Uncovering Functional Correlates of Structural Organization in Brain Networks at Multiple Scales : From the worm to the human

By

Anand Pathak

PHYS10201205010

The Institute of Mathematical Sciences, Chennai

A thesis submitted to the

Board of Studies in Physical Sciences

In partial fulfillment of requirements

for the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

July, 2020

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Anand Pathak** entitled "Uncovering Functional Correlates of Structural Organization in Brain Networks at Multiple Scales : From the worm to the human" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chair - Gautam Menon — Date: **22 December, 2020**

Supervisor/Convener - Sitabhra Sinha — Date: 22.12.2020

Member 1 - R. Rajesh — Date: 22/12/2020

Member 2 - Nita Sinha — Date: 22-12-2020

Member 3 - Pinaki Chaudhuri — Date: 22.12.2020

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

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

Date: 22.12.2020

Place: Chennai

Supervisor

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Anand Pathak

Name & Signature
of the student

DECLARATION

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

Anand Pathak

Name & Signature

of the student

List of Publications arising from the thesis

Journal

- **Published**

  1. **Pathak, A.**, Chatterjee, N. and Sinha, S. (2020). *Developmental trajectory of* Caenorhabditis elegans *nervous system governs its structural organization.* PLoS Computat Biol, 16(1): e1007602, (2020).

- **Communicated**

  1. **Pathak, A.**, Menon, S. N. and Sinha, S. (2020). *Mesoscopic architecture enhances communication across the Macaque connectome revealing structure-function correspondence in the brain.* Arxiv Preprint, 2007.14941

Conferences

1. "The Dynamical Worm : Analyzing spatio-temporal activity in a network model of *Caenorhabditis elgans* nervous system" at 8^{th} Dynamics Days Asia Pacific, 21-24 July, 2014 in Indian Institute of Technology Madras, Chennai, India.

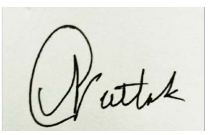
2. "Uncovering the mesoscopic organization of the macaque brain", at 3^{rd} Annual Conference of the Association for Cognitive Science (ACCS 2016), 3-5 October, 2016 in Indian Institute of Technology Gandhinagar, Gandhinagar, India

3. "Uncovering the mesoscopic organization of the macaque brain", at BrainModes 2017, 11-14 December, 2017 in National Brain Research Centre, Manesar, India

4. "Uncovering the mesoscopic organization of the macaque brain", at 27^{th} Annual Computational Neuroscience Meeting (CNS 2018), 13-18 July, 2018 in University of Washington, Seattle, WA, USA.

Anand Pathak

Name & Signature
of the student



Dedications

To my maternal grandfather **Shri Kastur Chand Jain** and my paternal grandfather **Shri Prabhakar Gopalrao Pathak**. One taught me love for science and the other taught me love for life.

Acknowledgements

First and foremost my heartfelt gratitude to my PhD supervisor **Prof. Sitabhra Sinha**, who not only constantly supported and guided me, and taught me everything I know about scientific research, but also let me grow by always challenging me to stretch my limits. I sincerely thank my collaborators and co-authors in my papers : **Dr. Nivedita Chatterjee** and **Dr. Shakti N. Menon** who were indispensable to my research that is presented in this thesis. Here I would specially like to acknowledge the efficient and helpful administrative, technical and housekeeping staff of IMSc for providing me with all possible infrastructural support all these years, with special mention to **Mr. Vishnu Prasad**, **Mrs. R. Indra** and **Mr. G. Srinivasan** for being ever dependable. The high performance computing facility at IMSc was crucial to my research and I want to duly thank the system administrators for maintaining the facility so efficiently. I cannot miss mentioning the good people of mess management who gradually became regular and jovial parts of our lives at IMSc. Further, I am grateful to all the members of my group for being so supportive, especially **Ashraf** and **Soumya** who were also my office-mates.

I want to thank my Doctoral Committee members for their advice and support : **Prof. R. Rajesh**, **Prof. Nita Sinha**, **Prof. Pinaki Chaudhuri** and specially **Prof. Gautam Menon**, who apart from providing me with regular encouragements, also taught a fantastic course on statistical mechanics in my coursework. There are so many people whom I want to thank for enriching me either through their academic discussions or by hosting fabulous conferences and schools that I attended: **Sandhya Kaushika, Upinder Bhalla, Raghavendra Singh, Sridhar Devarajan, Somdatta Sinha, Dipanjan**

Roy, Supratim Ray, Joby Joseph, Pranay Goel, Arnab Ghose, Nishikant Subedar, Nitin Williams, Areejit Samal and **James McClelland** are only few of them. In this regard I would specially mention the **Organization for Computational Neurosciences (OCNS), USA**, which gave me excellent opportunity to connect with the international community of scientists.

My friends have played an essential role in seeing me through my PhD. **Sheetal** and **Rampi**, who are my friends from antiquity, have been constant source of strength and moral support. Huge shout-out to my friends that I made at IMsc, with whom I have shared countless fond memories : **Madhu, Uday Kovur, Ashraf, Raghu** and **Prathik** to name a few. So many other colleagues who made my time at IMSc lively through their cheerful interactions, I thank all of them.

My family who has always stood solidly behind me : my **parents** and **brother**, **Sunil mama**, **Usha mami** and **Satej mama**, thank you from the bottom of my heart for keeping your faith in me. I am also sincerely grateful to all my in-laws, who happened to be in the same city, for providing their warmth and support all the while.

Finally, I want to thank my wife **Janaki**, my friend, colleague, partner in crime and more. I couldn't have done any of this without you being by my side.

Contents

Summary **1**

1 Introduction **3**

2 Developmental trajectory of *Caenorhabditis elegans* nervous system governs its structural organization **17**

2.1 Introduction 17

2.2 Materials and Methods 21

2.3 Results 31

2.4 Discussion 70

3 Mesoscopic architecture enhances communication across the Macaque connectome revealing structure-function correspondence in the brain **74**

3.1 Introduction 74

3.2 Materials and Methods 76

3.3 Results 84

3.4 Discussion 97

4 Invariances and diversity in the human structural connectome : A cohort study **101**

4.1 Introduction 101

4.2 Materials and Methods 103

4.3 Results 109

4.4 Discussion 127

5 The Hierarchy Index: A tool for revealing the sequential arrangement of processing levels in brain networks 131

5.1 Introduction 131

5.2 Methods 134

5.3 Results 140

5.4 Discussion 152

6 Conclusions 155

6.1 Summary of main results 156

6.2 Outlook 159

Appendix A 162

Appendix B 200

Appendix C 251

Appendix D 261

Bibliography 266

List of Figures

2.1 Birth time cohort membership and neurite lengths of neurons govern their connectivity. 32
2.2 Lineage of neurons affects their synaptic connectivity and spatial localization. 42
2.3 Lineage distance reveals developmental patterns of ganglia. 47
2.4 Birth times and lineage distances constrain connections between neurons whose cell bodies are spatially distant from each other. 51
2.5 Symmetrically paired neurons have a high probability of being connected and also exhibit strong association in their birth times and spatial positions. 53
2.6 Developmental histories of neurons show a bifurcation into early and late branches, with a predominance of motor neurons in the latter. 59
2.7 Neurons functioning as connectors between different network modules lead in development. 62
2.8 The developmental duration of functional circuit neurons are strongly indicative of their process length and connectivity. 65

3.1 Mesoscopic organization of the Macaque brain. 86
3.2 Classification of brain regions according to their intra- and inter-modular connectivity. 89
3.3 Physical distance between brain regions is seen to constrain their connectivity, but the modular organization of the network is independent of their three-dimensional spatial arrangement. 95

4.1 An increase in the occurrence of non-ubiquitous connections within a population leads to a steady shift in the corresponding link weight distribution towards higher values. 110
4.2 The variation of the weights of frequently occurring links over the population, as well as their frequency of occurrence, can both be described by a single link-specific Poisson process. 113
4.3 Exclusion of a small fraction of outliers can make deviating links fit a Poisson distribution. 116
4.4 Upon rescaling the weights, the structural connectivities consistently show a greater association with the corresponding functional connectivities across the population. 119
4.5 Dynamical simulation of whole brain activity using a neuronal population model suggests that structural connectivity with rescaled weights are better structural correlates for brain function than those with original weights. . . 122
4.6 The representative structural connectivity of a brain network can be resolved into two components. 126
4.7 "Basal" Network and "Superstructure" Network. 128

5.1 Schematic representation of hierarchy detection method. 137
5.2 Investigating the accuracy of the method in detecting hierarchical levels in random benchmark networks with embedded hierarchy. 142
5.3 Hierarchical structure of the *C. elegans* somatic nervous system. 145
5.4 Hierarchical structure of the macaque brain network. 148
5.5 Hierarchical structure of the human brain network. 150

A1 Birth cohort homophily is seen specifically for connections between neurons whose cell bodies are in close physical proximity. 192
A2 Spatial distribution of cell bodies of the neurons belonging to the somatic nervous system of *Caenorhabditis elegans*. 193

A3 Distribution of distances between cell bodies of synaptically connected pairs of neurons differs from that of all pairs. 194
A4 Distribution of distances between cell bodies of pairs of neurons connected by gap-junctions differs from that of all pairs. 195
A5 Dependence of the probability of connection between two neurons on the physical distance between their cell bodies. 196
A6 A stochastic branching model for the lineage tree of cells involved in the development of the *C. elegans* somatic nervous system. 197
A7 Developmental chrono-dendrograms for the Anterior (G1, top left), Dorsal (G2, top right), Lateral (G3, bottom left) and Ventral (G4, bottom right) ganglia, showing that each comprises multiple localized clusters of neurons. 198
A8 Developmental chrono-dendrograms for the Retrovesicular (G5, top left), Posterolateral (G6, top right), Preanal (G7, bottom left) and Dorsorectal (G8, bottom right) ganglia, showing that each comprises multiple localized clusters of neurons. 199
A9 Developmental chrono-dendrograms for the Lumbar ganglion (G9, left) and the Ventral Cord (G10, right), showing that each comprises multiple localized clusters of neurons. 200
A10 Connections between neurons born at different developmental epochs are over-represented when the cell bodies are far apart, suggesting the presence of active processes facilitating such links. 201
A11 Absence of segregated clusters in the developmental chrono-dendrograms for neurons having similar process lengths (viz., short, medium and long) suggest that process length is not exclusively determined by lineage. 202
A12 Physical proximity alone cannot explain the high degree of overlap between the cells that each member of a bilaterally symmetric pair of neurons connect to. 203
A13 Spatial distribution of the cell bodies of sensory, inter and motor neurons of the somatic nervous system of *Caenorhabditis elegans.* 204

A14 The number of synaptic connections of a neuron is influenced by its functional criticality, as well as, the developmental epoch in which it appeared. 205
A15 Comparison of the *Caenorhabditis elegans* connectome data obtained from Cook *et al.* and Varshney *et al.* 206
A16 Comparison of the *Caenorhabditis elegans* connectome data obtained from Varshney *et al.* and Haspel *et al.* 207
B1 The distributions of the in-degree, out-degree and total degree of the Macaque connectome. 210
B2 Visual representation of the comparison between the modular decomposition of the Macaque connectome obtained using spectral partitioning with that obtained using the Infomap method. 211
B3 Modularity of the Macaque connectome, shown as a function reconstructed from 10^3 partitionings obtained through simulated annealing method. . . . 212
B4 Robust modular membership in Macaque connectome. 214
B5 Modular decompositions of frontal, parietal, temporal and occipital lobes. . 227
B6 Modular decomposition of thalamus, basal ganglia, cingulate and insula. . 228
B7 Submodules of module #5. 238
B8 Distribution of the regions across the different categories R1-R7 for different modules. 240
B9 Extent of similarity between several anatomical subdivisions of the brain. . 242
B10 Connections between provincial hubs in Macaque connectome. 244
B11 Shift in intra- and inter-modular first passage times. 245
B12 Spatially independent modules for random surrogate networks ensembles. . 249
B13 Spatially independent modular organization for three types of random surrogate networks. 250
C1 Robustness of the detected community structure of the brain network. . . . 257
C2 Comparison of the accuracy of the community detection algorithms for different modular benchmark networks using normalized mutual information. 258

C3 Comparison of the accuracy of the community detection algorithms for different modular benchmark networks using the modularity parameter Q. . . 259
C4 Comparison of results obtained for modular benchmark networks with different values of CV 260
D1 The link-wise correlations between each structural connection and its corresponding functional connection over the population are extremely low. . . . 263
D2 Modules in the structural brain network of an individual subject, obtained using Newman Spectral Analysis. 264
D3 Modules in the rescaled structural brain network of an individual subject, obtained using the Newman Spectral Analysis. 265
D4 Similarity of Newman Spectral modules across different individual brain networks. 266
D5 Modules in the structural brain network of an individual subject, obtained using the *Infomap* method. 267
D6 Modules in the rescaled structural brain network of an individual subject, obtained using the *Infomap* method. 268
D7 Similarity of *Infomap* modules across different individual brain networks. . 269

List of Tables

5.1 Quantitative comparison between the hierarchical structures of neuronal/brain networks of *C. elegans*, macaque and human 152

A1 Process length homophily among neurons segregated into groups comprising cells with long, medium and short processes. 163

A2 Birth cohort homophily among neurons segregated into groups comprising cells differentiating before and after hatching (early and late-born, respectively). 165

A3 For synaptically connected neurons, process length of the pre-synaptic neuron primarily decides the average distance between the cell bodies. 166

A4 Neurons belonging to the somatic nervous system of *C. elegans* segregated into those which are born in the early (embryonic) and those born in the late (post-embryonic) developmental bursts. 168

A5 Time of appearance of neurons in the course of development of the *C. elegans* nervous system. 179

A6 Classification of *C. elegans* neurons according to their role in the mesoscopic organization. 185

A7 The neuronal composition of different functional circuits in the *C. elegans* somatic nervous system. 190

B1 List of nodes with robust modular memberships. 215

B2 Modular decomposition of the brain regions in different anatomical subdivisions of the Macaque brain. 224

B3 Functional characterization of modules 230

Summary

Networks, i.e., systems whose structure is completely described by specifying a set of vertices (or nodes), and the set of edges (or links) connecting these vertices to each other, are ubiquitous in nature. Among the large number of complex networks seen in the natural world, nervous systems, comprising neurons that communicate with each other through chemical (via synapses) or electrical (via gap junctions) means, are particularly fascinating because of the rapid and precise transfer of information they allow between different parts. Thus, explaining how the structural organization of networks in the brain aids in the attainment of its functional goals has become the aim of "connectomics", the collective term given to efforts at understanding various facets of the topological arrangement of connections between the cells and regions comprising the brain.

In this thesis, we have analyzed the network underlying the nervous system of different organisms and asked the following questions:

i. What are the key features that characterize the connection topology across the entire brain ?

ii. Can such structural features of the network be related to the functional goals of the nervous system (i.e., the "structure-function" relationship) ?

iii. What constraints ensure that a relatively invariant topological organization of the connections between neurons emerge over the course of development (i.e., the "wiring problem") ?

The different systems that we have considered in this thesis range between the somatic

nervous system of the hermaphrodite nematode *Caenorhabditis elegans*, the network of cortical and sub-cortical areas in the brain of the rhesus macaque monkey and an ensemble of networks of the whole brain reconstructed from diffusion tensor imaging of human subjects.

Our key findings include the determination of different kinds of homophily at play in the *C. elegans* nervous system. These tendencies of neurons with certain shared traits to connect to each other over the course of development appear to serve as key constraints during development that shape the structural organization of the nervous system. Our mesoscopic analysis of the macaque brain network reveals a robust modular structure, with each module corresponding to specific sensory modality or motor function. We reveal an underlying pattern of inter- and intra modular connectivity that promotes significantly faster communication across the network compared to other equivalent networks. Through analysis of a cohort of human structural brain networks we obtain a "representative network" characterizing the fundamental wiring patterns of the human brain. More informative than a simple averaging over the networks, this generic diagram indicates the frequency of occurrence and connection weight variability across the population for each link as well as provides better correspondence between the structural and functional connectivity. In this thesis we also propose an entirely new approach to analyze a specific type of mesoscopic organization in networks, viz., hierarchy. Using our hierarchy detection method, we uncover a robust arrangement of hierarchical levels in different nervous systems. Our analysis aims at contributing to a novel framework for understanding the structure-function relationship in networks of the brain.

Chapter 1

Introduction

"If the brain were so simple we could understand it, we would be so simple we couldn't."

–George Edgin Pugh in *The Biological Origin of Human Values* (1977)

Complexity as a fundamental paradigm to understand a system is not new. For instance, it was popularized by Philip Anderson in 1972 when he stated *"More is Different"* in his seminal paper [1]. Anderson had argued that in a complex aggregate of elementary components, understanding the properties of the components alone were not sufficient to understand or predict the behavior of the aggregate. Instead, at each successive level of complexity within a system, new features and phenomena emerge. Such *emergent behavior* is as fundamental as the properties of elementary components themselves. Thus, at every successive level we require a fresh set of ideas and paradigms to understand the system. Statistical mechanics and condensed matter physics were among the early fields that focused on understanding emergent phenomena in complex systems that arise due to interactions between the constituent elements. Properties like temperature, entropy, abrupt phase transitions, critical phenomena, ferromagnetism, superconductivity, etc., can be only understood through paradigms of statistical mechanics and many-body physics [2], rather than through the physics of elementary particles. This idea of multiple hierarchical levels of complexity has been later extended to biological systems as well. The hierarchy of

complexity in biological systems range from the molecular level (e.g., proteins, DNA), to the cellular level (e.g., organelles, unicellular organisms), to cell assemblies (e.g., neuronal circuits, cell signaling), extending upto the level of single organisms (e.g., nervous system, immune system) and populations of organisms (which form the domain of population ecology and social sciences).

Brain, which is a specific type of biological system and has neurons as its elementary components, is arguably one of the most complex of systems encountered so far by humans. The mammalian brain comprises extremely large aggregate of neurons interacting with each other via synapses and gap-junctions, allowing rapid and precise transfer of information from one part to another. It has specifically evolved to enable the organism to process sensory stimuli efficiently and respond fast to the changing environment in order to survive. In the scheme of hierarchical levels of complexity in the brain, starting from the level of neurons and their dynamical interactions, we observe the emergence of large scale spatio-temporal activity patterns in neural populations, which ultimately give rise to behavior and cognition. While there are multiple levels of complexity occurring inside an individual brain, there is a large variability in brain complexity across the species as well. In this, the nematode *Caenorhabditis elegans* with a small nervous system of ~ 300 neurons, lies at one end of the complexity spectrum, while the mammalian brain with billions of neurons is placed at the other extreme. Such high degree of complexity and variability does not simply arise from the size of the system alone, viz., the number of constituent neurons, but can also be attributed to the specific pattern in which the neurons interact with one another. Unlike many physical systems studied in many-body physics which also have equally large number of components, the interactions between neurons in a brain are neither regular (as in a lattice) nor completely random (as in disordered systems). Furthermore, the interactions are highly non-linear. Therefore, in order to understand the complexity of the brain, we have to first understand the organization of interactions between its components. To this end, the theory of complex networks comes across as a natural framework to describe the brain and over the recent past, has proven to be one of the most potent and useful paradigms in neuroscience [3, 4]. We now have a

established field of *Network Neuroscience* [5] that is dedicated to the study of brains though extensive network analysis over multiple spatio-temporal scales and across several species. Before proceeding further to discuss the central theme of this thesis, viz., how studying different brain networks of varying complexities might provide fundamental insights about the development, organization and functioning of the brain, the following section provides a brief overview of a few basic concepts in network science that are relevant for this thesis.

Fundamental concepts of network analysis: a primer

Representation of networks

A *network* or a *graph* is essentially a set of nodes (or vertices) connected to each other through links (or edges). In principle, a network representation can be formulated for any system that comprises sub-components (nodes) having a specific type of inter-relation to one another (links). The specification of nodes and corresponding links depends entirely on the context. For instance, for the world wide web, the web pages are the nodes and hyperlinks connecting the pages are the links whereas, for a neuronal network, the neurons are the nodes and the synapses and gap-junctions are the links. In its simplest form, a link between two nodes i and j is represented as (i, j), which only informs us about the existence of a link between i and j. One can also ascribe a direction to the link such that (i, j) refers to link from node j to node i where as (j, i) denotes a link from i to j. A compact way of representing a network is through the *adjacency matrix* A_{ij} whose components are described as:

$$A_{ij} = \begin{cases} 1 & \text{if there exists link between nodes } i \text{ and } j, \\ 0 & \text{otherwise.} \end{cases} \tag{1.1}$$

For a directed network, $A_{ij} = 1$ only if there exists a directed link from j to i. The information about existence or non-existence of links between all the node-pairs for a given

network is often called its *connection topology*. Furthermore, in many cases the links can also be ascribed a strength value which is also known as *weight*. In a weighted network, the adjacency matrix elements A_{ij} denoting an existing link between i and j can take any non-zero value. In the context of neuronal networks, the weight can describe the synaptic weights. In macro-connectomes where the brain regions form the nodes, the thickness of axonal pathways between them can be described by link weights. Thus, based on the types of links, a network can be classified as directed or undirected and also as weighted or unweighted. All these networks can be represented by their corresponding adjacency or weight matrices.

Structural and functional brain networks

The connectivity within brain is largely described using two fundamentally distinct types of networks, viz., *structural* and *functional*. A structural network comprises the anatomical connectivity between the constituent nodes such as synapses and gap-junctions between neurons, or, white matter tracts between the brain regions. They are obtained via in vivo imaging of physical connections between neurons or brain regions using techniques such as electron microscopy (for synapses and gap-junctions), tract tracing (for axonal pathways between brain regions in non-human animals such as macaque, rat etc.) and diffusion tensor imaging (DTI, for axonal pathways between brain regions in humans). Structural network forms the fundamental basis of interaction between the network components. Depending on the system and the method of recording, structural networks can be directed, undirected, weighted or unweighted.

A functional network describes the correspondence between the neurophysiological activities of constituent nodes. For brain regions, the electrophysiological activities are measured via electro-encephalograph (EEG), magneto-encephalograph (MEG) and functional magnetic resonance imaging (fMRI). The cross correlation (in time domain) or spectral coherence (in frequency domain) between the temporal evolution of these activities give the connection weight between the nodes. By construction the functional networks are

weighted undirected networks.

Properties of networks

Basic properties required to characterize any given network can be broadly grouped into three levels, viz., microscopic, macroscopic and mesoscopic, depending upon the scale of analysis. The scale in this context is meant in the topological sense, as opposed to the physical sense of length and time scales.

Microscopic Properties

Microscopic properties (or local properties) pertain to single nodes and their connected neighbors. Sometimes they are also used to describe the features associated with a small sub-network of a large network.

Node degree refers to the total number of connections that the node has with other nodes in the network. For directed networks, in-degree refers to total number of connections received by a node while out-degree refers to the number of connections sent out by a node.

Clustering coefficient quantifies how densely the neighbors of a node (i.e., those which are linked to it) are connected to each other, in terms of the ratio of the number of connections between the nearest neighbors to the total number of possible connections between them.

Motifs represent recurring patterns of connections between a few nodes $(2-5)$ that occur with significantly higher frequency within a given network as compared to chance [6].

Path length measures the minimum number of links required to be traversed to go from one node to another.

Centrality of a node measures the proportion of shortest paths between all other node pairs

that pass through it.

In general, nodes with high values of degree and centrality are considered to be of importance and are referred to as *hubs*.

Macroscopic Properties

Macroscopic properties (or global properties) pertain to various statistical properties that describe the network as a whole. The following are a few examples of such properties.

Degree distribution: The probability distribution of nodal degrees across the network provides one of the primary large scale descriptions for a given network. In some cases, the type of degree distribution seen in a network might indicate a certain type of generative mechanism. For instance, nodal degrees in randomly generated networks with homogeneous connection probabilities show Poisson distributions. Among the other frequently occurring types of degree distributions that are seen in naturally occurring networks are the exponential and power-law distributions. Networks with power-law distributions are often called *scale-free networks* [7].

Average path length & average clustering coefficient: Averaging the values for path lengths between all pairs of nodes and clustering coefficients over all nodes, provides us another set of important macroscopic descriptors for a network. Networks with high average clustering coefficient and high average path length indicate a regular lattice-like connectivity, where connections occur only between the nearest neighbors in physical space. On the other hand, networks with entirely random connectivity show low values of average path length and average clustering coefficient. However, in most naturally occurring networks, we see low average path-length but high average clustering coefficient, and such networks are also known as *small-world networks* [8].

Assortativity describes a tendency of high degree nodes to connect with other high degree

nodes and of low degree nodes to connect with other low degree nodes. It is measured as correlation between the degrees of connected nodes.

Mesoscopic Properties

Naturally occurring networks, such as the networks of the brain, are non-homogeneous and highly complex, often comprising multiple levels of organizational structures within them. These structures are revealed by analyzing the network at an intermediate topological scale between the microscopic and macroscopic descriptions given above. Thus, the properties describing these underlying structures within a network are known as mesoscopic properties. The following are few examples of mesoscopic properties commonly seen in networks.

