* **60**(6), 6519 (1999).

[119] L. Laloux, P. Cizeau, J. P. Bouchaud, M. Potters, *Phys. Rev. Lett.* **83**(7), 1467 (1999).

[120] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amaral, H. E. Stanley, *Phys. Rev. Lett.* **83**, 1471 (1999).

[121] V. Plerou, et al., *Phys. Rev. E* **65**, 066126 (2002).

[122] A. Utsugi, K. Ino, M. Oshikawa, *Phys. Rev. E* **70**, 026110 (2004).

[123] P. Gopikrishnan, B. Rosenow, V. Plerou, H. E. Stanley, *Phys. Rev. E* **64**, 035106(R) (2001).

[124] D. H. Kim, H. Jeong, *Phys. Rev. E* **72**, 046133 (2005).

[125] L. Giada, M. Marsili, *Phys. Rev. E* **63**(6), 061101 (2001).

[126] R. Morck, B. Yeung, W. Yu, *Journal of Financial Economics* **58**, 215 (2000).

[127] D. Wilcox, T. Gebbie, *Physica A* **375**(2), 584 (2007).

[128] V. Kulkarni, N. Deo, *Econophysics of Stock and Other Markets*, chap. A random matrix approach to volatility in an Indian financial market, pp. 35–48. Springer, Milan (2006).

[129] S. Cukur, M. Eryigit, R. Eryigit, *Physica A* **376**, 555 (2007).

[130] http://www.nseindia.com/.

[131] http://finance.yahoo.com/.

[132] A. M. Sengupta, P. P. Mitra, *Phys. Rev. E* **60**, 3389 (1999).

[133] F. Lillo, R. N. Mantegna, *Phys. Rev. E* **72**, 016219 (2005).

[134] R. K. Pan, S. Sinha, *Physica A* **387**(8-9), 2055 (2008).

[135] L. Bachelier, *Annales Scientifiques de l'École Normale Supérieure Sér* **3**(17), 21 (1900).

[136] G. Tononi, O. Sporns, G. Edelman, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5033 (1994).

[137] D. L. Riddle, T. Blumenthal, B. J. Meyer, J. R. Priess (eds.), *Celegans II*. Cold Spring Harbor Laboratory Press, New York (1993).

[138] M. Reigl, U. Alon, D. B. Chklovskii, *BMC Biology* **2**(1), 25 (2004).

[139] E. Bullmore, O. Sporns, *Nature Reviews Neuroscience* **10**, 186 (2009).

[140] L. Ferrarini, et al., *Human Brain Mapping* **doi:10.1002/hbm.20663** (2008).

[141] A. Schwarz, A. Gozzi, A. Bifone, *Magnetic Resonance Imaging* **26**(7), 914 (2008).

[142] M. Young, *Proc. R. Soc. B* **252**(1333), 13 (1993).

[143] M. D. Humphries, K. Gurney, T. J. Prescott, *Proc. R. Soc. B* **273**(1585), 503 (2006).

[144] B. L. Chen, D. H. Hall, D. B. Chklovskii, *Proc. Natl. Acad. Sci. U.S.A.* **103**(12), 4723 (2006).

[145] U. Deppe, et al., *Proc. Natl. Acad. Sci. U.S.A.* **75**(1), 376 (1978).

[146] A. Arenas, A. Fernandez, S. Gomez, *Lecture Notes In Computer Science* **5151**, 9 (2008).

[147] A. Perez-Escudero, G. G. de Polavieja, *Proc. Natl. Acad. Sci. U.S.A.* **104**(43), 17180 (2007).

[148] S. Itzkovitz, L. Baruch, E. Shapiro, E. Segal, *Proc. Natl. Acad. Sci. U.S.A.* **105**(27), 9278 (2008).

[149] M. Kaiser, C. C. Hilgetag, *PLoS Computational Biology* **2**(7), e95 (2006).

[150] Y.-Y. Ahn, H. Jeong, B. J. Kim, *Physica A* **367**, 531 (2006).

[151] N. Brenner, W. Bialek, R. de Ruyter van Steveninck, *Neuron* **26**(3), 695 (2000).

[152] R. F. Cancho, C. Janssen, R. V. Sole, *Phys. Rev. E* **64**(4), 046119 (2001).

[153] R. Milo, et al., *Science* **303**(5663), 1538 (2004).

[154] K. Klemm, S. Bornholdt, *Proc. Natl. Acad. Sci. U.S.A.* **102**(51), 18414 (2005).

[155] S. Maslov, K. Sneppen, *Science* **296**, 910 (2002).

[156] R. Guimera, M. Sales-Pardo, L. A. Amaral, *Nature Physics* **3**(1), 63 (2007).

[157] E. A. Leicht, M. E. J. Newman, *Phys. Rev. Lett.* **100**(11), 118703 (2008).