Community structure: Networks exhibiting community structure or *modularity*, have their nodes grouped into separate communities of nodes, also called modules, such that the connection density between nodes inside the same module is significantly higher than the connection density across nodes of different modules [9]. In many cases, the network can have hierarchical modularity, i.e., multiple levels of nested modules such that larger modules are divided into smaller modules which are further divided into even smaller modules and so on [10]. Modular structure has been observed in a many brain networks [11, 12, 13, 14, 15, 16, 17, 18, 19]

Core-periphery structure: The core periphery organization in a network comprises of a central core, whose members are connected to each other, as well as, to the remainder of the network, and a periphery, whose members are connected to the core but not to one another [20]. Such a structure can comprise multiple layers of peripheral nodes or shells around the core, where each successive outer shell is peripheral with respect to all its inner shells and the core. Core-periphery organization has been observed in several neuronal and brain networks such as *C. elegans* nervous system [21], macaque brain network [22] and human brain network [23, 24].

Rich-club is a type of mesoscopic organization in which a set of high degree nodes ("rich" nodes) form a densely inter-connected subnetwork as opposed to low degree nodes which are sparsely inter-connected. The density of connection between rich club nodes is significantly higher than that expected from a random homogeneous network with similar degree sequence [25]. A rich club ordering has been observed in a variety of brain networks such as *C. elegans* nervous system [26], macaque brain network [16] and human brain network [27].

Hierarchical organization: A highly relevant type of mesoscopic organization which has not got as much attention as the above described mesoscopic structures in brain networks is hierarchical organization. In a hierarchical network, the nodes are grouped into sequentially arranged hierarchical layers such that only consecutive layers have dense connections between them, whereas nodes of non-consecutive layers, as well as, nodes belonging to the same layer, are sparsely connected. In Chapter 5 of this thesis, we delve into this type of mesoscopic organization in detail.

Structural brain networks across species : a short overview

Quest for the connectome

Connectome, a term that was coined fairly recently [28], refers to a complete map of inter-neuronal connections within the brain, also called the *"wiring diagram"*. Often it has been flexibly used to represent the mapping of structural connectivity at higher resolutions such as brain regions as well [29]. Lately, there is a rapid growth in the community of neuroscientists who believe that the *connectome* might hold the key to understanding the repertoire of complex cognitive behavior, somewhat in the same spirit as the *genome*, i.e., the complete sequence of nucleotides in the DNA, was considered to hold the key for understanding how an organism develops and functions. This idea of structural connectome providing the basis for understanding brain function has been discussed in depth by Sebastian Seung in a popular book [30]. Put another way, the wiring diagram of the

brain indicating its structural connectivity is considered to be the "obligatory foundational model for understanding functional localization at molecular, cellular, systems and behavioral organizational levels." [31]. Thus, the significant role played by the anatomical architecture of brain connectivity at multiple scales in cognitive processes and behavior has been repeatedly stressed in the literature [32, 33]. With an increasing number of studies on model neural circuits, as well as, large scale connectivities in different species, the deep relation between the structural connectivity and various cognitive functions like learning and memory [34, 35, 36], motor control [37] and vision [38] is being gradually explicated.

One of the primary roadblocks on the path towards "solving" the brain is obtaining an accurate and complete structural connectome. Mapping the connectivity of the entire brain of an organism is an extremely difficult technological and computational challenge. Till date the only organism whose entire connectome has been mapped at cellular resolution is the nematode *Caenorhabditis elegans* [39]. For other species, ranging from insects to mammals including humans, attempts to completely map the connectome are yet to reach this goal. Towards this end, there have been huge investments on long term projects such as the Blue Brain Project [40], the Human Connectome Project [41], Japan's Brain/MINDS project [42] and the China Brain Project [43]. This is a testament to the promise of potential breakthroughs in our understanding of brain functions that a complete knowledge of the structural connectivity holds.

Universality in structural organization across species

In this thesis, we look at three different connectomes, viz., the somatic nervous system of the nematode *C. elegans*, the macaque brain and the human brain, that vary both in scale as well as complexity. However, there is a central unifying theme that ties the analyses of these varied systems that are presented in this thesis. This perceived unity is not simply because all of them are investigated using the paradigm of network theory. Rather, it emerges from a more fundamental universality which is intrinsic to neuronal systems in general. Here, universality is meant in the following sense. It is widely known

that as we look at progressively larger organisms in the animal kingdom, they do not simply grow in size and number of cells but also grow rapidly in terms of organizational and behavioral complexity as well [44, 45, 46]. However, in spite of the large degree of variation in scale and complexity across species, the underlying brain networks exhibit certain recurring organizational principles. This universality is particularly evident in the mesoscopic organization of the brain networks. In the preceding section, we have already pointed out a few of the common types of mesoscopic organization that have been observed in many brain networks, viz., modularity [14, 18, 19], core-periphery structure [21, 22, 23] and rich-club organization [26, 16, 27]. This has led to the emergence of the sub-field of *comparative connectomics* [47] which aims to uncover the general principles of network organization by carrying out cross-species quantitative studies.

The universality in network architecture could be primarily due to two possible reasons:

1. Fundamental constraints governing the wiring of neuronal networks, such as those arising from the trade-off between wiring length cost and communication efficiency, are conserved across all species.

2. The problem that a nervous system is trying to solve, viz., to efficiently process sensory stimuli and respond through appropriate action, is common for all species.

For the first of these, viz., developmental constraints in neuronal wiring arising from competition between metabolic cost and functional benefits, several mechanistic principles [48, 49, 13] and generative models have been examined [50]. However, the second, viz., the common problem which all neuronal networks are designed to solve, is not very straightforward to formulate. In this direction, Karl Friston has recently proposed a theoretical framework called the Free-energy Principle [51]. It formulates all neuronal systems as inference engines trying to optimize their free energy function through action and perception. It attempts to provide a "unified theory of brain" which would explain several fundamental aspects of brain structure and function [52].

From the above viewpoint, an in-depth analysis of structural organization in different brains is imperative not only for understanding the functioning of individual brains, but

also for uncovering the fundamental generative principles that might be common across all species. This was the primary motivation for the work reported in this thesis. Over the next few chapters, in addition to exploring the structure-function relationship within brain networks of varying complexities, we also draw concrete inferences about the generative mechanisms underlying their structural organization. The contents of this thesis are outlined in the following section.

Overview of the thesis

In this thesis, we have analyzed the network underlying the brain (or nervous system) of different organisms and asked the following questions:

i. What constraints ensure that a relatively invariant topological organization of the connections between neurons emerge over the course of development (i.e., the "wiring problem") ?

ii. What are the key mesoscopic features other than modularity that characterize the connection topology across the entire brain ?

iii. Can such structural features of the network be related to the functional goals of the nervous system (i.e., the "structure-function" relationship) ?

iv. What can we know from the intra-species variability of structural brain networks about the generative mechanisms and the functional interpretations of the links ?

The different systems that we have considered in this thesis range between the somatic nervous system of the hermaphrodite nematode *Caenorhabditis elegans*, the network of cortical and sub-cortical areas in the brain of the rhesus macaque monkey and an ensemble of networks of the whole brain reconstructed from diffusion tensor imaging of human subjects.

In **Chapter 2** we address the fundamental question of how does the characteristic wiring diagram of a neuronal network emerge during the development of an organism. For

this purpose, we focus on the nervous system of the hermaphrodite nematode *Caenorhabditis elgans*. While the organism does not possess a centralized 'brain', it has a somatic nervous system comprising 279 neurons. The fact that the connectome, as well as, the developmental chronology and lineage of all the cells, are invariant across different individuals, suggests that the emergent network structure is somehow generated by the developmental process itself. Using all available data about the connectivity and development, in conjunction with information about the spatial organization of every neuron belonging to the somatic nervous system of the adult hermaphrodite *C. elegans*, we show that the developmental trajectory itself might be providing the key constraints that shape the observed structure of the nervous system. These constraints may be thought of as strategies contributing to the eventual wiring of the nervous system, which is extremely precise and invariant from individual to individual. Insights obtained from studying how the structural and functional organization of the *C. elegans* nervous system is shaped by the developmental process might be helpful in understanding the mechanism of neuronal wiring in other organisms as many of the robust strategies seen here might be conserved across different species at multiple scales.

In **Chapter 3**, we probe the structure-function relationship in the macaque brain. The network is defined at the resolution of brain areas, with the vertices being anatomically defined regions of the brain and the links being the axonal tracts connecting these regions. While previously there have been studies on the community structures of different brain networks including that of the macaque, most of these considered only subsets of the entire network. As the brain is considered to be a system where information is processed in a highly distributed manner (rather than being localized in various compartments as previously thought), it is necessary to consider the entire brain in order to understand how the structural organization affects its function. In our study, we have focused on understanding how the mesoscopic organization of the network, specifically its modular arrangement, is reflected in the functioning of the brain. In accordance with previous studies, we find several network modules that are spatially localized. Different sensory modalities and motor functions can be associated with the modules, although we do find

intriguing exceptions. More importantly, we discover that the specific pattern of intra- and inter-modular connectivity between the brain areas allows extremely rapid diffusion across the network, that is fast compared to even homogeneous networks. Finally, we explore the possible relationship between the observed modular structure and the spatial configuration of the network which might point towards a generative mechanism for the emergence of the modules. Specifically, we ask whether the observed modules are simply an outcome of the spatial constraints in brain wiring. We show that even after taking into account the connection bias introduced by physical distance constraints to redefine the modules in a space-independent manner, an almost identical modular organization of the network is still obtained. This implies that the modular structure that is seen in the brain network is not a result of spatial constraint of connectivity but is possibly an outcome of the functional advantage of rapid communication.

In **Chapter 4**, we explore the intra-species variability of structural connectomes in humans. The human brain is apt for studying the variability of the underlying network structure as not only do studies on human cognition provide a much broader functional and behavioral repertoire compared to other species, but with the advent of state of the art non-invasive imaging techniques like diffusion tensor imaging (DTI) and functional magnetic resonance imaging (fMRI), we can complement functional studies with accurate, highly resolved maps of brain connectivity. The variability in brain networks across the population can occur not only in terms of their connection topology, but also in terms of the connection weights. Using a cohort of structural brain networks obtained from 196 human subjects, we show that a significant fraction of the connections have their weights follow a Poisson distribution. This suggests independent, discrete random processes as the generative mechanism. It also indicates that a single mechanism determines the wiring, as well as, the weights. This result provides a basis for rescaling the connection weights of each network, such that the structural connectome matches the functional connectome better - allowing a functional interpretation of the structural links. Our analysis also yields a generic representative network for the human brain which can be resolved into a "basal" network and a "superstructure" network, the former being relatively invariant

across individuals while the latter shows much more variability.

In **Chapter 5**, we examine a specific type of mesoscopic organization that is extremely relevant for brain networks but has received relatively less attention, namely, hierarchical organization. In hierarchical network topology, the nodes of the network are segregated into distinct levels that are sequentially positioned such that the density of connections is high only between consecutive levels. By contrast, the connections between other levels or among the nodes of the same layer are sparse. Such type of organization is found commonly in brain networks, e.g., the cortical columns in the cortex consist of hierarchically organized layers of neurons. While, modular organization in a network implies a form of compartmentalization at the mesoscopic level, a hierarchical organization gives the sense of directional and sequential flow across the network. We have developed a novel method to determine the hierarchical structure of a given network and detect the underlying levels. The benchmark tests show that the method is able to detect the embedded hierarchical levels with great accuracy. We also use the method to identify robust hierarchical organization in the somatic nervous system of *C. elegans* and the brain networks of macaque and human.

In **Chapter 6**, we conclude with a discussion about how studying brain networks at different scales allow us to appreciate the structural and functional commonalities across several orders of size and complexity. We also indicate possible future directions and further scope of the studies presented in this thesis.

Chapter 2

Developmental trajectory of *Caenorhabditis elegans* nervous system governs its structural organization

2.1 Introduction

The presence of an efficient machinery for responding immediately to changes in the environment with appropriate actions is essential for the survival of any organism. In almost all multicellular animals, this role is played by the nervous system comprising networks of neurons, specialized cells that rapidly exchange signals with a high degree of accuracy. It allows information about the environment obtained via sensory receptors to be processed and translated into output signals conveyed to effectors such as muscle cells. In even the simplest of such organisms, the structural description of the interconnections between neurons provided by the connectome presents an extremely complicated picture [53]. How the complex organization of the nervous system is generated in the course of development of an organism, occasionally referred to as the "brain wiring problem" [54], is one of the most

challenging questions in biology [55, 56]. Only over the past few decades is the intricate interplay of different developmental phenomena, including cellular differentiation, migration, axon guidance and synapse formation, responsible for the formation of the network, being gradually revealed [57, 58, 59, 60, 48].

The free-living nematode *Caenorhabditis elegans*, the only organism whose entire connectome has been reconstructed so far [39, 61], is the natural choice for a system in which to look for principles governing the development of complexity in the nervous system [62]. The nervous system of the mature hermaphrodite individuals of the species comprises 302 neurons, which is about a third of the total complement of 959 somatic cells in the animal. Their lineage, spatial position and connections to each other appear to be almost invariant across individuals [39, 63]. The small number of cells constituting the worm has made it a relatively tractable system for understanding the genetic basis of metazoan development and behavior. This, however, belies the sophistication of the organism which exhibits almost all the important specialized tissue types that occur in larger, more complex animals, prompting it to be dubbed as a "microchip animal" [64]. The availability of its complete genome sequence [65] along with detailed information about the cell lineage [66, 67] means that, in principle, the developmental program can be understood as a consequence of genetically-encoded instructions and self-organized emergence arising from interactions between diverse molecules and cells [68].

The "wiring problem" for the *C. elegans* nervous system had been posed early on with Brenner essentially raising the following questions: how are the neurons spatially localized in their specific positions, how they connect to each other through synapses and gap junctions forming a network with a precisely delineated connection topology, and what governs the temporal sequence in which different neurons appear over the course of development [69]. Subsequent work has identified several mechanisms underlying the guidance of specific axons and formation of synapses between particular neurons [70, 71, 72]. However, the minutiae of the diverse molecular processes at work may be too overwhelming for us to arrive at a comprehensive understanding of how the complexity manifest in the nervous system of the worm arises. Indeed, it is not even clear that all the guidance

cues that are involved in organizing the wiring are known [73]. An analogous situation had prevailed five decades earlier when *C. elegans* had been first pressed into service to understand how genetic mutations lead to changes in behavior of an organism. Brenner had responded to this challenge by analyzing the system at a level intermediate between genes and behavior [69]. Thus, the problem was decomposed into trying to understand (a) the means by which genes specify the nervous system (*how is it built ?*) and (b) the way behavior is produced by the activity of the nervous system (*how does it work ?*) [69, 68]. In a similar spirit, for a resolution of the "wiring problem", we may need to view it at a level intermediate between the detailed molecular machinery involving diffusible factors, contact mediated interactions, growth cone guidance, etc., and the organization of the neuronal network in the mature worm. Specifically, in this chapter, we have focused on uncovering a set of guiding principles that appear to govern the neuronal wiring and spatial localization of cell bodies, and which are implemented by the molecular mechanisms mentioned earlier (and thus genetically encoded). From the perspective of the three-level framework proposed by Marr [74, 75] for understanding the brain [56], viz., comprising (i) goals (and the logic of strategies for achieving them), (ii) algorithmic and (iii) implementation levels, such principles can be viewed as *strategies* for achieving specific network designs realized over the course of development [54].

For this purpose, we have used the analytical framework of graph theory, which has been successfully applied to understand various aspects of brain structure and function, in both healthy and pathological conditions [76, 3, 77, 78, 79, 80]. For the specific case of the *C. elegans* nematode, application of such tools has revealed the existence of network motifs [81], hierarchical structure [21], community (or modular) organization [82] and a rich club of highly connected neurons [26]. Comparatively fewer studies have focused on the evolution of the network during development of the nematode nervous system that we consider here [83, 84]. We have integrated information about spatial location of cells, their lineage, time of appearance, neurite lengths and network connectivity to understand how its developmental history constrains the design of the somatic nervous system of *C. elegans*, specifically the 279 connected neurons which control all activity of the worm except the

pharyngeal movements. Thus, our study complements existing work that has focused more on understanding the structural organization of the network using efficiency and optimality criteria such as minimization of the wiring cost, delineated by the physical distance between neurons [85, 86, 87, 88, 89, 90, 91].

The key questions related to development that we address here involve the spatial location of the cell bodies (*why is the neuron where it is, relative to other neurons ?*), the temporal sequence in which the cells appear (*why is it that certain neurons are born much earlier than others ?*) and the topological arrangement of their inter-connections (*why does a neuron have the links it does ?*). As reported in detail below, we find that these questions are related to the existence of general principles that can be expressed in terms of different types of homophily, the tendency of entities sharing a certain feature to preferentially connect to each other. We discern four different types of homophily, involving respectively, process or neurite length of neurons, the time of their appearance, their lineage history and bilateral symmetry. We also estimate the relative contributions of each of these four factors (which we show are linearly independent of each other) in determining the connectivity. Although it had been reported earlier that the probability of connection between two neurons decreases with increasing difference in their birth times [83], we show this result to be much more nuanced in that birth cohort homophily is predominant only for connections between neurons whose cell bodies are physically proximate. Our results also help reveal that the ganglia, anatomically distinct bundles into which the neurons are clustered in the nematode, are formed of several groups (or families) of cells, neurons within each group being closely related. Furthermore, as lineage relation between neurons is an important factor that influences the structure of the neuronal network, we have presented a stochastic generative model for the lineage tree of cells. By invoking a simple asymmetric branching process, such a model captures several features of the empirically observed lineage tree.

At a higher level of network organization, we show that neurons which play a vital role in coordinating activity spanning large distances across the network by connecting together distinct neuronal communities (or modules) also appear quite early in the sequence of development. This observation (along with others, such as linking the functional type

of neurons, viz., sensory, motor and inter, to their time of appearance) helps link the situation of a specific cell in the temporal hierarchy to which all neurons belong, with its function. We also provide an analysis of the inter-relation between functional, structural and developmental aspects, focusing on neurons identified to belong to different functional circuits, such as those associated with mechanosensation [92, 93, 94], chemosensation [95], etc. This provides us with a more nuanced understanding of the relation between the time of appearance of a neuron and the number of its connections. Our results suggest that developmental history plays a critical role in regulating the connectivity and spatial localization of neurons in the *C. elegans* nervous system. In other words, development itself provides key constraints on the system design. In addition, the tools we employ here for revealing patterns hidden in the lineage and connectivity information, including novel visual representations of developmental history, such as chrono-dendrograms, provide insights into principles governing the wiring of nervous systems that may be common across several organisms.

2.2 Materials and Methods

Data

Connectivity. Information about the connections between different neurons obtained using serial section electron micrography were first reported in Ref. [39]. Subsequent analysis of the images has led to discovery of many more connections which have been published in Refs. [61] and [96] (which differ marginally in the connections they report). A more recent re-analysis has added further connections to the connectome [97]. We have used the information about connectivity between 279 connected neurons of the *C. elegans* nervous system from the latest dataset which is published in Ref. [97]. We have explicitly verified that our results remain substantially unchanged on using the earlier connectome dataset from Ref. [61]. The differences between the connectivity reported in the above-mentioned databases are visually represented in Appendix A, Figure A15 and Figure A16.

Lineage. We have used information about the lineage distance between 279 connected neurons of the *C. elegans* somatic nervous system from the database published in Ref. [61], accessible from an online resource for behavioral and structural anatomy of the worm [98].

Time of birth for neurons. We have transcribed the time of appearance of each neuron over the course of development of the organism from lineage charts provided in Refs. [66, 67]. This is provided in Table A5.

Time of cell-division for progenitor cells The information about the time of each cell-division, starting from the zygote, that occurs over the course of development of the *C. elegans* somatic nervous system, and which has been used for generating the chrono-dendrograms shown here, are provided in Refs. [66, 67], accessible from an online interactive visualization application [99].

Physical distance. We have used information on the positions of the neurons from the database reported in Ref. [100], accessible online from `https://www.dynamic-connectome.org/`. The location information provides coordinates of each neuronal cell body projected on a two-dimensional plane defined by the anterior-posterior axis and the dorsal-ventral axis.

Neuronal process length. Process or neurite refers to any projection from the neuronal cell body, which can be either an axon or a dendrite. Lengths of the processes extending from each cell body has been estimated from the diagrams of the neurons provided in Appendix 2, Part A of Ref. [101] and from an online resource for the anatomy of the worm [98].

Ganglia and functional types. The information about the ganglion to which a neuron belongs and the functional type of each neuron (viz., sensory, motor or interneuron) has been obtained from the database provided in Ref. [102].

Functional circuits. The identities of the neurons belonging to each of the functional circuits have been obtained from the original references for each circuit. The functional circuits analyzed in this chapter and their respective original references are as follows : (F1) mechanosensation [92, 93, 94], (F2) egg laying [103, 104, 105], (F3) thermotaxis [106, 107, 108], (F4) chemosensation [95], (F5) feeding [39, 92, 109], (F6) exploration [39, 92, 109]

(F7) tap withdrawal [93, 110], (F8) oxygen sensation [111, 112] and (F9) carbon dioxide (CO_2) sensation [113, 114] The neuronal composition of each of these functional circuits is given in Table A7.

Modularity

A network can be partitioned into several communities or topological modules, defined such that neurons in a given module have a much higher probability of being connected to other neurons in the module compared to neurons that do not belong to it, by maximizing the modularity value Q [115] associated with a given partitioning, viz.,

$$Q = \frac{1}{L} \sum_{i,j} \Big[A_{ij} - \frac{k_i^{in} k_j^{out}}{L} \Big] \delta_{c_i c_j}. \tag{2.1}$$

Here, $\mathbf{A}$ is the adjacency matrix describing the connections of the network ($A_{ij} = 1$, if neuron i receives a connection from neuron j, and $= 0$ otherwise). The in-degree and out-degree of a node i are given by $k_i^{in} = \sum_j A_{ij}$ and $k_j^{out} = \sum_i A_{ij}$, respectively. The total number of links in the network is given by $L = \sum_{i,j} A_{ij}$. The Kronecker delta function $\delta_{ij} = 1$, if $i = j$, and 0, otherwise. The indices c_i, c_j refer to the modules to which the neurons i and j, respectively, belong. For an undirected network, such as that defined by the set of connections between neurons via gap-junctions, the adjacency matrix is symmetric (i.e., $A_{ij} = A_{ji}$) and $k_i^{in} = k_i^{out} = k_i$. The value of Q expresses the bias that a neuron has to connect to members of its own community (which could be defined in terms of any distinguishing characteristic of the cells, e.g., process length), relative to the null model. The latter corresponds to an unbiased, homogeneous network where the probability of connection between two nodes is proportional to the product of their respective degrees. A positive value of Q that is significantly higher than that obtained from an ensemble of randomized surrogate networks would indicate the existence of homophily.

To further establish homophily, we also show that the bias for connecting to neurons belonging to a particular class (e.g., short process length) can be observed only among

neurons in that class, and not between neurons belonging to different classes (e.g., between short and long process neurons). For this, we define Q values for each pair of classes (X and Y, say) as follows:

$$Q_{XY} = \frac{1}{L}\sum_{i,j}\Big[A_{ij} - \frac{k_i^{in}k_j^{out}}{L}\Big]\delta_{c_iY}\cdot\delta_{c_jX}, \tag{2.2}$$

where the symbols have the same meanings as in Eq. 2.1. Note that, the Q for the entire network calculated in terms of Eq. 2.1 is related to these class-specific modularity values as $Q = \sum_{X\in\text{all classes}} Q_{XX}$. Homophily is established upon demonstrating that the values of Q_{XX}, Q_{YY}, etc. are significantly higher than that of Q_{XY}, Q_{YX}, etc.

Bimodality coefficient

The bimodal nature of a probability distribution can be characterized by calculating its bimodality coefficient [116]:

$$BC = \frac{m_3^2 + 1}{m_4 + 3\cdot\frac{(n-1)^2}{(n-2)(n-3)}}, \tag{2.3}$$

where m_3 is the skewness, m_4 is the excess kurtosis and n represents the sample size. A distribution is considered to be bimodal if $BC > BC^*$ where $BC^* = 5/9$. This benchmark value corresponds to a uniform distribution, and if $BC < BC^*$, the distribution is considered unimodal.

Process length randomization

To establish statistically significant evidence for process length homophily, the empirical network is compared with an ensemble of networks obtained from the empirical one by randomly assigning process lengths (short, medium and long) to the neurons while ensuring that the total number of neurons in each process length category, viz., N_S, N_M and N_L, respectively, (as well as, all other properties of the network, such as connec-

tivity) remains unchanged. In practice, this is done by first partitioning the neurons into three communities according to process length and ordering the neurons in sequence according to the module they belong. Thus, neurons $i = 1, \ldots, N_S$ have short processes, neurons $i = N_S + 1, \ldots, N_S + N_M$ have medium length processes, and neurons $i = N_S + N_M + 1, \ldots N_S + N_M + N_L$ have long processes. Then, to create each member of the surrogate ensemble, this sequence is randomly permuted and the first N_S neurons are assigned short process length, the next N_M neurons are assigned medium process length and the remaining N_L neurons are assigned long process length. The modularity Q calculated for networks with such randomized module membership (corresponding to a null model where process length homophily is non-existent by design) is expected to be small.

Network randomization constrained by neuronal process lengths

An ensemble of surrogate networks is constructed by randomizing the connections of the empirical network, subject to different constraints. Each member of the ensemble is constructed by repeatedly selecting a pair of directed connections, e.g., $p \to q$ and $u \to v$, and rewiring them such that the in-degree and out-degree of each neuron remains invariant, i.e., $p \to v$ and $u \to q$. If these new connections already exist, this rewiring is disallowed and a new random selection for a pair of directed connections done. In addition, information about the spatial location of the cell bodies and that of the process lengths of neurons are used to further constrain the connections. This ensures that un-physical connections, such as between two short process neurons (i.e., each has a process length that is less than a third of the body length of the nematode) whose cell bodies are placed apart by more than $2L/3$ (L: total body length of the worm), do not appear through the randomization. In practice, this constraint is imposed as follows. In absence of precise knowledge of the length of each process, depending on the length process category to which each neuron belongs, an uniformly distributed value (lying between $[0, L/3]$ for short, between $[L/3, 2L/3]$ for medium and $[2L/3, L]$ for long process neurons) is assigned as the process length for a neuron. The distance between the cell bodies of a pair of neurons that have been selected

randomly for connection is then compared against the sum of their process lengths. If the latter is greater than the former, the connection is allowed, else not. The rewiring steps are repeated 5×10^5 times to construct each of the randomized networks belonging to the surrogate ensemble. The entire ensemble consists of 100 realizations of such randomized networks.

Lineage randomization

To establish that neurons belonging to the same ganglion are closely related in terms of their lineage, we compare the properties of the lineage distance distribution within and between ganglia obtained for the empirical network with those obtained upon randomizing the lineage relations. This is done by repeatedly selecting a pair of neurons at random on the lineage tree and exchanging their positions on the tree. This procedure is carried out 10^4 times for a single realization. This ensures that, in the randomized networks, the lineage relation between neurons is completely independent of whether they belong to the same ganglion or not. In order to compare the properties of the empirical network with its randomized version, an ensemble of 10^3 realizations is considered. To quantify the deviation of the empirical intra- and inter-ganglionic lineage distance distributions from their randomized counterparts, we measure the z-score of the corresponding means and coefficients of variation (CV). The z-score is a measure for the extent of deviation of an empirical property x_{emp} from the mean of the randomized counterparts, $\langle x_{rand} \rangle$, scaled by the standard deviation of the randomized counterparts, viz.,

$$z = \frac{x_{emp} - \langle x_{rand} \rangle}{\sqrt{\langle x_{rand}^2 \rangle - \langle x_{rand} \rangle^2}}. \tag{2.4}$$

Surrogate ensemble for comparison with average cell body distance between connected neuronal pairs

To see whether the distance d between cell bodies of connected pairs of neurons [where the members of the pair could belong to either the same or different process length categories,

viz., short (S), medium (M) and long (L)] is distributed in a significantly different manner from that between all pairs of neurons, we have constructed surrogate ensembles. For each realization belonging to such an ensemble, a number of cell body distances is sampled from the set of all distances D between each neuronal pair, such that the sample size is same as the number of connected neural pairs. The entire ensemble consists of 10^3 such sampled sets. To see whether the observed difference between $\langle d_{XY} \rangle$ and $\langle D \rangle_{X,Y}$, where $X, Y \in \{S, M, L\}$, can be explained simply as finite size fluctuation, we have evaluated the corresponding z-scores, viz.,

$$z_{XY} = \frac{d_{XY}^{emp} - \langle d_{XY}^{rand} \rangle}{\sqrt{\langle (d_{XY}^{rand})^2 \rangle - \langle d_{XY}^{rand} \rangle^2}}. \tag{2.5}$$

Lineage tree rung determination

The order of the rung in the lineage tree that a cell belongs to is obtained from the lineage information of the cell (available from Ref. [98]). This indicates the series of cell divisions, starting from AB (which results from the division of the single cell zygote) that leads to a particular neuron, e.g., ABprpapaap. The letters a (anterior), p (posterior), l (left) and r (right) which follow AB, indicate the identity of the progenitor cells that result from subsequent cell divisions eventually terminating in a differentiated neuron. As the rung that a neuron belongs to is given by the number of cell divisions (starting from the zygote) that leads to the differentiated cell, we simply count the total number of letters (AB is counted as a single letter) specifying the lineage of a cell to determines its rung.

Stochastic branching model for lineage tree

To theoretically describe the generative process leading to the observed lineage tree for the cells belonging to the *C. elegans* somatic nervous system, we have used a stochastic asymmetric branching model. Starting from the single cell zygote, each cell division leads to at most two daughter cells, with independent probabilities $P1$ and $P2$ ($P1 \geq P2$) for the occurrence of each of the two branches. Thus, based on the probabilities $P1$ and $P2$,

at each step of the generative process any one of the following three events can happen: (i) proliferation occurs along both branches, (ii) only one branch appears (the other branch leading to either apoptosis or a non-neural cell fate), and, (iii) there is no branching so that a terminal node of the tree is obtained (i.e., the cell differentiates into a neuron). Estimation of $P1$ and $P2$ from the empirical lineage tree suggests that proliferation markedly reduces after rung 10. Incorporating this in the model by decreasing the probabilities $P1, P2$ after rung 10 results in successive reduction of the branching, eventually coming to a stop. The ensemble of 10^3 simulated lineage trees produced by the process matches fairly well with the empirical lineage tree in terms of the number of terminal nodes, the distribution of the rungs occupied by each cell and the distribution of lineage distances between the differentiated neurons (see Fig. A6).

Classifying neurons according to their role in the mesoscopic structural organization of the network

The functional importance of a neuron *vis-a-vis* its own topological module (defined above in Sec. 2.2), as well as, the entire nervous system, can be quantified in terms of its intra- and inter-modular connectivity [82]. For this purpose we use the two metrics [117]: (i) the within module degree z-score (z) and (ii) the participation coefficient (P).

In order to identify neurons that have a significantly large number of connections to the other neurons belonging to their module, we calculate the within module degree z-score defined as

$$z_i = \frac{\kappa_{c_i}^i - \langle \kappa_{c_i}^j \rangle_{j \in c_i}}{\sqrt{\langle (\kappa_{c_i}^j)^2 \rangle_{j \in c_i} - \langle \kappa_{c_i}^j \rangle_{j \in c_i}^2}}, \tag{2.6}$$

where κ_c^i is the number of connections that a neuron i has to other neurons in its community (labeled c) and the average $\langle \ldots \rangle_{j \in c}$ is taken over all nodes in the community. Following Ref. [82], we identify neurons having $z \geq 0.7$ as hubs, while the remaining are designated as non-hubs.

The neurons are also distinguished in terms of how many well connected they are to

neurons belonging to other communities. For this purpose we measure the participation coefficient P of a neuron, which is defined as

$$P_i = 1 - \sum_{c=1}^{m} (\frac{\kappa_c^i}{k_i})^2, \tag{2.7}$$

where κ_c^i, as above, is the number of connections that the neuron has to other neurons in its own module (labeled c) and $k_i = \sum_c \kappa_c^i$ is the total degree of node i. Neurons that have their connections homogeneously distributed among all modules will have a P close to 1, while $P = 0$ if all of their connections are confined within their module. Based on the value of P, following Ref. [82] we have classified the non-hub neurons as ultra-peripheral (R1: $P \leq 0.05$), peripheral (R2: $0.05 < P \leq 0.62$), satellite connectors (R3: $0.62 < P \leq 0.8$) and kin-less nodes (R4: $P > 0.8$), while hub neurons are segregated into provincial hubs (R5: $P \leq 0.3$), connector hubs (R6: $0.3 < P \leq 0.75$) and global hubs (R7: $P > 0.75$).

Logistic Regression

We have employed logistic regression [118] to assess the relative contributions of different factors in determining the probability of connection between a pair of neurons. Here, the probability of occurrence of an event is expressed as a function of one or several independent predictor variables. In our case, the event Y denotes the presence ($Y = 1$) or absence ($Y = 0$) of a connection between two neurons. The independent predictor variables correspond to the attributes: (i) process length ($X_p = 1$ if the process lengths of the two neurons are in the same class [short, medium or long], $= 0$ otherwise), (ii) birth cohort X_b ($= 1$ if both neurons are born early or both are born late, $= 0$ otherwise), (iii) symmetric pairing X_s ($= 1$ if the two neurons correspond to a symmetric pair, $= 0$ otherwise) and (iv) lineage relation X_l (= lineage distance $l \in \{1, 2, ...25\}$). Thus, logistic regression models the probability of connection in terms of the above-mentioned variables as

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_p \cdot X_p + \beta_b \cdot X_b + \beta_s \cdot X_s + \beta_l \cdot X_l)}}, \tag{2.8}$$

where the regression coefficients $\{\beta_p, \beta_b, \beta_s, \beta_l\}$ are estimated from the empirical data. These can be interpreted as change in the logarithm of odds (i.e., $log(\frac{P}{1-P})$) that results from a change in the corresponding predictor variable by a single unit, with other variables kept unchanged. Therefore, the magnitude of the regression coefficients (determined by us using the `mnrfit` function in *MATLAB Release 2010b*) provides a measure of the relative contribution of each of the factors in determining the probability of connection. We have also explicitly ensured the absence of multicollinearity (i.e., correlations) among the predictor variables, as this would otherwise result in an inaccurate estimation of the regression coefficients. For this purpose, we have used Belsley collinearity diagnostics [119] as implemented in the `collintest` function in *MATLAB Release 2016b*.

Statistics

Two-sample Kolmogorov-Smirnov (KS) test [120] has been used to compare between pairs of samples (e.g., the degrees of neurons belonging to different categories) to determine whether both of them are drawn from the same continuous distribution (null hypothesis) or if they belong to different distributions. For this purpose we have used the `kstest2` function in *MATLAB Release 2010b*, with the value of the parameter α which determines threshold significance level set to 0.05.

Kernel smoothened density function [121] has been used to estimate the probability distribution functions of different quantities (e.g., distances between cell bodies of neurons). For this purpose we have used the `ksdensity` function in *MATLAB Release 2010b* with a Gaussian kernel.

Code Availability

The codes used in this analysis can be publicly accessed from the following link : `https://github.com/anandpathak31/C_elegans_development`.

2.3 Results

2.3.1 Homophily based on multiple cellular properties governs neuronal inter-connectivity

Direct contact between neurons whose cell bodies are located relatively far apart, through synapses or gap junctions located on their extended processes, plays a crucial role in reducing communication delay of signals across the entire nervous system [122]. This is particularly relevant for *C. elegans* where the majority of synapses occur *en passant* (forming at axonal swellings) between parallel nerve process shafts that can remain close to each other over long distances [63]. Therefore, in order to understand the principles governing the wiring organization of the nematode nervous system, it is appropriate to first focus on understanding how the connectivity of neurons is influenced by the length of their neurites.

It has also been observed that connected pairs of neurons very often differentiate close to each other in time [83]. This may suggest that preferential connectivity among neurons according to the time of their birth (i.e., *birth cohort homophily*) is a possible basis for guiding the network architecture. However, we need to explore the possibility that it could be a consequence of the restrictions on connections between neurons imposed by their respective process lengths. For instance, a large majority of the neurons that are born early, i.e., prior to hatching, are localized in the head region and have short processes extending to less than a third of the body length of the nematode. This could, in principle, be sufficient to explain the temporal closeness of connected neurons. We have accordingly investigated the joint dependence of the occurrence of connections (synapses and gap-junctions) between neurons on the lengths ℓ of their respective processes, as well as, their birth times t_b in Fig. 2.1 (A-B). The distance d between the cell bodies for each pair of connected neurons is also indicated, which makes apparent the restriction on connectivity imposed by the process lengths. This information adds a temporal dimension to our understanding of the organization of long-range connections (corresponding to high values of d) in the nematode

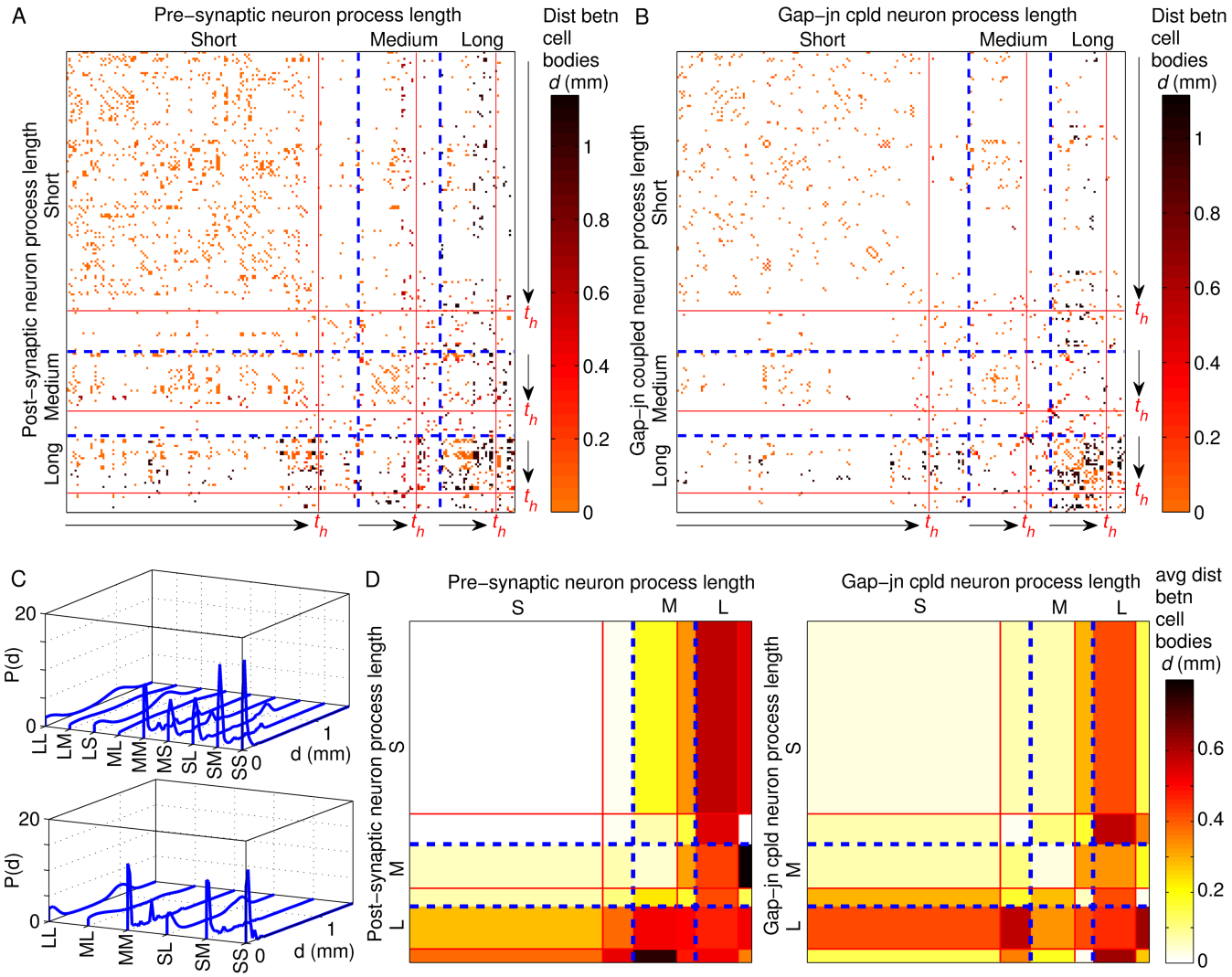


Figure 2.1: Birth time cohort membership and neurite lengths of neurons govern their connectivity. (A-B) Matrices representing synaptic (A) and gap-junctional (B) connections that exist between neurons, grouped into three classes [indicated by blue broken lines] according to their process lengths ℓ measured relative to the worm body length L, viz., short ($\ell \leq L/3$), medium ($L/3 < \ell \leq 2L/3$) and long ($\ell > 2L/3$), and ordered within each class according to birth time. Increasing birth time is indicated by arrows, with red lines marked t_h (time of hatching) separating neurons which differentiate in the embryonic stage from those born later. Matrix entries correspond to the existence of a connection, with its color representing the distance (measured in mm) between cell bodies of the corresponding neurons (see legend). We observe that there is evidence of birth time assortative mixing, with neurons born early(later) having a higher probability of connecting with other early(late) born neurons, which is particularly marked in the case of neurons having short processes. The gap junction matrix shows a large number of entries adjacent to the diagonal which correspond to connections between paired neurons [see Fig. 2.5 (A)]. (C) Distribution of distances d between cell bodies of pairs of neurons distinguished in terms of their respective process lengths (S: short, M: medium, L: long), which are connected by synapses (top) and gap junctions (bottom).

Figure 2.1 ***(previous page)*****:** As synaptic connections are directed, there are nine possible combinations of pairs of the classes (S/M/L) to which the pre- and post-synaptic neurons belong (e.g., SL refers to a synapse from a neuron with a short process to a long process length neuron). On the other hand, as gap junctions are undirected, only six possible combinations need be considered. We note the bimodal distributions of d when at least one of the two neurons connected by synapse or gap junction has a long (or medium) process. (D) The mean distance $\langle d \rangle$ between cell bodies of neurons connected by synapses (left) and gap junctions (right), grouped according to their process lengths (L/M/S) [indicated by blue broken lines] and further subdivided into those born early (i.e., embryonic stage) and those born late (i.e., L1, L2 or L3 stages) [separated by red lines]. Distances are expressed in mm (see legend for the color code). We note that pre-synaptic neurons with long processes tend to connect with post-synaptic neurons having short processes which are located far from them, corresponding to the higher peak in the bimodal distribution for LS in top panel of (C). Note that we have considered in this analysis the subset of 225 neurons for which information about process length is available.

nervous system. The neurons are grouped according to their process lengths ℓ. These are categorized as short ($\ell \leq L/3$), medium ($L/3 < \ell \leq 2L/3$) and long ($\ell > 2L/3$) relative to the total body length of the worm L. Moreover, within each category, the neurons are arranged by their time of birth in increasing order.

Process length homophily

Even a perfunctory perusal of the two matrices in Fig. 2.1 (A-B) makes it apparent that the diagonal blocks in the two matrices have relatively higher density of points. This observation indicates that there is a preponderance of connections between neurons having similar process lengths. However, to establish that there is indeed *process length homophily* which would imply an explicit preference for neurons to connect to other neurons whose neurites extend to similar distances as them, we will have to compare the empirically observed number of such connected pairs with that expected to arise by chance given the degree (i.e., the total number of connections) of each neuron. For this, we cluster the cells into three *communities* or *modules* which are characterized by all their members having short, medium or long processes, respectively. This allows us to calculate the *modularity* Q, a measure of the extent to which like prefers connecting to like in a network [115, 123] (see Methods for details). A positive value of Q for a particular module would suggest that

there is a bias for its members to preferentially connect to each other, while $Q \sim 0$ indicates the absence of any evidence for homophily. For the entire synaptic network, we measure Q to be 0.125, while for the network of neurons connected by gap junctions, it is 0.18. We find these empirical Q values to be significantly higher than the corresponding values, viz., -0.003 ± 0.008 and -0.003 ± 0.013, calculated for ensembles of randomized surrogates for the synaptic and gap-junctional networks, respectively, obtained by randomly permuting the process length category membership of each neuron (see Methods). This suggests that neurons having similar process lengths do indeed have many more of their connections with each other than would be expected simply on the basis of the number of synapses and gap junctions possessed by each of them. We have additionally considered another surrogate ensemble obtained by randomly permuting the connections of the network keeping the degree of each neuron unchanged, subject to constraints imposed by the neuronal process lengths given the spatial positions of the cell bodies (see Methods). With respect to this ensemble also the empirical Q values are seen to be significantly higher (see Appendix A, Table A1). Thus, although the empirical values of the modularity appear to be small, they cannot be attributed simply to noise and suggests the existence of specific mechanisms that make connections between two neurons, both of which have short (or long) processes, more likely. Moreover, individually considering the three communities comprising neurons having short, medium and long processes, respectively, yields class-specific Q values (see Methods) which are also significantly higher than the Q values obtained from the corresponding randomized surrogates. In contrast, the class-specific Q values obtained for connections between neurons belonging to different categories of process lengths are either lower or about the same as the Q values obtained from the corresponding randomized surrogates, as indicated by the respective z-scores (see Appendix A, Table A1). This further underlines the existence of a significant bias for connections to occur between neurons having similar process lengths.

Birth cohort homophily

We observe a relatively high density of points in Fig. 2.1 (A and B) in the blocks corresponding to connections between cells having short processes (i.e., SS) that are *born at the same epoch*, i.e., either pre- or post-hatching. This is also seen for connections between neurons having medium length processes (MM), as well as, those between neurons having short and other neurons having medium process lengths (MS or SM). This suggests that, apart from process length, the time of birth of the cells also determine neuronal interconnectivity. Indeed, earlier studies [83] have shown that most of the neurons that are connected to each other happen to be born close in time, with the probability of connection between contemporaneous neurons being much more than what is expected by chance. However, we find that the actual temporal separation between the time of birth of different neurons does not have any significant correlation (viz., $p \gg 0.05$) with the probability of there being a connection between them, either synaptic or gap-junctional. This apparent contradiction is resolved on noting the following. While, within the group of neurons born in the embryonic stage and those born post-embryonic there may be a great diversity in terms of birth times (thereby significantly weakening any correlation with connection probability), these differences are minor when viewed from the perspective of membership in the cohorts of those born pre- and post-hatching, respectively. As reported earlier [83], these correspond to two distinct, temporally separated bursts of neuronal differentiation, which provides a natural demarcation of the neurons into early and late-born categories.

We observe that neurons prefer to connect to other members of their cohort (viz., early or later-born). This is indicative of *birth cohort homophily*, which is quantitatively established by segregating the neurons born pre- and post-hatching into two communities and then calculating Q values. As shown in Appendix A, Table A2, the Q for synapses is 0.09 and that for gap junctions is 0.07, which are significantly higher than the corresponding values obtained from the randomized surrogate ensemble (obtained by keeping the degree sequence unchanged, subject to constraints imposed by process lengths given the spatial position of cell bodies), viz., 0.02 ± 0.005 for synaptic and 0.03 ± 0.01, respectively.

We note that this homophily is restricted to neurons whose cell bodies are located in close physical proximity (see Appendix A, Figure A1). By comparing with randomized surrogates, we observe that connections between neurons are not significantly enhanced if they are born in the same epoch except for the case when the distance d between their cell bodies is short ($d < L/3$).

Process lengths affect the spatial arrangement of neurons

So far, in our consideration of how connections between neurons is affected by their process lengths, we have not considered the information concerning the spatial position of the cell bodies of the connected neurons. Consideration of this information is important if we want to understand how activity of spatially distant parts of the organism are coordinated through long-range connections that allow signals to be rapidly transmitted across relatively large physical distances. Figs. 2.1 (C) and (D) show how the distance d between cell bodies of connected pairs of neurons are distributed differently according to their respective process lengths.

The top panel in Fig. 2.1 (C) corresponds to the probability distribution function of distance between cell bodies d for neurons connected by synapses, while the bottom panel considers gap junctions. When both the pre- and post-synaptic neurons have short processes (indicated by SS in the figure), it is expected that the cell bodies will be located close to each other. This is indeed what is observed, with a prominent peak of $P(d)$ occurring at extremely low values of d. On the other hand, when at least one of the neurons has a long or medium length process, we observe that the distributions for neurons connected through synapse are much more extended towards higher values. For SL, LS and LL connections, we in fact observed a distinct bimodal character in the corresponding distribution of d. This can be linked to the observation that neurons having short as well as long processes tend to predominantly have their cell bodies located at the head or in the tail of the worm. In contrast, neurons whose processes are intermediate in length have cell bodies distributed relatively more homogeneously across the body of the organism (see

Appendix A, Figure A2). This can be quantified by measuring the extent to which the cell bodies themselves are distributed along the longitudinal axis of the nematode body in a bimodal manner using the Bimodality Coefficient (BC) metric [116] (see Methods). A distribution is said to be prominently bimodal if its $BC \gg BC^*$ ($= 5/9$), the value of the metric for an uniform distribution. We find that while the spatial positions of the cell bodies of neurons having short, as well as, long process are distributed in a bimodal manner ($BC_S = 0.93$ and $BC_L = 0.83$, respectively), that of neurons with intermediate length process ($BC_M = 0.67$) are relatively more uniformly distributed. Accordingly, we observe that synaptically connected pairs, in which at least one neuron has process of medium length, exhibit distributions of d where bimodality is either muted (as in SM, MM and MS) or absent (ML and LM), even though all of these distributions span a much larger range of d than SS. This indicates that process length is an important determinative factor for the occurrence of long-range connections in the nematode nervous system.