[158] M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8577 (2006).

[159] O. Hobert, *Specification of the nervous system.* WormBook, ed. The C. elegans Research Community (2005).

[160] M. Chalfie, et al., *J Neurosci.* **5**(4), 956 (1985).

[161] S. Wicks, C. Rankin, *J Comp Physiol [A]* **179**(5), 675 (1996).

[162] E. R. Sawin, *Genetic and cellular analysis of modulated behaviors in Caenorhabditis elegans*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts (1996), page-316.

[163] L. E. Waggoner, G. T. Zhou, R. W. Schafer, W. R. Schafer, *Neuron* **21**(1), 203 (1998).

[164] I. A. Bany, M. Q. Dong, M. R. Koelle, *J Neurosci* **23**(22), 8060 (2003).

[165] I. Mori, Y. Ohshima, *Nature* **376**(6538), 344 (1995).

[166] E. R. Troemel, B. E. Kimmel, C. I. Bargmann, *Cell* **91**(2), 161 (1997).

[167] J. M. Gray, J. J. Hill, C. I. Bargmann, *Proc. Natl. Acad. Sci. U.S.A.* **102**(9), 3184 (2005).

[168] S. R. Wicks, C. H. Rankin, *Journal of Neuroscience* **15**(3), 2434 (1995).

[169] C. J. Franks, L. Holden-Dye, K. Bull, S. Luedtke, R. J. Walker, *Invert Neurosci.* **6**(3), 105 (2006).

[170] B. Shtonda, L. Avery, *J Exp Biol.* **209**(1), 89 (2006).

[171] S. Daniels, M. Ailion, J. H. Thomas, P. Sengupta, *Genetics* **156**(1), 123 (2008).

[172] A. J. Bretscher, K. E. Busch, M. de Bono, *Proc. Natl. Acad. Sci. U.S.A.* **105**(23), 8044 (2008).

[173] E. R. Sawin, R. Ranganathan, H. R. Horvitz, *Neuron* **26**(3), 619 (2000).

[174] S. L. McIntire, E. Jorgensen, J. Kaplan, H. R. Horvitz, *Nature* **364**(6435), 337 (1993).

[175] N. Chatterjee, S. Sinha, *Progress in Brain Research* **168**, 145 (2007).

[176] A. Broder, et al., *Computer Networks* **33**(1-6), 309 (2000).

[177] L. C. Freeman, *Sociometry* **40**(1), 35 (1977).

[178] A. Vazquez, R. Pastor-Satorras, A. Vespignani, *Phys. Rev. E* **65**(6), 066130 (2002).

[179] K.-I. Goh, E. Oh, B. Kahng, D. Kim, *Phys. Rev. E* **67**(1), 017101 (2003).

[180] R. Guimera, S. Mossa, A. Turtschi, L. Amaral, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7794 (2005).

[181] D. Van Essen, J. Maunsell, *Trends in Neurosciences* **6**(9), 370 (1983).

[182] D. Hubel, T. Wiesel, *The Journal of Physiology* **160**(1), 106 (1962).

[183] D. Hubel, T. Wiesel, *Journal of Neurophysiology* **28**(2), 229 (1965).

[184] J. Nicholls, A. Martin, B. Wallace, P. Fuchs, *From neuron to brain.* Sinauer Associates Inc, Massachusettes, 3rd edn. (1992).

[185] O. Hobert, *Journal of Neurobiology* **54**(1), 203 (2003).

[186] J. C. Coates, M. de Bono, *Nature* **419**, 925 (2002).

[187] J. Hertz, A. Krogh, R. G. Palmer, *Introduction to the Theory of Neural Computation.* Addison Wesley Publishing Company, Reading, MA (1991).

[188] M. Muller-Linow, C. Hilgetag, M. Hutt, *PLoS Computational Biology* **4**(9), e1000190 (2008).

[189] C. C. Hilgetag, G. A. P. C. Burns, M. A. O'Neill, J. W. Scannell, M. P. Young, *Phil. Trans. R. Soc. B* **355**, 91 (2000).

[190] G. M. Shepherd (ed.), *The Synaptic Organization of the Brain.* Oxford University Press, New York, 4 edn. (1998).

[191] O. Sporns, *Biosystems* **85**(1), 55 (2006).

[192] M. Sales-Pardo, R. Guimera, A. A. Moreira, L. A. N. Amaral, *Proc. Natl. Acad. Sci. U.S.A.* **104**(39), 15224 (2007).

[193] P. Dodds, D. Watts, C. Sabel, *Proc. Natl. Acad. Sci. U.S.A.* **100**(21), 12516 (2003).

[194] P. Holme, J. Karlin, S. Forrest, *Proc. R. Soc. A* **463**(2081), 1231 (2007).