When we consider the distribution of distances between cell bodies of neurons connected by gap junctions [lower panel of Fig. 2.1 (C)], we observe that connections are more likely to occur between spatially adjacent cell bodies. This is manifest in the distributions of d being much less extended than those seen in the case of synapses, with the exception of SL and LL which exhibit bimodality. The distinction between the situations seen in the upper and lower panels may arise from the fact that while synapses between two neurons can in principle be located anywhere on their processes, gap junctions predominantly occur close to the cell body of at least one of the participating neurons.

The detailed nature of the information about the number of neuronal pairs with given process lengths whose cell bodies are placed a specific distance d apart that is provided by the distributions shown in Fig. 2.1 (C) tends to obscure certain gross features. The latter can impart important insights into how process length facilitates connections between spatially distal neurons. Therefore, in Fig. 2.1 (D) we display the average physical distance between cell bodies of *connected* neurons which are distinguished in terms of their process lengths (short/medium/long), and further subdivided into those appearing in the embryonic stage, i.e., prior to hatching (referred to as early), and those which appear at

the post-embryonic stage (referred to as late). For synaptic connections (shown at left), the average d for neurons with long processes (pre-synaptic) connected to neurons having short processes (post-synaptic) is the highest ($\langle d_{LS} \rangle = 0.57$ mm) of all the categories considered, higher even than that when both neurons in a connected pair have long processes ($\langle d_{LL} \rangle = 0.50$ mm). Intriguingly, both of these values are larger than the average distance between cell bodies for connected neurons when the pre-synaptic neurons have short processes while the post-synaptic ones have long processes, viz, ($\langle d_{SL} \rangle = 0.33$ mm). This is consistent with the two peaks of the bimodal distribution of d corresponding to these connections differing substantially in amplitude - the peak at lower d being higher for SL, while the one at higher d being larger for LS. To a lesser extent, a similar asymmetry is seen for the average distance between connected cell bodies when one has short process while the process of the other is of medium length (viz., $\langle d_{MS} \rangle = 0.22$ mm as compared to $\langle d_{SM} \rangle = 0.09$ mm).

We can compare these values with the average distance between cell bodies of *all* neurons, whether connected or not. For instance, the mean separation D between cell bodies of all neurons with long process lengths is $\langle D \rangle_{L,L} = 0.55$ mm which is almost the same as the average distance between every pair of neurons in which one has a short process and the other has a long one ($\langle D \rangle_{L,S} = 0.54$ mm). To ensure that the difference between $\langle d_{XY} \rangle$ and $\langle D \rangle_{X,Y}$ (where $X, Y \in \{S, L, M\}$) is statistically significant, we show that it is extremely unlikely that the observed values of d will arise by chance if random surrogates are constructed having the same number of connected neurons as is observed empirically (by sampling the set of all neuronal pairs without replacement). For instance, the z-score (see Methods) for the distance between cell bodies of pre-synaptic neurons with long processes connected to post-synaptic neurons with short processes is $z_{LS} = 1$. By contrast, considering the reverse, i.e., synapses from neurons with short processes to those having long processes, we obtain $z_{SL} = -7.8$. Thus, neurons with long processes appear to form a synapse with neurons having short processes whose cell bodies are located far away from their own much more often than that expected by chance given the spatial positions of the cell bodies. On the other hand, neurons with short processes prefer to connect to neurons

with long processes whose cell bodies are much closer to their own. Indeed, excepting the class of LS and ML synaptically connected neuron pairs (i.e., pre-synaptic neurons with long process with post-synaptic neuron with short process and pre-synaptic neurons with medium process with post-synaptic neuron with long process), all other connected neural pair classes, distinguished in terms of the process lengths of the two neurons, have negative values for z-score (see Appendix A, Table A3, and Figures A3,A4). The results indicate that the process length of the pre-synaptic neuron is a dominant influence deciding the average distance between cell bodies connected by synapses. It is also consistent with the possibility that a high proportion of synaptic contacts are occurring close to the cell body of the post-synaptic neuron (which is closer to the classical concept of the pre-synaptic axon connecting to a dendrite close to cell body of the post-synaptic neuron and not just making a synaptic contact anywhere on the process). Such asymmetry between LS and SL may also have the advantage of functional efficiency in that the resulting connection architecture allows signals to rapidly travel large distances across the nematode body through long processes - thereby spreading globally using L to S connections - and then being disseminated locally using neurons with short processes.

If we now consider the case of neurons connected by gap-junctions [Fig. 2.1 (D, right)], we note that the average value of d is highest for the case of cells with long processes connecting to each other. In particular, unlike the situation seen above for synaptically connected neurons, $\langle d_{LS} \rangle (= 0.44$ mm) is lower than $\langle d_{LL} \rangle (= 0.5$ mm). The z-score for the distance between cell bodies of neuron pairs whose members belong to any of the classes S,M and L are seen to be strongly negative, ranging between $z_{LL} = -1.6$ and $z_{SS} = -12.7$. The high statistical significance of $\langle d \rangle$ when compared against the average separation between neurons $\langle D \rangle$ suggests that gap junctions occur between neurons whose cell bodies lie close to each other far more often than expected by chance (given their positions). This is consistent with the belief that gap junctions predominantly act to coordinate activity locally between neurons [124]. We also note in passing another feature of gap junctional connections between neurons which is manifest in Fig. 2.1 (B) as a large number of entries in the adjacency matrix immediately neighboring the diagonal. These correspond to a very

high proportion of connections between bilaterally symmetric pair of neurons, e.g., AVAL and AVAR, that is discussed later (see under bilateral symmetric pairing homophily). These connections may have the possible functional goal of coordinating response of the nematode nervous system to sensory inputs between the left and right sides of the body [125].

The process length homophily between neurons that we demonstrated above can be attributed to multiple possible factors. For instance, the preference of neurons having long process for connecting to other neurons with long processes could be an outcome of the geometry resulting from parallel fibers extending over relatively large distances, which have a proportionately higher probability of forming *en passant* synapses with each other. On the other hand, the preference of neurons having short processes to connect to each other could be tied to the fact that many of their cell bodies are located in close physical proximity. This suggests an important role for the physical distance d between cell bodies in deciding connectivity between neurons. When we look at the correlations between d and the probability that the cells are connected, we do not find any statistically significant correlation for either synapse or gap junctions. Focusing only on neuron pairs whose cell bodies are located close to each other (i.e., $d \leq L/3$ where L is the total body length of the worm), however, we observe a very strong correlation of -0.92 ($p = 0.003$) between d and the probability of a synaptic connection between the two (for gap junctions, the correlation is -0.89 with $p = 0.007$). This high value indicates that synapse formation between neurons whose cell bodies are located near each other is indeed strongly dependent on the distance between them. Moreover, it cannot be explained in terms of simple physical limits imposed by the process lengths of neurons on the farthest distance allowed between cell bodies of connected neurons. This is because if we consider the correlation between d and probability of connection only between neurons having short processes, we obtain a value of -0.87 ($p = 0.012$) for synapses and -0.89 ($p = 0.008$) for gap junctions (see Appendix A, Figure A5).

A possible explanation for the weakening of the relation between connection probability and the physical distance separating the cell bodies when all neurons are considered could be because, even though neurons born in close physical proximity have a higher probability

of getting connected, it is masked by the cells moving apart subsequently over the course of development. In the absence of information about the location of the cell bodies at the time synaptogenesis happens, we can probe this indirectly by considering how the probability of connection between two cells depends on how closely they are related in terms of lineage - as cells having common ancestry also tend to be born adjacent to each other.

Lineage homophily

Cell lineage provides knowledge of the developmental trajectory in all metazoa, being defined by successive divisions starting from the zygote to the final differentiated cell. In most animals, the identity of any terminal node of the lineage tree, known as cell fate, is determined by intrinsic and extrinsic factors, as well as, interactions with neighboring cells. This introduces sufficient variability in the developmental path so as to make lineage relationships discernible only at the level of cell groups rather than individual cells [101]. However, some organisms such as nematodes exhibit an almost invariant pattern of somatic cell divisions that is identical across individuals, and in the case of *Caenorhabditis elegans*, is known in its entirety [66, 67]. Thus, the lineage tree of the organism provides us with a complete fate map at single-cell resolution [126]. The schematic representation of such a tree shown in Fig. 2.2 (A) depicts successive mitotic cell divisions starting from a zygote that, through intermediate progenitor cells, eventually differentiate into mature neuronal cells. Each successive cell division (beginning from the zygote) corresponds to different rungs in the tree used to label the resulting daughter cells. The difference between any two cells in terms of their lineage can thus be quantified by their lineage distance, i.e., their separation on the tree measured as the total number of cell divisions that leads to each of them from their last common progenitor.

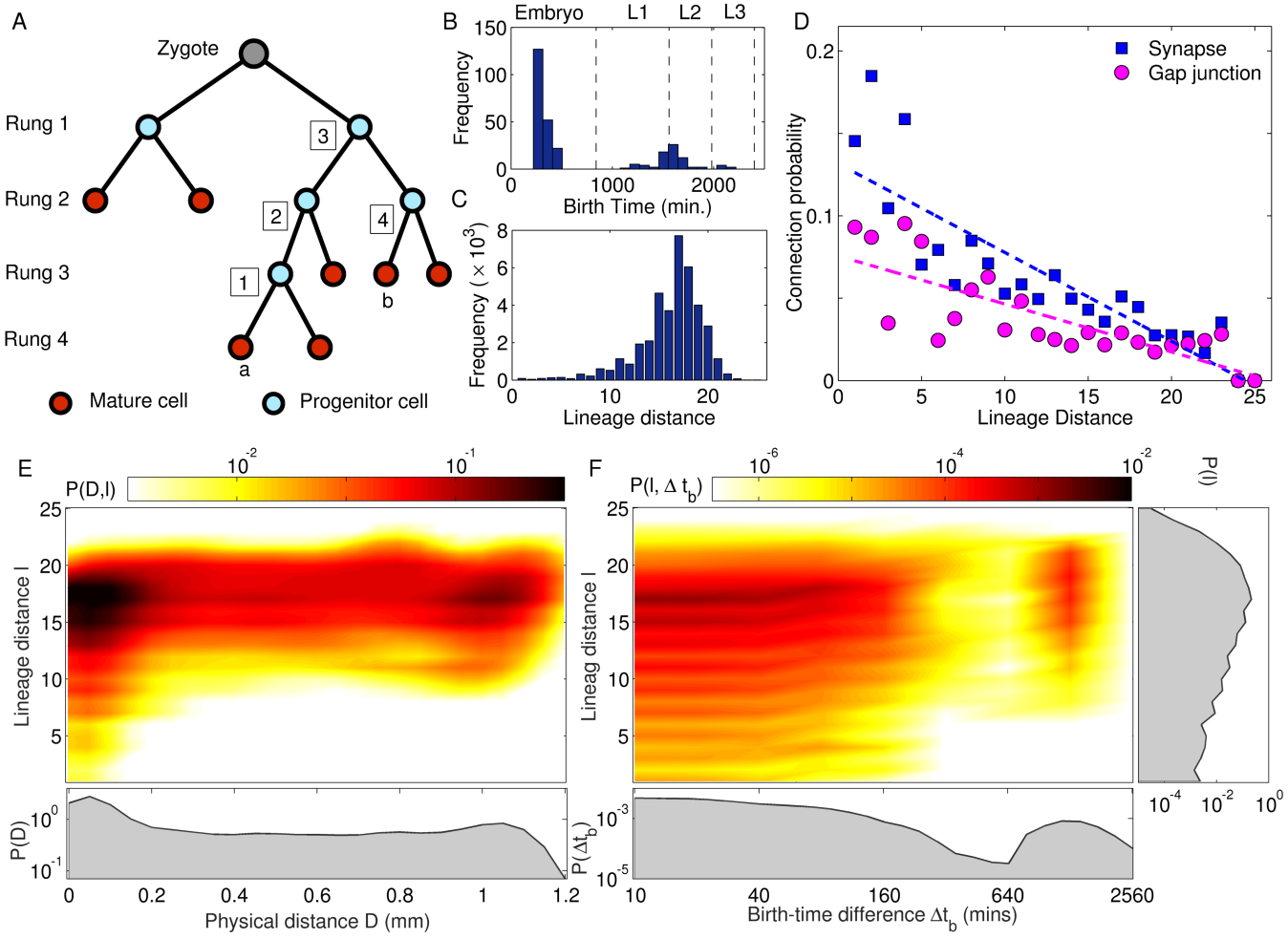


Figure 2.2: Lineage of neurons affects their synaptic connectivity and spatial localization. (A) Schematic diagram of a lineage tree of cells resulting from consecutive mitotic divisions of the zygote. The terminal nodes of the tree correspond to terminally differentiated mature cells (shown in red) while other nodes represent progenitors (shown in blue) that appear at different rounds of cell division. Cells born at each round of cell division are indicated by the corresponding rung of the tree they belong to, the numerical value for the rung (shown at the left) being the number of divisions starting from the zygote. The lineage distance l between a pair of mature cells is measured as the total number of cell divisions leading to each from their common progenitor. An example of lineage distance measurement is shown in the figure for the pair of cells a and b which are separated by four cell divisions (the distance of a from each of the intermediate dividing progenitors is indicated in the figure). (B-C) Frequency distributions of the birth time of different neurons (B, separated into the different developmental stages) and the lineage distances for each pair of neurons (C). (D) The probability of a pair of neurons to be connected through a synapse decreases with increasing lineage distance between them, as indicated by a statistically significant linear correlation between the two ($r = -0.87$, $p < 10^{-7}$). For gap junctional connections, the correlation is marginally weaker ($r = -0.79$, $p < 10^{-5}$). (E-F) Joint probability distributions of lineage distance l along with distance between cell bodies D (E) and birth time difference Δt_b (F) between all pairs of neurons.

Figure 2.2 ***(previous page)*:** The marginal distributions for the corresponding quantities are indicated in the bounding surfaces. Contours for the distributions are indicated at the base of each figure. We notice that the distribution of physical distances in (E) exhibit a bimodal nature. However, cells which are closely related in terms of lineage ($l < 5$) also has a high probability of being physically located nearby (indicated by a prominent peak at the lower end of the distribution of D) which suggests that lineage influences spatial localization of cells. In panel (F), the distribution shows peaks at odd values of the lineage distance (particularly for low Δt_b) suggesting that neurons born close in time are located at the same rung on the lineage tree.

Apart from the lineage tree, crucial information on the relationships between different cells that stem from their developmental history is provided by the knowledge of birth times of the individual mature neurons, i.e., the specific instant in developmental chronology of the nervous system at which each neuron differentiates. Fig. 2.2 (B) shows the distribution of birth times for all cells belonging to the somatic nervous system of *C. elegans*, indicating that development of the system occurs in two bursts clearly separated in time [83]. The 'early burst', during which the bulk, viz., 72%, of the neurons are born, occurs at the embryonic stage of development, while the more temporally extended 'late burst' spans across the L1 and L2 stages. This information, in conjunction with a simple generative model for reconstructing the lineage tree through successive cell divisions, can be used to explain the distribution of lineage distance shown in Fig. 2.2 (C). As at each node of the lineage tree a cell divides into at most two daughter cells, we can view it - at least in the first few rungs belonging to the early proliferative phase - as a balanced binary tree, with the number of cells that appear in each rung R increasing exponentially with R (upto $R = 10$ in *C. elegans*, see Appendix A, Figure A6. Within the AB sub-lineage of cells to which almost all the neurons belong, the maximum lineage distance that can occur between two cells which are placed in rungs R_1 and R_2, respectively, is given by $l_{max}(R_1, R_2) = (R_1 - 1) + (R_2 - 1) - 1$. Thus, the distribution of lineage distances has an exponential profile upto $l = 17$. Beyond rung 10, the subsequent branching of the nodes in the binary tree reduce markedly as many of the divisions terminate in differentiated neurons (and occasionally programmed cell death) or lead to non-neuronal fates (so that their further divisions are not considered for the purpose of this study). This can be seen to result in the lineage

distance distribution *decreasing* exponentially for $l > 17$, with a maximum lineage distance of 25. A more detailed theoretical model of the lineage relationships between neurons resulting from their developmental history can be constructed as an asymmetric stochastic branching process (see Methods). Here, beginning with a single node that corresponds to the zygote, at each iteration every node that appeared during the preceding iteration is considered in turn for giving rise to each of two possible branches with probabilities $P1$ and $P2$ ($P1 \geq P2$) that result in further nodes. By considering the actual lineage tree, these asymmetric branching probabilities in the model were fixed as $P1 = 1$ and $P2 = 0.85$ until rung 9 and for later rungs they were set to $P1 = 0.25$ and $P2 = 0.2$. For these values of $P1$ and $P2$, the trees generated by the model exhibited properties that were statistically similar to the empirical lineage tree (see Appendix A, Figure A6).

Going back to the question we had posed earlier, viz., how does the lineage distance l between cells affect the probability that they are connected by synapses, we observe from Fig. 2.2 (D) that there is indeed a strong correlation of -0.87 ($p < 10^{-7}$) between the two. For gap-junctions, we again observe a correlation between lineage distance and connection probability that is only marginally weaker, viz., -0.79 ($p < 10^{-5}$), than that seen for synapses. This observation provides evidence of *lineage homophily* being one of the key principles governing connectivity of the nematode nervous system. The linear fits for the dependence of the connection probabilities on lineage distance (shown using broken lines) imply that synaptic connection probability has a slightly stronger dependence on the lineage distance compared to gap junctions, as indicated by the higher slope of the regression line for the former. These observations suggest that changes in the locations of cell bodies from that they occupied initially (i.e., at the time the corresponding neurons differentiated) which are brought about by the appearance of cells born later through subsequent cell-divisions, result in a weak correlation between connection probability and physical distance separating the cell bodies, as alluded to earlier.

Lineage relation between neurons constrains distance between their cell bodies

The connection between lineage distance l and physical distance D between cell bodies of neurons (whether connected or not), which has been mentioned earlier, is illustrated by the joint probability distribution $P(D, l)$ shown in Fig. 2.2 (E). In particular, cells having short lineage distance, viz., $l \leq 5$, tend to have their cell bodies located close to each other, as indicated by the function being peaked towards lower values of D. However, cells that are farther apart in terms of lineage can occur at different distances from each other, resulting in the overall bimodal form for the marginal distribution of D. A similar nuanced relation between lineage distance for two neurons and the difference of the times Δt_b in which they are born is indicated by the joint probability distribution $P(l, \Delta t_b)$ shown in Fig. 2.2 (F). We note that for small l ($l \leq 5$), the distribution peaks at low values of Δt_b indicating that closely related neurons tend to be born within a short time interval of each other. We also observe that the distribution of l between neurons that differentiate at around the same time (i.e., for low Δt_b) tends to alternate between peaks and troughs for odd and even values, respectively. This is easy to explain if neurons that are contemporaneous occur at the same rung (as, by definition, neurons at the same rung will have odd values of lineage distance between themselves).

The different ganglia comprise clusters of closely related neurons

The compelling association between lineage and physical proximity of neurons alluded to above is manifest in the spatial organization of the cell body locations. It is particularly conspicuous in the clustering of neurons into anatomically distinct bundles that are referred to as ganglia. These structures, characteristic of nematode nervous systems, contain only cell bodies of the neurons with their axonal and dendritic processes located outside of the bundles [127]. The somatic nervous system comprises nine such spatially localized clusters, viz., anterior, dorsal, lateral, ventral, retrovesicular, posterolateral, preanal, dorsorectal and lumbar ganglia, with the remainder belonging to the ventral cord. Comparison of the distributions of intra-ganglionic lineage distances (i.e., between pairs of neurons located in

the same ganglion) with that of inter-ganglionic lineage distances (i.e., between neurons in different ganglia) provides an insight into how these bundles can be interpreted from a developmental perspective.

We first note that the mean of the lineage distances $\langle l \rangle$ within a given ganglia are typically much smaller than those between different ganglia. Moreover, as seen from Fig. 2.3 (A), the mean of the intra-ganglionic lineage distances for most ganglia are significantly small, which we determine by comparing with values of $\langle l \rangle$ obtained from ensembles of 10^3 surrogate lineage trees where the identity of each of the leaf nodes (i.e., the differentiated neurons) has been randomly permuted. This randomization decouples the ganglionic membership of the neurons from their position on the lineage tree while keeping the lineage distances between cells invariant, consistent with our null hypothesis that the ganglion to which a neuron belongs is independent of its developmental history. The observed mean intra-ganglionic lineage distances deviate markedly from those obtained from the surrogate trees (as measured by z-score, see Methods), indicating that neurons in a ganglion are much more closely related to each other than expected by chance. We note however that, the Posterolateral (G6) and Lumbar (G9) ganglia are exceptions to the rule, in that they do not exhibit a significantly negative z-score like the other ganglia.

However, when we consider the coefficient of variation (CV), a relative measure of the dispersion in the lineage distances within a ganglion or between two ganglia, we note that this is almost always greater for intra-ganglionic, compared to the inter-ganglionic, lineage distances [Fig. 2.3 (B)]. We can again establish the statistical significance by measuring the same quantities for the ensemble of surrogate lineage trees mentioned above and quantifying the difference between the actual tree and the randomized ensemble using z-scores. The large values of z for CV in most of the diagonal blocks (corresponding to intra-ganglion dispersion) shown in Fig. 2.3 (B), suggests that the relatedness between neurons in a ganglion shows a much larger variability than expected by chance.

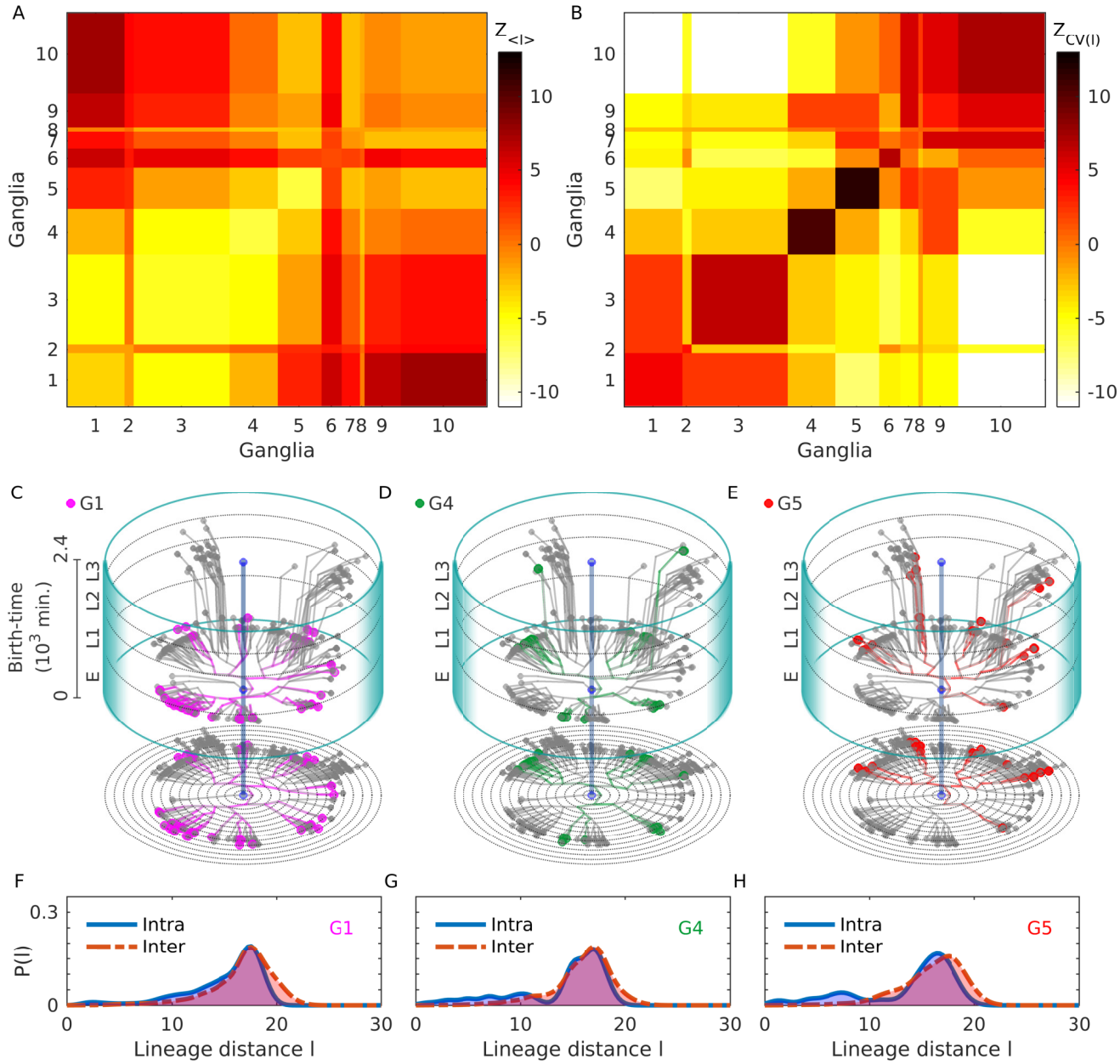


Figure 2.3: Lineage distance reveals developmental patterns of ganglia. (A-B) Statistically significant features of the distribution of intra and inter-ganglionic lineage distances, quantified by deviations of the mean $\langle l \rangle$ (A) and coefficient of variation CV (B), from a surrogate ensemble of randomized lineage trees of neurons in the C. elegans somatic nervous system. These deviations (measured by z-score) show that the mean intra-ganglionic lineage distances (represented by diagonal blocks of the matrix) are significantly lower than that of the inter-ganglionic lineage distances (off-diagonal blocks), with the exception of G6 and G9. By contrast, CV for the intra-ganglionic lineage distances are significantly higher than that of the inter-ganglionic lineage distances. (C-E) Developmental chrono-dendrograms for three representative ganglia (viz., G1, G4 and G5) show that each comprises multiple localized clusters of neurons occurring at different locations on the developmental lineage tree, explaining the statistically significant deviations

Figure 2.3 ***(previous page)*****:** of the mean and CV for intra-ganglionic lineage distances. Colored nodes represent neurons belonging to the specified ganglion while gray nodes show the other neurons. Branching lines trace all cell divisions starting from the single cell zygote (located at the origin) and terminating at each differentiated neuron. The time and rung of each cell division is indicated by its position along the vertical and radial axis respectively. The entire time period is divided into four stages, viz., Embryo (indicated as E), L1, L2 and L3. A planar projection at the base of each cylinder shows the rung (concentric circles) of each progenitor cell and differentiated neuron. (F-H) The probability distribution functions for the intra-ganglionic lineage distances show bimodality (unlike that of the inter-ganglionic distances), which is consistent with the segregation of a ganglion into multiple clusters along the chrono-dendrogram. The different ganglia are indicated by symbols G1-G9 (1: Anterior, 2: Dorsal, 3: Lateral, 4: Ventral, 5: Retrovesicular, 6: Posterolateral, 7: Preanal, 8: Dorsorectal and 9: Lumbar) and the Ventral cord as G10.

The apparent contradiction between the results mentioned above, viz., that a majority of the neurons in a ganglion have a shared lineage while, at the same time, exhibit a high degree of diversity in their lineage relations, is easily resolved on inspecting the chrono-dendrograms that visually represent the complete developmental trajectory for each of the ganglia [shown in Fig. 2.3 (C-E), for the anterior, ventral and retrovesicular ganglia; see Appendix A, Figures A7-A9 for the others]. While the lineage tree shown in each of these figures is, of course, identical, the neurons that belong to a particular ganglion are distinguished (by color) in the corresponding chrono-dendrogram, allowing us to note at a glance how all the members of the given ganglion relate to each other. We note that the differentiated neurons that constitute a ganglion are typically organized into multiple clusters, each of which are highly localized on the lineage tree. In other words, a ganglion comprises several 'families' of neurons emanating from different branches of the tree, with each family composed of closely related cells sharing a last common ancestor separated from them by only a few cell divisions.

The grouping of the cells belonging to a particular ganglion into distinct clusters, which are widely separated on the lineage tree, is reflected in the bimodal nature of the distribution of intra-ganglionic lineage distances [Fig. 2.3 (F-H)]. In contrast to the unimodal distribution seen for inter-ganglionic lineage distances, the neurons within a ganglion could either have (i) extremely low distances to cells which belong to their own 'family' or (ii)

large distances to cells belonging to the other 'families' that constitute the ganglion. These manifest, respectively, as a smaller peak at lower values and a larger peak at higher values of l seen in Fig. 2.3 (F-H). The bimodality gives rise to a large dispersion and hence a value for the CV of lineage distances that is higher than expected. Note that the peak at higher l for this distribution almost coincides with the peak of the inter-ganglionic l distribution, which is expected as the latter is dominated by cells that are not closely related. Thus, the presence of the second peak at lower values of l in the intra-ganglionic distribution reduces the mean lineage distance for cells within a ganglion, compared to that for cells belonging to different ganglia. Conversely, the absence of multiple peaks in the inter-ganglionic distribution provides for a smaller value of the CV compared to the case for the intra-ganglionic distribution. Thus, these results explain the apparently contradictory coexistence of low mean value and high CV for lineage distances of neurons within a ganglion, which is related to the localization of the developmental trajectories of cells belonging to it into distinct groups visible in the lineage tree. This clearly demonstrates that the spatial segregation of neurons into ganglia is shaped by the relations between the constituent cells which arise from their shared developmental history.

Birth time and lineage relation together constrain the physical distance between cell bodies of connected neurons

Having considered the distribution of physical distance, lineage distance and birth-time differences between all neuronal pairs in the somatic nervous system, we now focus on the subset of connected pairs to see how the above factors may constrain the probability that a neuron has a direct interaction with another. Fig. 2.4 shows the inter-relations between similarity of ancestry, spatial separation and birth times for each pair of neurons that are linked either by synapses (top row) or gap junctions (bottom row). The clustering of mean birth times of the connected pairs into three distinct groups (seen in panels A-B and E-F) is a consequence of the two bursts of neuronal differentiation widely separated in time [seen in Fig. 2.2 (B)]. Thus, the lower and upper clusters correspond to connected

neurons both of which appear in the course of the same developmental burst (early and late, respectively), while connections between neurons that arose during different bursts populate the intermediate cluster.

In Fig. 2.2 (F) we had already seen that closely related neurons tend to have similar birth times. This helps explain why, as seen in Fig. 2.4 (B), whenever synaptically connected neurons have short lineage distance to each other, they also happen to belong to the same developmental burst epoch. However, apart from the relative differences in the birth times, the actual time of differentiation also determines the occurrence of a synapse between neurons. Indeed, it is known from Ref. [83] that about 68% of long-range synaptic connections occur between neurons both of which are born in the early burst of neuronal differentiation. This is complemented by Fig. 2.4 (A) which shows that synapses between neurons, whose cell bodies are separated by large distances, mostly occur when at least one of the neurons was born early. Conversely, when both neurons are born in the late burst, such long-range links become extremely unlikely. Indeed, the distribution of distances between cell bodies of connected neurons (see Appendix A, Figure A10, that compares the empirical data with degree-preserved randomized networks where the connections are made according to constraints imposed by the length of processes of each neuron) show that long-range connections in the nematode typically do not occur significantly more often than that expected by chance, given the process lengths of the neurons. Thus, specific mechanisms for explaining the occurrence of such connections maybe unnecessary given that *en passant* synaptic contacts form between neighboring parallel neuronal processes. In contrast, short range connections are much more numerous than that seen in the random surrogate networks. This suggests that active processes may be driving synaptogenesis [71, 72] between neurons lying in close proximity, for example, chemoattractant diffusion [128, 58, 60]. Furthermore, the exceptional feature of early pre-synaptic neurons having long-range connections to late post-synaptic neurons much more often than is expected by chance could suggest a possible role of fasciculation in this process [48]. For instance, late-born neurons could be following the extended processes of earlier neurons to connect to cell bodies placed far away.

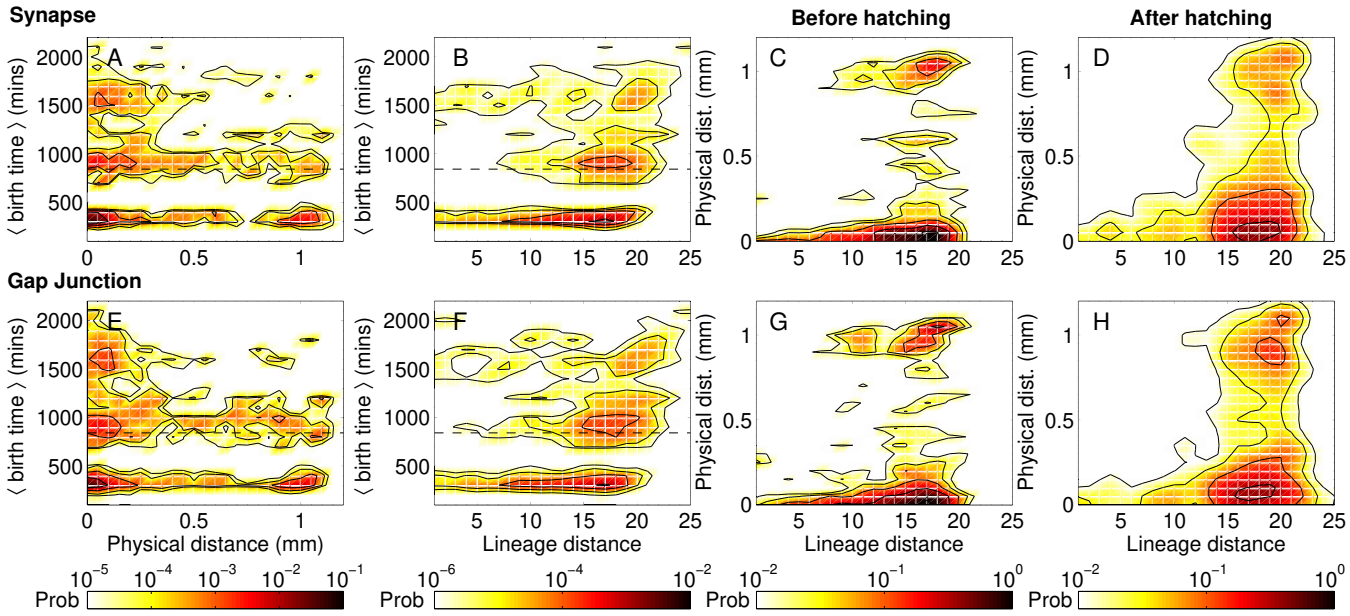


Figure 2.4: Birth times and lineage distances constrain connections between neurons whose cell bodies are spatially distant from each other. (A-B) The mean birth time of synaptically connected pairs of neurons exhibit a trimodal distribution, with connections clustering into three temporal groups corresponding to those (i) between neurons that are both born early, i.e., in the embryonic stage, (ii) between one born early and the other born late (i.e., in the post-embryonic stage), and (iii) between neurons that are both born late. The hatching time h_t separating the embryonic from other developmental stages is indicated by the broken line. We note from panel (A) that when both neurons are born late (corresponding to the uppermost cluster of connections), synaptic connections are more likely to occur between neurons whose cell bodies are located close to each other. (C-D) Synaptic connections between neurons that are closely related to each other in terms of lineage ($l < 10$) occur almost always when their cell bodies are in proximity, regardless of the time of birth of the neurons. We note that this restriction is more pronounced than observed in Fig. 2.2 (E), where $P(D, l)$ shows a prominent peak at the lower end of D for small l suggesting that most closely related neurons (whether connected or not) typically have short distances between their cell bodies. (E-H) Neurons connected by gap junctions show patterns similar to those seen in the case of synaptic connections.

In Fig. 2.4 (C) and (D) we compare explicitly the pre- and post-hatching scenarios in order to see whether early and late-born neurons differ in terms of how the synaptic connections between them are influenced by the lineage and/or physical distances between them. We note that for both groups of cells, closely related neurons that are connected by synapse also happen to occur at spatially proximate locations. This is consistent with Fig. 2.2 (E) where the peak in the joint probability distribution of all neuronal pairs with lineage distance l and physical distance D is observed to occur at low D when l is small. Qualitatively similar results are observed when we consider neuronal pairs connected by gap junctions [see panels E-H of Fig. 2.4].

The results reported above provide remarkable evidence for the role that developmental attributes (viz., lineage distances and birth-times of neurons) play in shaping the spatial organization of cell bodies and the topological structure of the connections in the somatic nervous system of the worm. However, the process length homophily described earlier appears to be independent and cannot be explained as a consequence of lineage homophily. The chrono-dendrograms (see Appendix A, Figure A11) showing the positions of neurons with short, medium and long processes, respectively, on the lineage tree indicate that neurons having a particular process length do not cluster together. This suggests that neurons with extremely similar lineage may have very different process lengths (and vice versa), so that the observed bias in the connection probability between neurons having processes of similar length cannot simply be attributed to a common lineage.

Bilateral symmetric pairing homophily

The major fraction ($\approx 66\%$) of neurons belonging to the somatic nervous system of *C. elegans* occur in pairs. These are located along the left and right sides of the body of the nematode in a bilaterally symmetric fashion. While there are instances of bilaterally symmetric neurons exhibiting functional lateralization (e.g., ASEL/R, see Ref. [129]), the vast majority of the left/right members of such pairs remain in the symmetrical "ground state", i.e., they are indistinguishable functionally, as well as, in terms of anatomical features and

gene expression [130].

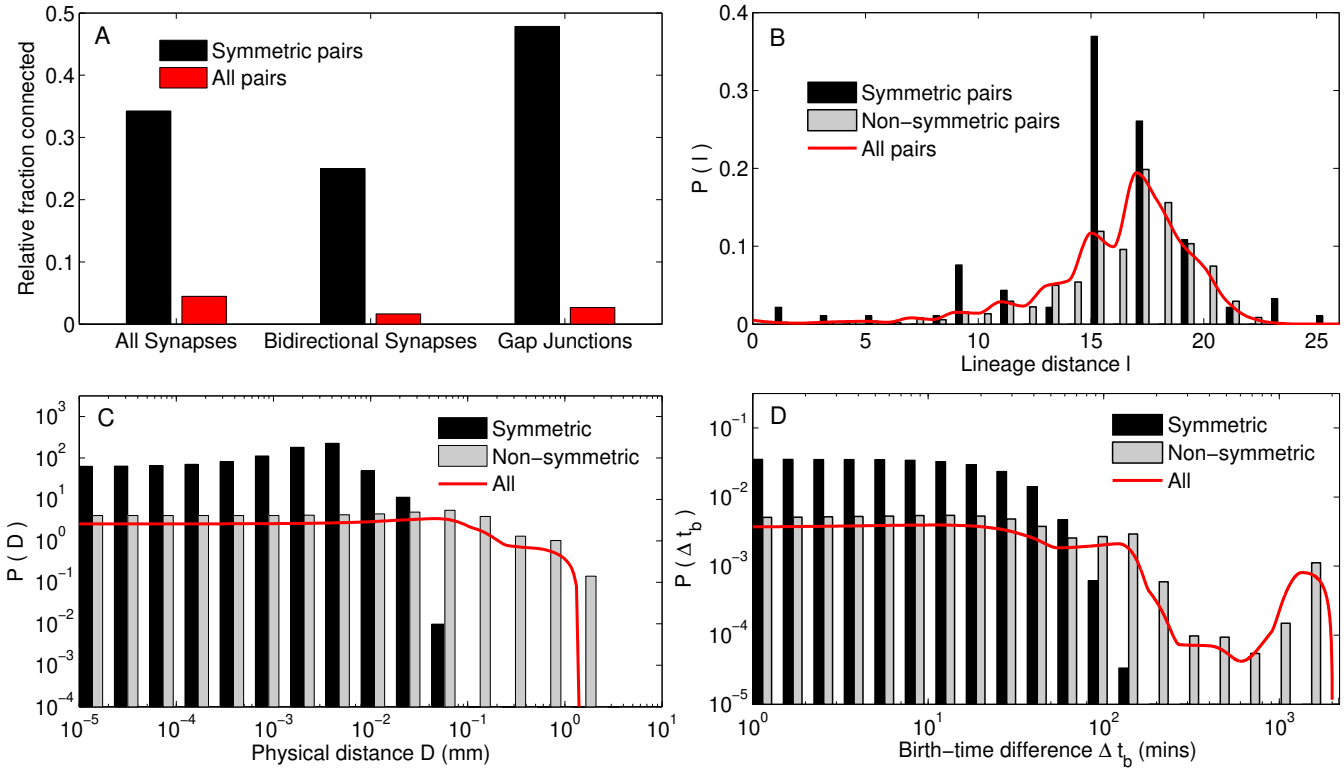


Figure 2.5: Symmetrically paired neurons have a high probability of being connected and also exhibit strong association in their birth times and spatial positions. (A) Bilaterally symmetric neurons that are positioned on the left and right of the body axis of the organism tend to have a much higher probability of synaptic, as well as, gap junctional connections between them, compared to that for all pairs of neurons. In addition, the synapses are highly likely to be reciprocal (bidirectional). (B) The distribution of lineage distances between paired neurons show that the mean value is lower than that for all neurons. We note that almost all lineage distances between symmetric neurons are odd-valued suggesting that they occur at the same rung of the lineage tree. (C-D) Symmetrically paired neurons have cell bodies located in physical proximity of each other (C) and are born close in time as indicated by low birth-time differences Δt_b (D), compared to all pairs of neurons.

In particular, whenever one member of a bilaterally symmetric pair occurs in any of the known functional circuits obtained through behavioral assays [131, 132, 133] (discussed later), the other also appears in it without exception. While it is known that this symmetric nature is manifested in the spatial arrangement (e.g., location of the cell bodies) and connection structure of paired neurons, here we ask whether bilaterally symmetric neurons share a similar network neighborhood, i.e., whether there is a high degree of overlap between

the neurons that each of them connect to, or indeed whether they have a significantly higher probability of being connected to each other. The latter assumes importance in view of the fact that it is the direct contact between the paired cells AWCL/R that trigger asymmetrical gene expression resulting in differential expression of olfactory-type G-protein coupled receptors in the neurons [134].

Fig. 2.5 (A) shows that indeed the left/right members of a symmetric pair have a much higher probability of connection between them than any two arbitrarily chosen neurons belonging to the somatic system. Moreover, 25% of the bilaterally symmetric pairs have reciprocal synaptic connections with each other, compared to less than 2% of all neuronal pairs being connected in such a bidirectional manner. We can further distinguish the symmetric neuron pairs into those which originate from an early division across left/right axis of the common ABp blastomere (i.e., they have similar lineage differing only in the early cell division event ABpl/r) and those where members of a pair originate from non-symmetric blastomeres (e.g., ABal and ABpr) [129]. These two distinct origins of the bilaterally symmetric neurons are reflected in the two peaks of the distribution of lineage distance between the left/right members of each pair seen in Fig. 2.5 (B), with only the latter category of paired neurons that do not share a bilaterally symmetric lineage history having low values of l. The synaptic connection probability between the members of pairs belonging to these two classes differ only by a small amount (0.34 for the former and 0.35 for the latter, with the corresponding numbers reducing to 0.13 and 0.24, respectively, when we consider reciprocal synapses). The occurrence of gap junctions between bilaterally symmetric neurons is seen to be exceptionally high (47.8% of such pairs being connected) compared to that for the entire system, with no distinction in numbers being observed between the two categories of symmetric pairs. This preponderance of gap-junctional connections between bilaterally symmetric neurons (also indicated by the band diagonal structure of the connectivity matrix shown in panel (B) of Fig. 2.1) suggests that their activity is highly coordinated. This may possibly explain the co-occurrence of both members of a symmetric pair in the different functional circuits.

In addition to exhibiting a high probability of being connected directly, bilaterally

symmetric neurons are also characterized by a high degree of neighborhood similarity. Figure A12 in Appendix A shows the magnitude of overlap between the neurons that each member of a pair is connected by a synapse (either pre- or post-synaptically) or a gap junction, which is seen to be much higher than that for any two arbitrarily chosen neurons. This is consistent with the left/right neurons in the majority of bilaterally symmetric pairs having an identical role in terms of the mesoscopic organization of the network (see discussion related to mesoscopic functional roles). The large number of neighbors that paired neurons share in common is a striking feature that cannot be explained from their physical proximity alone.

We note that almost all lineage distances between symmetric neurons are odd-valued suggesting that they are born at the same rung of the lineage tree. The only exception is the pair AVFL/R, whose members have distinct non-symmetric lineage history, with a lineage distance of 8. Given their shared lineage, it is perhaps unsurprising that most bilaterally symmetric paired neurons also exhibit strong associations in their physical locations and birth times. Panels (C-D) show that a large fraction of the left/right members have cell bodies that are located in close physical proximity of each other (C) and are also born close in time as indicated by low birth-time differences Δt_b (D), compared to all pairs of somatic neurons. Indeed we note that the only exception is the late-born pair SDQL/R with bilaterally symmetric history whose members are located in the anterior and posterior (respectively) parts of the organism, the physical distance between the cell bodies being 0.5 mm.

Relative importance of the different types of homophily in determining the network connectivity

We have demonstrated here the existence of four different types of homophily, i.e., preference of neurons to connect to other neurons having identical or similar attribute(s). We identify these attributes to be (i) process length, (ii) birth cohort, (iii) shared lineage and (iv) bilateral symmetric pairing. While (i) and (ii) are properties characterizing individual

cells that allow neurons to be classified into distinct categories, (iii) is measured in terms of the lineage distance between two neurons and is, hence, an attribute of a pair, as is (iv). Thus, in order to quantitatively demonstrate homophily for these four attributes, we have had to use different measures, viz., modularity Q in connections between neurons belonging to the same (as opposed to different) categories for (i) and (ii), correlation between connection probabilities and lineage distances for (iii) and comparison of connection probability between symmetrically paired neurons with that for the entire network for (iv). We have also quantitatively established that these attributes are not dependent on each other (see below).

We can now ask about the relative contributions of the four attributes in determining the connectivity of the *C. elegans* nervous system. While, the lack of a common measure means that we cannot directly compare numerical values characterizing these attributes, we can estimate how strongly each of them affect the connection probability by using logistic regression analysis (see Methods). Here, the connection probability P between a pair of neurons is expressed as a function of four independent predictor variables X_p, X_b, X_l and X_s corresponding to the four attributes that show homophily. For this, we first establish that the predictor variables are not correlated by using Belsley collinearity diagnostics [119] (see Methods). The condition indices for all predictors are less than 5, indicating very weak dependencies among them. Furthermore, following the logistic regression analysis we observed that the p-values for each of the predictor variables are extremely low ($p \sim 0$) which is indicative of very high significance for the dependence of the connection probability on all of the predictors. For synaptic connections, the regression coefficients estimated from the empirical data are $\beta_p^{syn} = 0.35$ (for process length), $\beta_b^{syn} = 0.71$ (for birth cohort), $\beta_l^{syn} = -0.06$ (for lineage relation) and $\beta_s^{syn} = 1.78$ (for symmetric pairing). Magnitudes of these coefficients indicate the extent by which connection probability is affected upon altering the numerical value of the corresponding predictor variable by a single unit (keeping the other predictors unchanged). Thus, symmetric pairing seems to have the strongest influence in determining the synaptic connection probability. Birth cohort homophily appears to have the next highest contribution followed by process length. Lineage homophily

has the weakest contribution, which may be surprising given the almost linear dependence between lineage distance and connection probability [Fig. 2.2 (D)]. This is possibly related to the fact that X_l is the only predictor variable whose numerical values are not confined to be binary but instead ranges between 1 and 25. This suggests that the connection probability for a pair of neurons having a lineage distance of 2 will not differ much from that for a pair having lineage distance 3, as compared to, for instance, the difference between neurons belonging to the same and to different birth cohorts. We note that, increasing the lineage distance between a pair of neurons by 6 units would have approximately the same effect on connection probability as the difference in the probability of connections between neurons belonging to the same process length category and different process length categories, given that $|\beta_p| \approx 6 \times |\beta_l|$. Using similar arguments, we can see that increasing lineage distance by 12 units would lead to an approximately equivalent change in the connection probability as seen between neurons belonging to the same birth cohort and to different cohorts ($|\beta_b| \approx 12 \times |\beta_l|$). For gap-junctions, the regression coefficients estimated from the empirical data are $\beta_p^{gap} = 0.22$ (for process length), $\beta_b^{gap} = 0.16$ (for birth cohort), $\beta_l^{gap} = -0.08$ (for lineage relation) and $\beta_s^{gap} = 3.22$ (for symmetric pairing). Thus, symmetric pairing and lineage relation have the strongest and weakest contributions, respectively, for this case also. However, unlike synapses, process length homophily has a larger effect on gap-junction connection probability than birth cohort homophily.

2.3.2 Temporal hierarchy of the appearance of neurons during development is associated with their functional identity

We have been focusing, so far, on the various properties related to the developmental history of neurons which govern their spatial organization as well as their inter-connectivity. The latter, as we have shown above, is guided by several types of homophily, i.e., the tendency of neurons which are similar in terms of certain features - viz., process length, lineage, birth-time and bilateral symmetry - to be connected via synapses or gap junctions. We shall now see how the functional identities of neurons are related to their developmental

histories. In particular, we show that classes of neurons distinguished by their (i) type (viz., sensory, motor and interneurons), (ii) functional role in the mesoscopic structural organization of the network and (iii) membership in distinct functional circuits, strongly influences the temporal order of their appearance in the developmental chronology of the nervous system.

Sensory, Inter and Motor neurons

One of the simplest classifications of neurons is according to their position in the hierarchy along which signals travel in the nervous system. Thus, *sensory* neurons receive information from receptors located on the body surface of the organism and transmit them onward to *interneurons*, which allow signals arriving from different parts to be integrated, with appropriate response being eventually communicated to *motor* neurons that activate effectors such as muscle cells. In the mature *C. elegans* somatic nervous system, the motor neurons form the majority (106), while sensory (77) and interneurons (83) are comparable in number. The remaining neurons are polymodal and cannot be uniquely assigned to a specific functional type. In Fig. 2.6 (A) we show how the sub-populations corresponding to each of the distinct types evolve over the course of development of the organism.

We immediately note that while the bulk of the sensory and interneurons differentiate early, i.e., in the embryonic stage, followed by a more gradual appearance of the few remaining ones in the larval stages, more than half of the motor neurons appear much later after hatching. Moreover, of the 48 motor neurons which appear early, approximately half (23) belong to the nerve ring while the rest are in the ventral cord, where they almost exclusively innervate dorsal muscles (the positions of neurons, classified according to function type and birth time, is shown in Appendix A, Figure A13). On the other hand, the 58 late-born motor neurons primarily belong to the ventral cord (with only 4 appearing in the nerve ring). In addition, the majority of them (41) innervate ventral body muscles (see Appendix A, Table A4 for details). The few (11) late-born motor neurons that do innervate dorsal muscles differ from the early-born ones in that they do not have

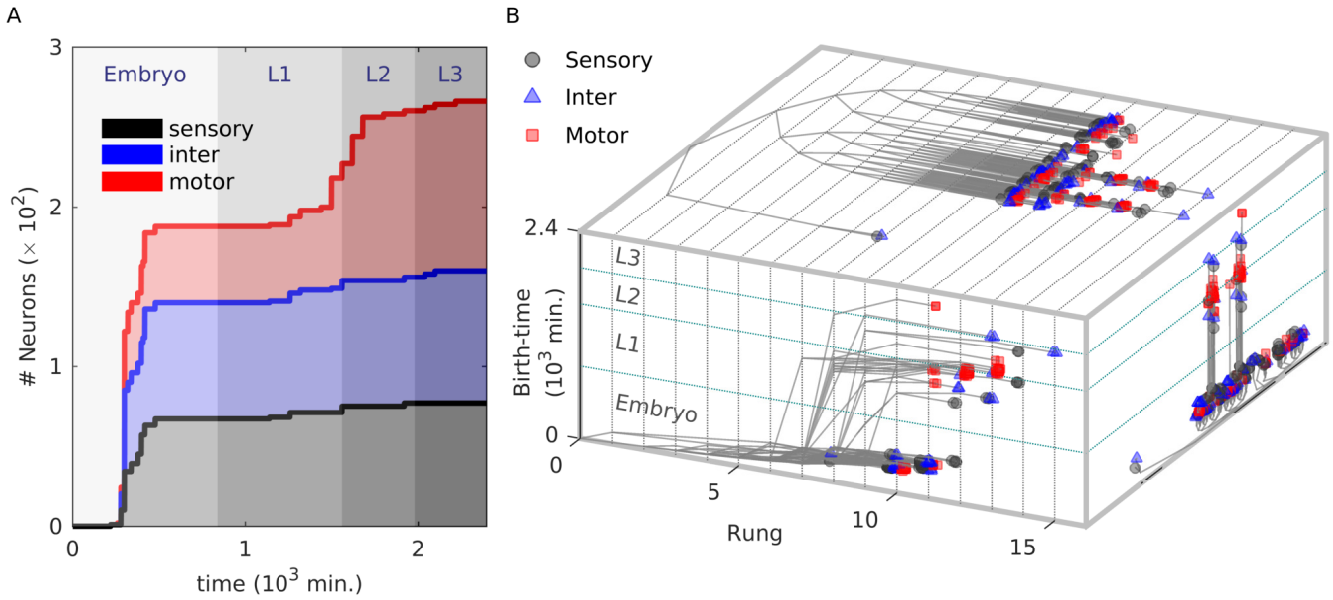


Figure 2.6: Developmental histories of neurons show a bifurcation into early and late branches, with a predominance of motor neurons in the latter. (A) Bulk of the sensory and interneurons appear early, i.e., during the embryonic stage, while a large fraction of motor neurons differentiate much later (L2 or L3) during development. (B) Planar projections of a three-dimensional representation of the developmental history of the entire somatic nervous system of *C. elegans*. Different colors and symbols have been used to denote distinct neuron types (viz., sensory, motor and interneurons). The projection on the top surface shows the lineage tree with branching lines connecting the single cell zygote (shown at rung 0) to each of the differentiated neurons located on their corresponding rungs. At higher rungs (> 11) we see that the differentiated cells are tightly clustered into two bundles of branches with a predominance of motor neurons (also seen in the chrono-dendrogram projection shown at the right face of the base). We note the absence of segregated clusters comprising exclusively the same functional type of neurons (viz., sensory, motor or inter), suggesting that the progenitor cell can give rise to neurons of different types. This in turn implies that commitment to a particular neuron function occurs quite late in the sequence of cell divisions. The projection along the base (left face) shows trajectories representing the developmental history of each final differentiated neuron, indicating the time of each cell division starting from the zygote along with the corresponding rung. For the first few rungs, cell division across different lineages appear to be synchronized and occur at regular time intervals, which is manifested as an almost linear relation between time of division and rung. However, between rungs 6-9, we observe a bifurcation of the trajectories into two clusters widely separated in time. One of these comprises cells which differentiate in the embryonic stage (termed as the "early branch") while the other consists of cells that differentiate much later ("late branch"). This is manifested in a bimodal distribution of birth times for neurons occurring in rungs ≥ 10. In contrast to the regularly spaced cell divisions in the early branch, the trajectories belonging to the late branch are widely dispersed, with relatively little correlation between birth time of neurons and their rungs.

complementary partners and bring about asymmetric muscle activation [135]. This early innervation of dorsal muscle but late, larval-stage innervation of ventral muscles could embody developmental constraints that deserve further exploration in the future.

Having looked at how neurons emerge according to their functional type at different times and at different locations in the physical space described by the body of the worm, we now consider the appearance of such neurons in the developmental space defined by lineage and birth time [Fig. 2.6 (B)]. The projections of the chrono-dendrogram that are shown on the top and the extreme right surfaces, both correspond to representations of the lineage tree that are demarcated by rung and birth time, respectively. We note immediately that the developmental trajectories of the neurons appearing in the late burst of development are clustered into two distinct branches that originate in an early division across left/right axis of the common ABp blastomere (i.e., cells in one branch originate from ABpl, while those in the other emanate from ABpr). Unlike the case seen for neurons belonging to a specific ganglion, we observe that neurons of the same functional type do not form localized clusters in the tree that would have suggested a common ancestry. Thus, progenitor cells can give rise to neurons of each of the different functional types, suggesting that the commitment to a sensory/motor/interneuron fate happens later in the sequence of divisions during development.

The projection on the remaining bounding surface (left face of the base) shows the trajectories followed by cells to their eventual neuronal fate across a space defined by the rung of the lineage tree along one axis and the time of cell division along the other. These trace the developmental history of the entire ensemble of neurons comprising the somatic nervous system. We observe that in the early phase of embryonic stage (corresponding to rungs ≤ 6) there is a linear relation between the time at which a cell divides and the rung occupied by the resulting daughter cells. This implies that cell divisions across different branches of the lineage tree occur at regular time intervals in a synchronized manner. Following this, we observe that the trajectories bifurcate and cluster into two branches that are widely separated in time. The 'early branch', which results in cells differentiating to a neuronal fate much before hatching, continues to follow the trend seen in the earlier

rungs. However, several progenitor cells (that can occur in rungs ranging between 6 and 9) suspend their division for extremely long times, i.e., until after hatching. These comprise the 'late branch' where the final neuronal cell fate is achieved in the larval stages (L1-L3). The occurrence of these two branches gives rise to the bimodal distribution of birth-times shown in Fig. 2.2 (B). In contrast to the regular, synchronized cell divisions across different lineages seen in the 'early branch', the 'late branch' exhibits a relative lack of correlation between rung and birth time, manifested as a wide dispersion of trajectories followed by individual cells. We note that the majority of differentiated neurons that eventually result from the late branch are motor neurons, which corresponds to the late increase in the subpopulation of motor neurons seen in Fig. 2.6 (A). Although there is little information as to when synapses form, the late appearance of the majority of the motor neurons could suggest that stimuli from neighboring neurons are playing an important role in shaping their connectivity in comparison to that of sensory and interneurons that are primarily guided by molecular cues.

Mesoscopic functional roles

Turning from the intrinsic features of neurons to the properties they acquire as a consequence of the network connection topology, we observe that it has been already noted that neurons that have a large number of connections are born early [83],[26]. This could possibly arise as a result of the longer time available prior to maturation of the organism for connections between these early born neurons to be formed with other neurons, including those that differentiate much later. However, as many neurons which have relatively fewer connections are also born in the early stage, there does not seem to be a simple relation between the degree of a neuron and its place in the developmental chronology. To explore in more depth how the connectivity of a neuron is related to the temporal order of their appearance, we therefore consider the role played by it in the mesoscopic structural organization of the network.

Specifically, we focus on the six previously identified *topological modules* of the

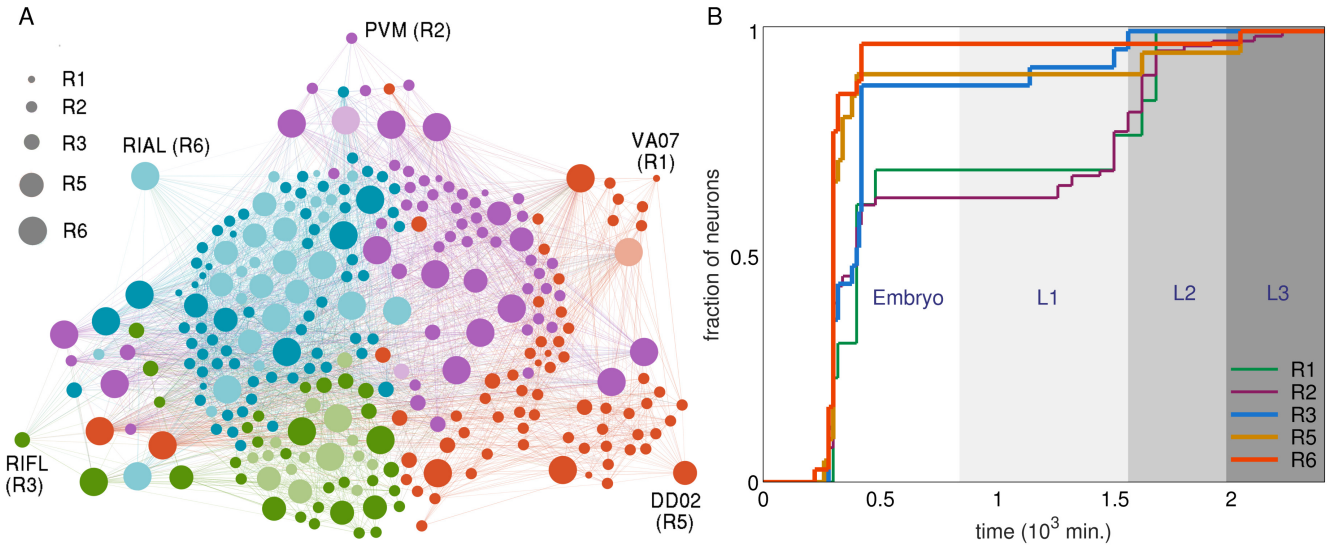


Figure 2.7: Neurons functioning as connectors between different network modules lead in development. (A) Schematic representation of the network of neurons belonging to the somatic nervous system of *Caenorhabditis elegans*, indicating the role of each neuron (indicated by the node size, see legend) in the mesoscopic structural organization of the network. This organization is manifest in the partitioning of the entire network into six structural modules [82] which are characterized by relatively dense connections among neurons in each module compared to the connections between neurons belonging to different modules (node color representing the identity of a module to which a neuron belongs). Within each module, neurons can be further distinguished into those which have significantly higher number of connections to neurons within their own module (hubs) and those which do not (non-hubs). According to their intra- and inter-modular connectivity, every neuron is then classified into one of seven possible categories (see Methods), viz. R1: ultra-peripheral (non-hub nodes with all their connections confined to their own module), R2: peripheral (non-hub nodes with most of their connections occurring within their module), R3: satellite connectors (non-hub nodes having with many connections to other modules), R4: kin-less (non-hub nodes with connections distributed uniformly among all modules), R5: provincial hubs (hub nodes with a large majority of connections within their module), R6: connector hubs (hub nodes with many connections to other modules) and R7: global hubs (hub nodes with connections distributed uniformly among all modules). One representative neuron from each of the categories is separately indicated with a label identifying them by name (note that there are no neurons in the *C. elegans* somatic nervous system which belong to categories R4 or R7). Neurons which function as connectors, e.g., RIAL (R6) and RIFL (R3), are seen to have links to neurons belonging to many different modules (as indicated by the node color of their network neighborings) while neurons belonging to other categories are connected predominantly to neurons within their own modules (indicated by their network neighborhood being almost homogeneous in terms of node color). Neighbors of labeled neurons are either shown clustered around them (for VA07, PVM, RIFL and DD02) or indicated by a lighter shade of node color (for RIAL).

Figure 2.7 ***(previous page)*****:** (B) Distribution of differentiation times of neurons belonging to the different network functional role categories indicate that the development of those functioning as connectors and/or hubs (i.e., R3, R5 and R6) lead the other classes of neurons in the embryonic, as well as, L1 stages. In particular, more than 90% of satellite connectors, provincial hubs and connector hubs have appeared before hatching, while for the peripheral categories (R1 and R2), 70% or less of their members would have differentiated by that time.

C. elegans neuronal network, which are groups of neurons that have markedly more connections with each other than to neurons belonging to other modules [82]. We classify all the neurons by identifying their function in terms of linking the elements belonging to a module, as well as, connecting different modules to each other [117]. This is done by measuring (i) how significantly well connected a neuron is to other cells in its own module by using the within-module degree z-score, and (ii) how dispersed the connections of a neuron are among the different modules by using the participation coefficient P [136]. Cells are classified as hub or non-hub based on the value of z (see Methods for details). The hubs can be further classified based on the value of P as (R5) *local or provincial* hubs, that have most of their links confined within their own module and (R6) *connector* hubs, that have a substantial number of their connections distributed among other modules. The measured value of P is also used to divide the non-hub neurons into (R1) *ultra-peripheral* nodes, which connect only to members of their own module, (R2) *peripheral* nodes, most of whose links are restricted within their module and (R3) *satellite connectors*, that link to a reasonably high number of neurons outside their module.

Fig. 2.7 (A) shows the roles (indicated by node size) played by each neuron in the somatic nervous system of *C. elegans* using a schematic representation of the network. In principle, while it is also possible to have (R7) *global* hubs and (R4) *kinless* nodes, viz., hub and non-hub nodes that may connect to other neurons homogeneously, regardless of their module, none of the neurons appear to play such roles in the network.

Earlier investigation [82] has already established that the connector hubs are crucial in coordinating most of the vital functions that the *C. elegans* nervous system has to perform. Their importance to the network is further reinforced by observing from Fig. 2.7 (B) that

all but one of the neurons belonging to the R6 category appear early in the embryonic stage, exception being PVNR, which differentiates in the L3 stage. About 90% of (R3) *satellite connectors* and (R5) *provincial* hubs are differentiated before hatching. By contrast, peripheral categories (R1) and (R2) have a much smaller fraction of their members appear in the early burst of development and have to wait till the L1, L2 or L3 stage for the development of their full complement. In particular, *satellite connectors* (R3) in spite of having relatively lower degrees than (R5) or (R6) hubs, develop at par with the hubs. Previous studies suggested that hubs, having high degrees, are expected to develop early [83], but *satellite connectors* developing as early as hubs suggests that not only the degree, but also the distribution of the connections of a neuron among the different modules (quantified by the participation coefficient P), and thus its functional role in coordinating activity across different parts of the network, which is an important determinative factor for its appearance early in the developmental chronology of the nervous system.

Membership in functional circuits

In order to delve deeper into a possible association between the function(s) that a neuron performs in the mature nervous system and its developmental characteristics, specifically its place in the temporal sequence of appearance of the neurons, we now focus on several previously identified functional circuits of *C. elegans* [131, 132, 133]. These are groups of neurons which have been identified by behavioral assay of individuals in which the cells have been removed (e.g., by laser ablation). As their absence results in abnormal or impaired performance of specific functions, these neurons are believed to be crucial for executing those functions, viz., (F1) mechanosensation [92, 93, 94], (F2) egg laying [103, 104, 105], (F3) thermotaxis [106, 107, 108], (F4) chemosensation [95], (F5) feeding [39, 92, 109], (F6) exploration [39, 92, 109] (F7) tap withdrawal [93, 110], (F8) oxygen sensation [111, 112] and (F9) carbon dioxide (CO_2) sensation [113, 114] Note that several neurons belong to multiple functional circuits. Fig. 2.8 (A) shows that one can classify these nine functional circuits into two groups based on whether or not all the constituent neurons of a circuit

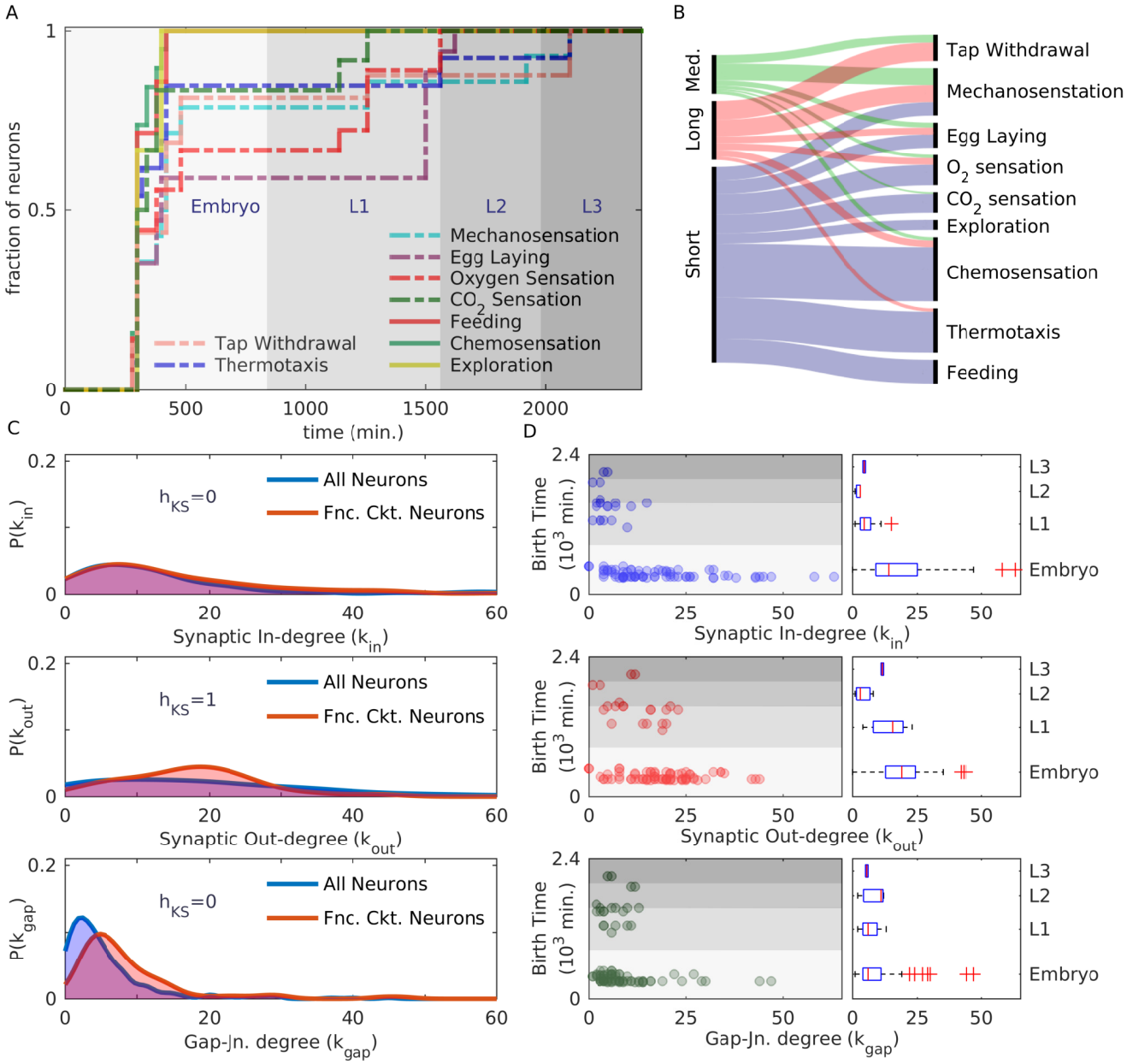


Figure 2.8: The developmental duration of functional circuit neurons are strongly indicative of their process length and connectivity. (A) Distribution of differentiation times of neurons that belong to any of nine functional circuits identified from behavioral assays. Note that the entire complement of neurons belonging to three functional circuits [shown using solid lines] have differentiated before hatching, while those for others [shown using broken lines] are completed later. (B) The distribution of neurons having short, medium and long processes [indicated at left], among the different functional circuits [right]. We note a correlation between the morphological feature of neurite length and the development time of functional circuit neurons, viz., those in the circuits completed before hatching predominantly have short processes, while those in circuits that are completed later mostly have medium to long processes (the exceptions being thermotaxis and CO_2 sensation circuits that comprise a majority of short process neurons).

Figure 2.8 ***(previous page)*****:** (C) Comparison between the distributions of the number of incoming and outgoing synaptic connections (k_{in} [top panel] and k_{out} [middle panel], respectively), as well as, gap junctions (k_{gap} [bottom panel]) of neurons in the entire somatic nervous system (blue) and of the subset of functional circuit neurons (red). We note that the distribution of outgoing synaptic connections for the functional circuit neurons is significantly different from that for the entire network, as indicated by the result of a two-sample Kolmogorov-Smirnov test at 1% level of significance ($h_{KS} = 1$), but this is not the case for incoming synaptic connections or gap junctions ($h_{KS} = 0$). (D) Dispersion of k_{in} (top panel), k_{out} (middle) and k_{gap} (bottom) for the functional circuit neurons differentiating at various times is shown in terms of the adjoining box plots where neurons are clustered into four groups according to the developmental stage during which they are born, viz., Embryo, L1, L2 or L3. In general, the distributions are far more broad for the early born neurons (Embryo) compared to those born later (L1-L3). Focusing on the functional circuit neurons that develop in the embryonic stage, we note that the distribution of incoming connections is more skewed than that for outgoing connections. The distribution of gap junctions is even more skewed, with outliers lying very far from the median.

appear during the early burst of development in the embryonic stage. Thus, while circuits for F4-F6 (shown using solid lines in the figure) have their entire complement of cells differentiate prior to hatching, circuits for F1-F3 and F8-F9 lag behind (broken lines), with less than 60% of the egg laying circuit having appeared at the time of hatching. Indeed, for the entire set of neurons for the latter circuits to emerge one has to wait until the late L1 (for F8 and F9), L2 (for F2) or L3 (for F1, F3 and F7) stages [note that out of the 16 neurons in the F7 circuit, 15 are common to those belonging in the F1 circuit, making the former almost a subset of the latter]. While it makes intuitive sense that egg laying circuit does not have all its components in place by the time of hatching (as the function is required only in the adult), it may appear surprising that many circuits mediating functions vital for the survival of the organism (such as mechanosensation, thermotaxis and oxygen sensation) are not completed at the embryonic stage itself.

However, upon taking a closer look at the late-born neurons belonging to these functional circuits we note that even though these neurons are essential for the normal execution of the corresponding function, their absence have also been individually shown not to result in a significant decline in the function. For example, it has been shown that the ablation of the late born neurons PVDL/R and AVM, which belong to both the tap withdrawal and mechanosensation circuits, do not significantly impair the response of the worm to a

tap stimulus [93]. The same is the case for the late born neurons PDEL/R and PVM belonging to the mechanosensation circuit, whose removal does not significantly affect touch response [110]. The late born neurons PLNL/R and SDQL/R in the oxygen sensation circuit and PQR and AQR belonging to the CO_2 sensation circuits have all been identified as "minor" sensory neurons for the respective gases [98]. It is to be noted that the principal sensory neurons belonging to both of these circuits appear before the worm hatches. In the thermotaxis circuit, the ablation of the late born neurons PVDL/R have been shown not to cause any significant impairment of normal thermotactic behavior [108]. While the thermosensory neurons PHCL/R at the tail that are also born late are indeed essential for thermal avoidance behavior [108], the corresponding neurons FLPL/R in the head are present before hatching ensuring that thermosensory behavior of the worm is not seriously compromised immediately after hatching. Thus, the apparent paradox of how the worm manages to survive after hatching even though several of its functionally critical circuits are not yet complete by that time, is answered by the fact that the role of the late-born neurons belonging to these circuits is often relatively minor. Possibly the sole exception is the egg-laying circuit, in which some of the motor neurons (VC1, VC2, VC3, VC4, VC5) responsible for generating muscle movement appear only after hatching.

An intriguing relation between process lengths of neurons and their occurrence in different functional circuits is suggested by Fig. 2.8 (B), from which we see that circuits which have their entire complement of neurons differentiate early, viz., F5-F7, are dominated by neurons having short processes. In contrast, circuits such as F1, F2, F7 and F8 that take much longer to have all their members appear comprise a large number of neurons with medium or long processes (the exceptions being circuits F3 and F9 that have predominantly short neurons). This association between a morphological feature (viz., neurite length) of a functionally important neuron and its time of appearance suggests a possible connection with the process length homophily, viz., preferential connection between neurons having short processes, mentioned earlier. Neurons with short processes that belong to the "early" functional circuits are mostly chemosensory or interneurons that are all located in the head region. To perform their task these neurons only need to connect to each other, whose

cell bodies are mostly in close physical proximity of each other. Moreover, having their synapses localized within a small region allows them to be activated by neuromodulation through diffusion of peptides and other molecules [137, 138]. This assumes significance in light of our observation that process length homophily between short process neurons is marginally enhanced in the head. The value of Q, a quantitative measure of homophily introduced earlier, is 0.1 (for synapse, for gap junctions it is 0.13) for early born short process neurons which have their cell bodies located in the head region. In contrast, when we consider all short process neurons, Q is 0.067 for synapse and 0.085 for gap junction, respectively. Thus, the process length homophily we reported earlier could arise in short process neurons because of functional reasons.

We shall now see how consideration of functional circuits help in obtaining a deeper understanding of the nuanced relation between the degree of a neuron and the time of its birth that was discussed above (in the context of mesoscopic functional roles of neurons). As seen in Fig. 2.8 (C), neurons belonging to the functional circuits show a significantly different distribution for the number of synaptic connections (both incoming and outgoing) from that of the entire system, as indicated by the results of two-sample Kolmogorov-Smirnov test (test statistic $h_{KS} = 1$) at 1% level of significance. Thus, the set of functionally critical neurons - which, on average, have a larger number of connections than a typical neuron in the somatic nervous system - may need to be treated separately from the other neurons when we examine how synaptic degree correlates with birth time. In contrast, their gap junctional degree distributions cannot be distinguished from that of all neurons, as indicated by the result ($h_{KS} = 0$) for the statistical test of significance.

Considering only the neurons that appear in functional circuits, we observe that most of the neurons having a large number of synaptic connections (particularly, incoming ones) do tend to appear early [Fig. 2.8 (D)]. On comparing the distributions of synaptic in-degree separately for early and late appearing functionally critical neurons (see Appendix A, Figure A14) we note that their difference is indeed statistically significant. When we consider the distribution of the synaptic out-degree we see a very different result. The distributions for the early and late born functionally critical neurons turn out to be sta-

tistically indistinguishable (despite the appearance of a few extreme outliers such as the command interneurons AVAL/R. In contrast, the rest of the neurons show a much broader distribution (statistically distinguishable using a two-sample Kolmogorov-Smirnov test) for the neurons that are born early, compared to those which are born late. This is consistent with the assumption that pre-synaptic neurons that exist for a longer period during development, are able to form many more connections than those neurons which appear later (the latter presumably having less time to form connections before the maturation of the nervous system). From this perspective, it is thus striking that the late born functionally critical neurons have as many connections as they do (making them statistically indistinguishable from the early born set), and is possibly related to their inclusion in the functional circuits.

When we consider the gap-junctional degree distributions, we observe that there is no statistically significant difference between the distributions for early and late born neurons, whether they be functionally critical or other neurons. The box plots showing the nature of the distribution at different developmental stages are all fairly narrow [bottom panel of Fig. 2.8 (D)], even though the embryonic one shows several outliers with the four farthest ones being the command interneurons AVAL/R and AVBL/R that appear in four functional circuits, viz., those of mechanosensation, tap withdrawal, chemosensation and thermotaxis. These, in fact, correspond to the outlying peaks of the k_{gap} distribution, located on the extended tail at the right of the bulk [bottom panel of Fig. 2.8 (C)]. Indeed, the outliers in each of the distributions (for k_{in}, k_{out} and k_{gap}), that appear only at the embryonic stage, almost always happen to be command interneurons. Of these, AVAL/R are common across the distributions and the fact that they occur in four of the known functional circuits underlines the relation between function, connectivity and the temporal order of appearance of neurons that we have sought to establish in this chapter.

2.4 Discussion

The nervous system, characterized by highly organized patterns of interactions between neurons and associated cells, is possibly the most complex of all organ systems that is assembled in an animal embryo over the course of development [139]. For this neural network to be functional, it is vital that the cells are able to form precisely delineated connections with other cells that will give rise to specific actions. This raises the question of how the "brain wiring problem" is resolved during the development process of an organism. In addition to the processes of cellular differentiation, morphogenesis and migration that are also seen in other tissue, cells in the nervous system are also capable of activity which modulates the development of the neighboring cells they may interact with. Processes extending from the neuronal cell bodies are guided towards designated targets by molecular cues, and the resulting connections are subsequently refined (e.g., by pruning) through the activity of the cells themselves. In this chapter we have looked at a more abstract level of guiding principles that can help in connecting the details of cellular wiring at the implementation level of molecular mechanisms with the final result, viz., the spatial organization and connection topology of an entire nervous system. Using the relatively simple nervous system of the model organism *Caenorhabditis elegans*, whose entire developmental lineage and connectivity are completely mapped, we have strived to show how development itself provides constraints for the design of the nervous system.

One of our key findings is that neurons with similar attributes, specifically, (i) the lengths of the processes extending from the cell body (short/ medium/ long), (ii) the birth cohort to which they belong (early/ late), (iii) the extent of shared lineage and (iv) bilateral symmetry pairing (left/ right), exhibit a significant preference for connecting to each other (homophily). Moreover, each of these are manifested by both the connection topology of the network of chemical synapses, as well as, that of electrical gap-junctions, despite the fundamental differences in the nature of these distinct types of links.

We have already discussed earlier a plausible mechanism by which homophily based on lineage would be observed. As neurons are displaced from their initial locations while

retaining their connections that have formed already, cells that share common lineage tend to move apart. An alternative possibility that may explain lineage homophily for synaptic connections is related to the suggestion that synaptogenesis could be guided by cellular labels that are specified by a combinatorial code of neural cell adhesion proteins [68]. In this scenario, cells that are close in terms of their lineage will be likely to share several of the recognition molecules that will together determine the label code. Thus, if a sufficiently large number of these determinants match each other, it could promote synapse formation between such cells, resulting in lineage homophily. We would also like to note that, apart from playing an important role in determining the topological structure of the synaptic network, lineage relations between neurons also appear to shape the spatial organization of neurons by segregating them into different ganglia.

In addition to investigating the probability that a connection will occur between a pair of neurons during development, our study also considers how the distance between cell bodies of the neurons thus connected is distributed. Our results suggest that for synapses, the process length of the pre-synaptic neuron is a decisive factor in determining the separation that is allowed between the neuronal partners. Birth time also appears to play a role, particularly, in the case of long-range connections, i.e., between neurons whose cell bodies are separated by more than two-thirds of the worm body length. Specifically, such connections occur between pre-synaptic neurons that are born early and post-synaptic neurons that are born late, much more often than is expected by chance. This suggests the existence of an active process for the formation of such long-range connections, for example, using fasciculation as an axon guidance mechanism [39, 48]. The latter involves a few pioneer neurons with long process lengths acting as supporting pathways that guide axons of the later developing neurons. This may also underlie a triadic closure-like phenomenon in the network [123, 140] (viz., two neurons having links to one or more common neighbors that have an increased likelihood of being connected to each other). Such a process is known to yield strongly clustered networks with high communication efficiency [141, 142] and could be responsible for the appearance of the so-called "common neighbor rule" that has been reported in the *C. elegans* connectome [143].

Our results also indicate that the temporal sequence in which neurons appear during development of the nervous system is linked to their functional identities. The simplest of these identities is simply the basic functional type of a neuron, viz., whether it is sensory, motor or interneuron. The neurons belonging to these different types not only differ in terms of the cells they connect to (for instance, only motor neurons connect to muscles and sensory neurons are the only ones to receive connections from receptors, while interneurons connect to all types of neurons) but also in their molecular inventory. While their lineage does not show any significant differences, the different functional types of neurons do appear to segregate to a certain extent in terms of their time of appearance. Specifically, we find that the bulk of neurons that are born in the late, post-embryonic burst of development are motor neurons. Our results suggest that there may be temporal cues that appear late in the process of development which are responsible for the specialization of neurons into different functional types according to their time of birth.

At a higher, mesoscopic level of organization of the network structure, we have considered the functional role of neurons in coordinating the activity of different topological modules of the network. We show that this allows us to obtain a much more nuanced picture of how the number of connections that a neuron has with other neurons, affects its place in the temporal sequence in which neurons appear during development. Thus, rather than a simple case of just the degree (the total number of connections) of a neuron deciding its precedence in the sequence, it is both the hub (provincial, as well as, connector hubs) and the connector neurons (satellite connectors, in addition to the connector hubs) that appear early. We also examine in detail the subset of neurons that have been identified as belonging to one or more functional circuits in the *C. elegans*. We observe that membership of a specific functional circuit does determine the order in which these neurons appear, with certain circuits, such as those responsible for chemosensation, emerging early (before hatching) while others circuits, such as those for mechanosensation and egg-laying, appear much later (after hatching). In turn, the time of appearance of functional circuits determines to an extent the morphological properties, such as the process lengths, of their constituent neurons.

The observations reported in this chapter are an attempt at resolving the "wiring problem" for the *C. elegans* nervous system by focusing at a level that is intermediate between the molecular mechanism-level details of developmental processes and the resulting structural organization of the entire somatic nervous system. Specifically, we have attempted to uncover general strategic principles governing the design of the neuronal network, which will allow linking the complicated molecular machinery involved to the equally complicated spatial and topological description of the nervous system. The next step in this approach will involve delineating exactly how these governing principles (such as the various types of homophily) are implemented by molecular mechanisms, and how genetics may be relating the temporal sequence of appearance of neurons to their functional identities. Experimental and theoretical progress towards this direction would enable us to achieve a seamless understanding of nervous system development involving different scales.

Chapter 3

Mesoscopic architecture enhances communication across the Macaque connectome revealing structure-function correspondence in the brain

3.1 Introduction

Cortical localization, which refers to specific regions of the cerebral cortex being associated with distinct functions such as vision and language, has long been a dominant paradigm in neuroscience [144]. As the connectome provides the physical substrate for cognition and behavior [145], it would seem intuitive that such localization would be reflected in the structural attributes of the network [31]. However, brain imaging studies show that a large number of regions become active during any cognitive task, ruling out a simple one-to-one correspondence between a certain set of vertices of the connectome and a particular function [146]. This suggests the necessity for a theoretic framework that investigates

the dynamics of the brain in terms of how different areas connect and interact with each other [147]. Such an approach should integrate complementary perspectives that focus on (a) dynamics, where *distributed* activation of the entire network converges to different attractors, and (b) computation, in which *localized* processing of information occurs in a sequential manner, allowing us to interpret cognitive processing as dynamical computation [148].

An integrated view of how local and global coordination of activity across the brain can arise may be obtained by adopting a mesoscopic approach to analyzing the connectome. Such an approach focuses on understanding the interactions within and between communities of densely inter-connected brain areas (*modules*) that have been identified in nervous systems of different organisms [11, 12, 13, 14, 15, 16, 17, 18, 19, 149]. Such structural modularity of the brain is expected from the advantages that such an architecture may confer during evolution and development [150, 19], such as imparting robustness in the presence of constraints on wiring and performance [151, 145]. Traditionally, modules have been viewed in functional terms, associated with innate, domain-specific mental faculties (such as language) that are believed to be relatively independent of each other [152]. Examining how such cognitive modules relate to the structural communities of the connectome addresses the fundamental issue of structure-function correspondence in the brain [77, 153].

In this chapter, we focus on the structure-function relation as evident in the modular organization of the Macaque connectome, which balances specialized and integrated processing by allowing rapid communication at both local and global scales. This is striking in view of the role that modularity plays in promoting information encapsulation in other network architectures [154]. In performing this analysis, we have added curated spatial and functional information concerning the brain areas to the existing database of brain connectivity, which can serve as a resource for the community. The modules revealed by our analysis extends earlier work [12, 16, 149] by including sub-cortical regions. We show that each module comprises both cortical and sub-cortical components, which is intriguing in view of the proposal that the thalamo-cortical loop plays a central role in the computational architecture of the neocortex [155]. More crucially, we show that the

empirically determined pattern of intra- and inter-modular connectivity facilitates local, as well as global, dissemination, complementing studies showing that maximizing information flow may cause model networks to evolve towards a modular structure [156, 157]. Furthermore, while it has been suggested earlier that physical space constraints cannot exclusively account for modules [14, 50, 158], our determination of the space-independent modules and their relatively high overlap with the original communities clearly indicate that the modularity of the Macaque brain has functional significance, viz., the facilitation of communication across the connectome.

3.2 Materials and Methods

Data

Connectivity. We have used as the basis for reconstructing the Macaque connectome a directed network of brain regions (cortical and sub-cortical) that was compiled in Ref. [159] using several hundred tract tracing studies obtained from CoCoMac - a comprehensive neuroinformatics database [160, 161, 162]. The original network comprised 383 vertices, representing regions in the cortex, basal ganglia and thalamus, at different levels of spatial resolution, and 6602 directed edges corresponding to tracts, i.e., myelinated bundles of axons connecting different brain regions, which may span large distances. In this hierarchically organized arrangement of subdivisions starting from the level of the entire brain, the same region may occur multiple times as a vertex could represent an area that is part of a larger area corresponding to a different vertex. For example, the hippocampus is a vertex of the network, as are its subdivisions CA1, CA3 and Dentate Gyrus. Consequently, there is no unique mapping between brain regions and vertices of this network. It also leads to ambiguity in interpreting edges connecting vertices that occur at any of the levels other than the lowest one in the hierarchy. For instance, if both vertices A and B link to C, but B is a sub-division of A, it is unclear if the two edges are distinct. These issues make it difficult to interpret any results obtained by analyzing the original network.

In the connectome we consider here, these issues are avoided by considering only those

nodes that occur at the lowest hierarchical level, i.e., corresponding to regions with no further sub-divisions, in the original network. This yields a network comprising 266 nodes, representing brain areas that span a range of spatial scales ranging from the visual cortex area V1 (which has a volume of $\sim$ 2000 mm^3) to the thalamic region PT#2 (which has a volume of less than 2 mm^3). The network that we consider, consequently, consists of the 2602 directed links that occur between these nodes. Note that this procedure leads to the network having a largest connected component of 261 nodes (as the following five regions do not have any reported connections to the other areas at the lowest hierarchical level: PT#2, 6b-beta, 4a, 4b and Sub.Th). Despite the reduction in the size of the network upon removal of the aforementioned redundancies, the resulting connectome has similar macroscopic properties as the original network, such as the exponential nature of the degree distribution (Fig. S1).

Spatial Positions. As the brain connectome is a spatially embedded network, it is important to consider geometric information such as physical locations and extent of the different brain regions, in addition to the connection topology. As the original network [159] did not contain any spatial information, we have compiled a comprehensive database of the positions of the areas corresponding to each of the nodes, as well as, the volumes spanned by them. We have obtained the stereotaxic coordinates of each brain region in our connectome from several sources. Information about 134 of the 266 regions included in the connectome has been obtained from the website [163] associated with the Paxinos Rhesus Monkey Atlas [164]. For the remaining regions, we manually curated the requisite data from the relevant research literature. The position of a region is identified with the approximate location of its center obtained from the online three-dimensional visualization platform in the website mentioned above. The volume spanned by a particular region was estimated by approximating the cross-sectional area occupied by the region in each of the coronal sections of the brain in which it appears and obtaining the sum of these areas weighted by the thickness of the sections measured along the rostral-caudal axis.

Modularity

A prominent mesoscopic structural property associated with many networks that occur in nature is modular organization. Modules (or communities) are subnetworks that are characterized by a higher density of connections between the constituent nodes compared to that between nodes belonging to different modules [115]. One of the most well-known approaches for determining the modules of a network is to maximize a quantitative measure, Q, defined for a given modular partitioning of the network as, $Q = L^{-1}\Sigma_{i,j}B_{ij}\delta_{c_ic_j}$, where $B_{ij} = A_{ij} - (k_i^{\rm in}k_j^{\rm out}/L)$ are elements of the modularity matrix $\mathbf{B}$ [165, 9]. The adjacency matrix $\mathbf{A}$ ($A_{ij} = 1$, if a directed link exists from j to i, and 0, otherwise) specifies the connection topology of the network, while the number of incoming and outgoing connections of node i are indicated by the in-degree $k_i^{\rm in} = \Sigma_j A_{ij}$ and out-degree $k_i^{\rm out} = \Sigma_j A_{ji}$, respectively, with L ($= \Sigma_j k_j^{\rm in} = \Sigma_j k_j^{\rm out}$) being the total number of connections in the network. The Kronecker delta function δ_{ij} yields 1 if the communities c_i and c_j to which nodes i and j belong respectively, are identical, and is 0 otherwise.

Spectral analysis and its refinement. In order to achieve an optimal partitioning of the network through the maximization of Q we have used the spectral method [9]. Here, we first bisect the network by assigning nodes to one of two communities according to the sign of the elements of the eigenvector corresponding to the largest positive eigenvalue of the symmetrized modularity matrix $\mathbf{B}+\mathbf{B}^{\rm T}$. Subsequently we refine the partition by swapping the nodes between communities in order to achieve the highest possible value of Q. The above procedure is carried out recursively on each of the communities to further subdivide them until Q cannot be increased further [9]. This approach yields a maximum value of Q for a partitioning of the network into 5 modules with $Q_{\rm spectral} = 0.485$.

Robustness of the partitioning. To ensure that the modular partitions of the network obtained using the deterministic spectral technique (described above) are not sensitively dependent on the specific method used for maximizing Q, we have used the stochastic simulated annealing approach to obtain an ensemble of 10^3 optimal partitions. The dis-

similarity between the different partitions generated by each realization of the annealing technique reflects the extent of degeneracy (and hence, ambiguity) inherent in the modular decomposition of the network. Following Refs. [166, 167], for each realization of the simulated annealing approach we begin with an arbitrary partition of the network and iteratively change the modular composition by implementing one of three types of operations: (i) move a randomly chosen node to any other module including a newly created one, (ii) merge two randomly chosen modules and (iii) split a randomly chosen module into two parts so as to minimize the number of connections between the two parts. Any one of the possible operations (across all types) is chosen at each step with equal probability. The resulting partition associated with a change ΔQ in the modularity is accepted with a probability $\exp(-|\Delta Q|/T)$ if $\Delta Q < 0$ and $p = 1$ otherwise. Here, the parameter T, which is analogous to temperature, is decreased over time according to a specified cooling schedule. The process terminates when the number of successive failures at altering the modules exceeds a threshold value. While the Q values corresponding to the partitions obtained for different realization span a wide range, most of them cluster around that obtained from the spectral method, Q_{spectral}. We focus on the 291 partitions whose Q value deviates from Q_{spectral} by less than 3%. As shown in Appendix B (see Fig. B3-B4 and Table B1, Appendix B), the modular membership of 70% of the nodes remain invariant across all of these partitions, and are in fact identical to that obtained from the spectral method, underlining the robustness of the modular decomposition. We have also used alternative methods of module identification that do not rely on maximizing Q, viz., the Infomap method [168], and have obtained qualitatively similar results (see Fig. B2, Appendix B). For a detailed comparison of performance between the Newman's spectral method and the Infomap method, see Appendix C.

Classification of brain regions according to their role in the mesoscopic structural organization of the connectome

The importance of a given region within the topological organization of the Macaque brain network is indicated by its connectivity within its own module (as defined above), as well

as that across the entire brain, which is evident from its connections to regions belonging to other modules. These can be quantitatively measured by the metrics (i) the within module degree z-score (z) and (ii) the participation coefficient (P), respectively [117, 169]. To identify regions that have significantly more connections within their own module, we determine a within module degree z-score:

$$z_i = \frac{k^i_{c_i} - \langle k^j_{c_i} \rangle_{j \in c_i}}{\sqrt{\langle (k^j_{c_i})^2 \rangle_{j \in c_i} - \langle k^j_{c_i} \rangle^2_{j \in c_i}}}, \tag{3.1}$$

where $k^i_{c_i}$ is the number of links between region i and other regions belonging to its module (c_i) and the average $\langle \dots \rangle_{j \in c}$ is taken over all regions in a module c. As in Ref. [14], nodes (regions) having $z > 0.7$ are identified as hubs, the remainder being classified as non-hubs.

In order to distinguish between brain regions in terms of their inter-modular connectivity we calculate the participation coefficient P_i of region i as:

$$P_i = 1 - \sum_{c=1}^{m} \left(\frac{k^i_c}{k^i} \right)^2, \tag{3.2}$$

where k^i_c is number of links that region i has with those regions belonging to module c and $k^i = \sum_c k^i_c$ is the total degree of the i-th node (region). A region whose connections are restricted within its own module has $P_i = 0$ while one whose links are uniformly distributed among the different modules has P_i closer to 1. Based upon the value of P_i, which provides a measure of how well a node (region) bridges different modules, the non-hub regions are classified as ultra-peripheral (R1, $p \leq 0.05$), peripheral (R2, $0.05 < p \leq 0.62$), satellite connectors (R3, $0.62 < p \leq 0.8$) and kinless nodes (R4, $p > 0.8$), while the hubs can be demarcated into provincial hubs (R5, $p \leq 0.3$), connector hubs (R6, $0.3 < p \leq 0.75$) and global hubs (R7, $p > 0.75$).

Degree- and modularity-preserved network randomization

We construct an ensemble of 10^3 networks obtained by randomizing the empirical network preserving the in-degree and out-degree of each node (region) as well as the modular orga-

nization of the network [14]. Each network is obtained by selecting directed connections, e.g., $i \to p$ and $j \to q$, such that the source nodes i, j belong to the same module A and target nodes p, q belong to the same module B (which could be same as A), and then rewire them so as to have $i \to q$ and $j \to p$. This procedure is repeated for 10^6 times for each realization of a randomized network. To randomize the network preserving the degree alone, we follow the same procedure as above with the difference that there is no constraint on the modular membership of the nodes.

Diffusive spreading model

As the function of the connectome is to facilitate communication between the different brain regions, we investigate the role of the empirically observed pattern of intra- and inter-modular connections on the diffusion of information across the system. For this purpose, we consider discrete random walks that, starting from a given node on the network, proceeds at each time step from one node to a randomly chosen node that receives an outwardly directed link from the former. The rate at which spreading occurs in different parts of the system can be analyzed by obtaining the distribution of first passage times (FPTs) for a random walk to reach a target node starting from a source node. For this, we have measured the FPTs τ to all nodes that are visited by a walk initiated from a given node of the network. The process is repeated 10^3 times starting from each of the 266 nodes, with a walk terminating when either every node has been visited at least once or a node with no outgoing connections is reached. Separate distributions for intra-modular FPTs (τ^{intra}) and inter-modular FPTs (τ^{inter}) can be obtained by considering the source and target nodes to be in the same module or in different modules, respectively. For comparison, we also compute the distributions of FPTs τ_D and τ_{DM} for randomized surrogates in which either the degrees, or both the degrees and modular memberships, of the nodes are preserved, respectively. In each case, the distribution is averaged over 20 network realizations. The deviation of the empirical FPT distribution from those obtained from the randomized surrogates by averaging over multiple realizations is quantified in terms of

a z-score measure defined as:

$$z = \frac{P_{emp}(\tau) - \langle P_{rand}(\tau) \rangle}{\langle P_{rand}(\tau)^2 \rangle - \langle P_{rand}(\tau) \rangle^2}, \tag{3.3}$$

where $P_{emp}(\tau)$ and $P_{rand}(\tau)$ are the empirical and randomized surrogate FPT distributions, respectively.

Role of spatial geometry in the modular organization of the connectome

The physical distance d_{ij} between two brain regions i and j, whose centers are indicated by the vectors $\mathbf{x}$ and $\mathbf{y}$, respectively, has been measured in terms of the Euclidean metric $d(\mathbf{x},\mathbf{y})$ and scaled by the geometric mean of the radii r_i, r_j of the two regions (the radius of each region being estimated from the its volume, see Appendix B).

Space-independent partitioning of the network into communities. For networks whose nodes are embedded in a space associated with a metric, it can be argued that the network properties, such as modularity, could be a consequence of the constraints imposed by the underlying geometry. We therefore need to modify the method for determining the modular structure of a network outlined above, in order to take into account the role of the physical space in which the network is embedded. This is done by re-defining the modularity matrix $\mathbf{B}$ in the definition of the quantity Q (given above), so that the expectation of a pair of nodes (i, j, say) being connected by chance in the null model incorporates the physical distance (d_{ij}) between the nodes. Thus, following Ref. [170], we re-define $B_{ij} = A_{ij} - (k_i^{\text{in}} k_j^{\text{out}} f(d_{ij})/L)$, where $f(d) = \Sigma_{d_{ij}=d} A_{ij}/(k_i^{in} k_j^{out})$ is referred to as the *deterrence function*. This function, which is estimated from empirical data for the network, contains information about how the physical distance between a pair of nodes modulates their connection probability. Note that if the communities in the network arise entirely because of spatial dependence, measuring Q taking into account the physical distance between nodes does not yield any modular structure. Moreover, comparing the space-independent modular decomposition of the network obtained using this technique with the

communities determined using exclusively information about the connection topology (as described earlier), we can infer whether the observed modularity is primarily driven by physical distance constraints (see Appendix B). The similarity between the communities obtained using the two methods is quantified using *normalized mutual information*.

Normalized mutual information. To quantify the similarity between two modular decompositions $\{c_i^A\}_{i=1}^{M_A}$ and $\{c_j^B\}_{j=1}^{M_B}$ resulting from different partitionings A and B of a network (that yield M_A and M_B modules, respectively) we have used the normalized mutual information [171]

$$I_{\mathrm{norm}}(A,B) = \frac{2\sum_i \sum_j P(c_i^A, c_j^B) \ln[P(c_i^A, c_j^B)/P(c_i^A)P(c_j^B)]}{-\sum_i P(c_i^A)\ln P(c_i^A) - \sum_j P(c_j^B)\ln P(c_j^B)}, \tag{3.4}$$

where $P(c_i^A)$ is the probability that a randomly chosen node lies in module c_i^A in partition A, $P(c_j^B)$ is the probability that a randomly chosen node lies in module module c_j^B in partition B, and $P(c_i^A, c_j^B)$ is the joint probability that a randomly chosen node belongs to module c_i^A in partition A, as well as, to module c_j^B in partition B ($i = 1, \ldots, M_A$, and $j = 1, \ldots, M_B$). Each of the probabilities can be estimated from the ratio of the community sizes to the size of the entire network.

Surrogate networks. In order to explicitly show that the modular organization is not primarily driven by the constraints imposed by the physical distance d between brain regions, we have demonstrated how spatial embedding affects the modular decomposition of a network, using three classes of surrogate random network ensembles (of size 100 each) having different underlying spatial dependences. The three ensembles, in increasing order of importance of d in governing the connection probability P between nodes, comprise networks with (a) $P \sim d^0$, (b) $P \sim d^{-1}$, which is the case in the empirical network, and (c) $P \sim \exp(-d)$, with nodes in each network occupying the same spatial position as in the empirical network. Each network (comprising an identical number of nodes and links as in the empirical network) was subject to community detection using information

about the connection topology alone, as well as, space-independent modular decomposition, following the two approaches described above. The difference between these two sets of partitions provides a measure of the role that spatial embeddedness of the networks plays in determining the modular nature of their connectivity (see Appendix B).

3.3 Results

3.3.1 Mesoscopic organization of brain areas in the Macaque

Fig. 3.1 (a-c) shows the modular organization of the Macaque brain network spanning regions from the cortex, basal ganglia and thalamus, revealed by our analysis (for details see Methods). The network is seen to comprise 5 modules, each module i being composed of m_i densely inter-connected brain regions (their numbers ranging between 39 and 71, see the color key to the right of Fig. 3.1, a-c, containing the list of brain regions in each module). The membership of the individual regions in these modules is seen to be robust (see Methods). Given that the network is embedded in a specific geometry, namely that of the Macaque brain, it is noteworthy that each of the modules are spatially clustered as is clearly seen from the projections shown in Fig. 3.1(a-c). To understand the implications of the spatial location of these modules, we visually represent the mapping between the modules and the major anatomical subdivisions of the brain in Fig. 3.1 (d) [see also Table B2, Appendix B].

We observe that every module comprises sizable number of both cortical and sub-cortical regions. With the exception of #3, the modules have their sub-cortical components located almost exclusively in the *Thalamus*. We note that each of these modules are associated with different sensory modalities (discussed in detail later), consistent with one of the primary functions of the Thalamus, namely, relaying information from the sensory organs to cortical areas for further processing. As the Thalamus is also involved in sleep-wake regulation coordinated via extensive reciprocal connections with the cortex [172, 173, 174], it is reasonable to expect that the each of the network modules will

have thalamic components along with cortical ones, with dense intra-modular connectivity representing thalamo-cortico-thalamic circuits [175, 176]. However, none of the sub-cortical components of module #3 (displayed in green in Fig. 3.1) belong to the Thalamus and instead constitutes almost the entirety of the *Basal Ganglia*.

The locations of the cortical components of the different modules across the principal lobes of the cortex, viz., *frontal*, *temporal*, *parietal* and *occipital*, are indicated in Fig. 3.1 (d). We observe that there is no simple correspondence between the modules, which are topological partitions of the connectome, and the gross anatomical subdivisions of the cortex. While the regions comprising the frontal and temporal lobes are split between several modules, those in the parietal and occipital are dominated by single modules (modules #2 and #5, respectively), indicating the relative homogeneity of the latter lobes in the mesoscopic organization of the network. This assumes importance in light of a possible connection between the modular divisions and functional specialization in the brain - a point that we discuss below.

As mentioned in the Introduction, the term *module* has been primarily used in the neuroscience literature to refer to a functionally integrated set of areas [178, 179, 180] that allows for "information encapsulation" [152], whereas we employ the term in the sense of a specific meso-level structural feature in the connectome [11, 12, 13, 14, 15, 16, 17, 18, 19]. In analogy with other biological networks where a structure-function correlation has been established for modules [181, 117], we now ask whether the network modules that appear as separate structural units of the brain can be considered as distinct functional units as well. Using information about the known functions of different cortical and sub-cortical areas obtained from decades of experimental studies, we have created a mapping between the regions belonging to each module and the specific functionalities attributed to them (see Table B3, Appendix B). A perusal of this information reveals that the different regions belonging to a module complement each other in carrying out various cognitive functions. For example, several cortical areas in module #5, viz., 45a and 8Ac of the pre-frontal cortex, and V1 and V2 of the occipital lobe, are related through their involvement in vision, even though they may be part of distinct lobes and have disparate functions (controlling

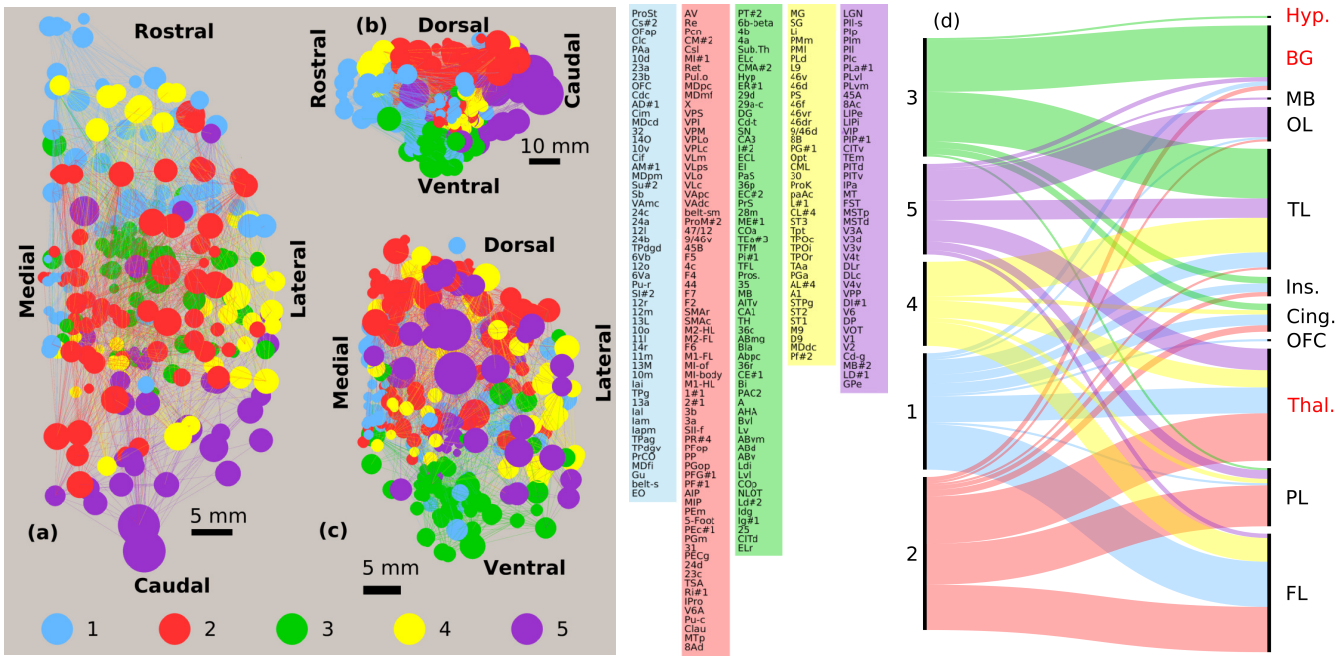


Figure 3.1: Mesoscopic organization of the Macaque brain. The network of brain regions, shown in (a) horizontal, (b) sagittal and (c) coronal projections, clearly indicate that the nodes (filled circles) are organized into five modules, each characterized by dense intra-connectivity. The modular membership of each node is represented by its color (see color key to the right, containing the list of brain regions in each module), while node sizes provide a representation of the relative volumes of the corresponding brain regions (the spatial scale being indicated by the horizontal bar in each panel). The spatial positions of the nodes are specified by the three-dimensional stereotaxic coordinates of the corresponding regions (see Methods). Links indicate the directed nerve tracts connecting pairs of brain regions, and are colored in accordance with their source nodes. For details of each of the brain regions see Appendix B. (d) Visual representation of the association between the network modules and cortical (in black), as well as, sub-cortical (in red) subdivisions of the brain, viz., *FL*: Frontal Lobe, *PL*: Parietal Lobe, *TL*: Temporal Lobe, *OL*: Occipital Lobe, *Cing.*: Cingulate, *Ins.*: Insula, *BG*: Basal Ganglia, *Thal.*: Thalamus, *Hyp.*: Hypothalamus, *OFC*: Olfactory complex, and *MB*: Mid-brain. For a detailed breakdown of the major subdivisions of the brain in terms of their module membership, see Appendix B (Fig. B5-B6, Table B2). This alluvial diagram has been created using the online visualization tool RAW [177].

saccadic eye movements in the case of 45a and 8Ac, and processing of visual information in the case of V1 and V2). This suggests a general scheme of organization in which the regions associated with each of the principal sensory modalities are localized in specific modules, viz., visual in module #5, auditory in module #4, somatosensory (along with the principal motor area M1) in module #2 and olfactory (as well as, gustatory) in module #1. We show below that the known behavior of the regions comprising each of the modules is consistent with the broad functions attributed to that module.

First, we observe that module #5 (diplayed in purple in Fig. 3.1) consists of the primary visual area in the occipital lobe and association areas in the parietal (e.g., LIP, VIP etc.) and temporal lobe (e.g., CIT, PIT, etc.). In addition, its thalamic component includes *lateral geniculate nucleus* (LGN), which relays visual information to the cortex from the retina. We note that these regions are all involved in various aspects of visual cognition, which is consistent with the sensory modality associated with this module, viz., vision. Second, module #4 (displayed in yellow in Fig. 3.1), consistent with its attributed sensory modality, is seen to comprise the auditory cortex lying in the superior temporal gyrus of the temporal lobe (as well as, the corresponding association areas), and the medial geniculate nucleus in the thalamus, which is the relay for all auditory information destined for the cortex from the brainstem [182]. Third, module #2 (displayed in red in Fig. 3.1), contains the primary and secondary somatosensory areas (S1, S2) in the parietal lobe, while its thalamic component contains all the regions which together comprise the ventral posterior nucleus that relays somatosensory information to the cortex. Apart from its sensory function, as noted earlier it also consists of primary and supplementary motor areas which are associated with planning, control and execution of voluntary movements [183]. Finally, we note that module #1 (displayed in blue in Fig. 3.1), has the *olfactory complex* and the *gustatory cortex*, both located in the frontal lobe, as well as, a few other regions (e.g., the olfactory field of the entorhinal cortex, EO, in the temporal lobe) involved in the sensory processing of smell. However, the module is dominated by association areas located in the prefrontal cortex which are involved in high-level multi-modal sensory integration and decision-making [184, 185, 186, 187, 188].

In contrast to the other modules, #3 neither contains motor areas nor does it include any primary or secondary sensory areas. This is possibly related to our earlier observation that this module has a distinct structural arrangement, in that its sub-cortical components do not have any contribution from the thalamus, but instead comprise regions belonging to the basal ganglia. In particular, the module contains the entire *amygdala* which is known to regulate emotional responses and fear-conditioning in mammals [189, 190, 191, 192]. This gains significance in light of the fact that both the *Hippocampus* and the *Parahippocampus*, which are primarily involved in the formation of memory, feature prominentaly among this module's cortical components. It resonates with the known relation between emotional state and formation of memories in individuals that have been established by several studies [193, 194, 195, 196].

As the brain is characterized by structures occurring at several scales [197], it is pertinent to ask whether further levels of organization can be identfied in the connectivity

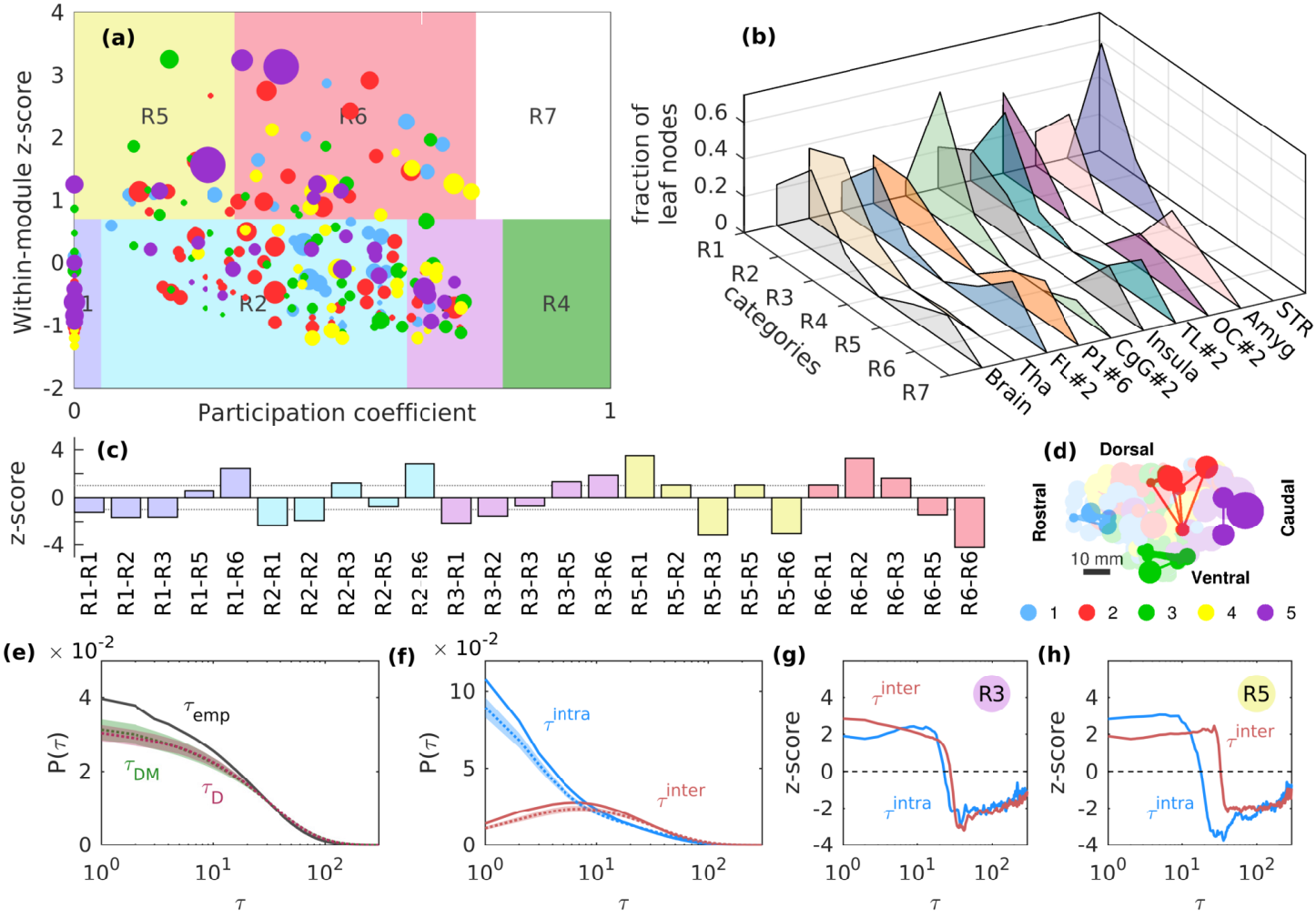


Figure 3.2: Classification of brain regions according to their intra- and inter-modular connectivity. (a) Nodes of the Macaque brain network [colored and scaled as per Fig. 3.1 (a-c)] are displayed in accordance with their within-module degree z-score (z) and participation coefficient (P), which provide a measure of their intra- and inter-modular connectivity respectively. This allows the brain regions to be categorized into one of seven possible categories (see Methods), viz., R1: ultra-peripheral, R2: peripheral, R3: satellite connector, R4: kinless, R5: provincial hub, R6: connector hub, and R7: global hub. Note that there are no regions in the Macaque brain belonging to the categories R4 and R7. (b) The distribution of the regions of the entire Macaque brain across the different categories R1-R7 is similar to the corresponding distributions observed in several anatomical subdivisions, viz., *Tha*: Thalamus, *FL#2*: Frontal Lobe, *P1#6*: Parietal Lobe, *CgG#2*: Cingulate Gyrus, Insula, *TL#2*: Temporal Lobe, *OC#2*: Occipital Lobe, *Amyg*: Amygdala and *STR*: Striatum. (c) The connectivity pattern between regions belonging to the different categories R1-R7 indicated by the z-scores for abundance of links between each pair of categories (the first symbol in Ri-Rj refers to the category of the source region and the second to that of the target), measured with respect to degree- and modularity-preserved randomized ensemble of networks (see Methods).

Figure 3.2 ***(previous page)*:** Large positive (or negative) z-scores, i.e., $z > 1$ (or $z < -1$), indicated by the dotted lines, imply that the corresponding connection types occur significantly more (or less) often than expected from random networks that have degree sequence and community structure identical to the empirical network. (d) Sagittal projection of the network of brain regions [see Fig. 3.1 (b)] showing that connections between provincial hubs (highlighted nodes) are localized within each module. (e) Temporal evolution of spreading processes, quantified in terms of distributions of first passage times (τ) of random walkers starting from one node to reach another, contrasted between the empirical brain network (solid line, τ_{emp}) and randomized ensembles of networks, generated by preserving either the degrees alone (red, τ_D), or both the degree and the modular membership of each node (green, τ_{DM}). (f) The distribution of τ differs significantly, depending on whether the target and source nodes belong to the same module (blue, τ^{intra}) or different modules (red, τ^{inter}). As in (e), spreading occurs significantly more rapidly in the empirical network (solid lines) compared to the networks belonging to the randomized ensemble (obtained by preserving degree and modular membership). In both (e) and (f), the dotted lines and the shaded regions around them represent the mean and standard deviation of $P(\tau)$ calculated over the randomized ensembles. To see how the different categories R1-R7 of brain regions allow spreading to occur faster in the empirical brain network than in equivalent randomized networks, we focus on the cases where the source nodes are either satellite connectors R3 (g) or provincial hubs R5 (h). The z-score indicates that there is a statistically significant shift in the empirical distribution towards lower values of τ in both cases. However, while for R3 the increase in the rate of spreading is similar, irrespective of whether the target is in the same module or in a different one, we observe that for R5, there is a relatively larger shift at lower values for τ^{intra} as compared to τ^{inter}. This is consistent with the connectivity pattern of provincial hubs with the other categories of nodes [shown in (c)] which particularly favors intra-modular communication.

pattern within each of the modules described above. Indeed, when we consider module#5, the most robust under different realizations of network partitioning (see Methods and Appendix B), and subject it to further modular decomposition, we observe that it comprises three communities which we refer to as sub-modules. The largest of these contains the visual cortex and almost the entirety of the sub-cortical components, while the other two (which are comparable to each other in terms of the number of constituent regions) are dominated by regions belonging to the superior temporal sulcus and the intraparietal sulcus, respectively (see Fig. B7, Appendix B). Intriguingly, we note that the latter two communities appear to correspond to regions identified with different visual processing pathways, viz., the dorsal and ventral streams [198, 199].

3.3.2 Distribution profile of nodes in terms of their intra- and inter-modular connectivity is conserved across cortical and sub-cortical subdivisions

Having described the overall organizational structure of the network at the mesoscopic level, we now focus on understanding the role played by the individual brain regions in connecting other regions within their own module, as well as, across modules. The importance of each region is quantified in this framework by measuring the within-module degree z-score and the inter-modular participation coefficient P (see Methods for details).

As seen in Fig. 3.2 (a), the z-score allows regions to be distinguished between *hubs*, i.e., those having significantly higher number of connections to other regions in their module, and *non-hubs*, while P further classifies the hubs into *provincial* (R5), *connector* (R6) and *global* (R7) categories and the non-hubs into *ultra-peripheral* (R1), *peripheral* (R2), *satellite connector* (R3) and *kinless* (R4) classes. We note that regions in each module have a similar distribution across R1-R3 and R5-R6 (with the sole exception of module #4 which has no region playing the role of a provincial hub, see Fig. B8, Appendix B). Uniformity of this nature can also be observed in Fig. 3.2 (b) where we compare the distributions of constituent regions across the different categories for the entire brain with that of the various subdivisions of the cortex, such as the Frontal (FL#2), Parietal (P1#6), Temporal (TL#2) and Occipital (OC#2) lobes, the Insula and the Cingulate Gyrus (CgG#2), as well as, the Amygdala (Amyg) which belongs to the basal ganglia. However, the Striatum which is also in the basal ganglia, and the Thalamus (Tha) have the distinctive characteristic of being essentially devoid of regions that act as hubs, indicating a relative lack of heterogeneity in the number of connections that their constituent regions have with others in their modules.

We have also analyzed the relative frequency with which regions belonging to the different categories connect to each other in the Macaque brain, compared to the corresponding connectivity pattern observed in surrogate networks obtained by degree- and modularity-preserving randomization (see Methods) [169]. The profile of connection preferences be-

tween the various caregories shown in Fig. 3.2 (c), with under-representation of connections between R1-R1, R5-R6 and R6-R6 which has been related to the occurrence of multi-star structures, resembles other networks involved in information propagation [169]. As can also be seen from the figure, non-hubs prefer in general to connect to hubs and vice versa. This is indicative of degree disassortativity, i.e., connections between nodes having dissimilar characteristics (in this case, the number of connections) are favored. However, on investigating the connectivity between pairs of these categories, we notice that source regions belonging to peripheral (R2) and provincial hub (R5) categories show a distinct bias in their connections in terms of the participation coefficient of the target regions. Specifically R2 regions prefer to connect to connectors, both hubs (R6) and non-hubs (R3), while avoiding regions that are localized in their modules (R1, R2 and R5). The trend is reversed for R5 regions. In particular, they show a slight preference for connecting to each other, which is in contrast to the other categories which exhibit a marked tendency to avoid others of their own kind.

This homophily between provincial hubs could arise from two different patterns of connectivity between them, viz., one in which connections between the R5 regions are confined within the same module and another in which the corresponding regions across different modules are connected. Fig. 3.2 (d) shows that the empirical evidence supports the former arrangement where, within each module, provincial hubs connect to each other preferentially. We note that the three R5 regions indicated in module #5 occur, respectively, in the three different sub-modules that were identified in the previous subsection. This intra-modular connectivity within provincial hubs, taken together with the observation that they preferentially connect to peripheral regions while avoiding connectors, suggest that they help co-ordinate activity locally within each module while limiting the spread of information over the network.

3.3.3 Information spreading within the brain is enhanced by the specific pattern of intra- and inter-modular connections

The roles played by regions belonging to different categories in facilitating the transmission of information within and between modules can be investigated by considering a process of diffusive propagation across the network (see Methods). The distribution of first passage times τ, i.e., the time elapsed between initiating a random walk from any source node and the earliest arrival to any given target node, is shown in Fig. 3.2 (e). While, in general, presence of modules in networks leads to slower global diffusion [154], surprisingly we observe that the distribution for the empirical network is markedly shifted towards lower values of τ compared to randomized networks with an identical degree sequence that may or may not have modular organization. This indicates that, as opposed to information encapsulation, the specific pattern of intra- and inter-modular connections between brain regions belonging to different categories actually promotes faster communication across the network. Moreover, as seen from Fig. 3.2 (f), the enhancement of the rate of diffusion in the connectome (in comparison to the randomized surrogates) can be seen both for transmission within a module, as well as, between different modules.

We also investigate how nodes having distinct intra- and inter-modular connectivity roles contribute to enhancing communication in the network. This is achieved in each case by having the source node belong to the respective category and comparing the corresponding distribution of τ with that obtained from randomized surrogates (quantified using z-score, see Methods). Fig. 3.2 (g) shows that starting from a satellite connector R3, diffusion to other nodes belonging both within its module or to other modules is significantly faster compared to randomized networks with identical modular organization and degree distribution. In contrast, as seen from Fig. 3.2 (h), when starting from a provincial hub R5, the increase in the rate of diffusion within a module, compared to that in the surrogate networks, is even higher than the increase in the rate of diffusion across modules. This resonates with the observation of homophily between provinical hubs in a module reported earlier (Fig. 3.2 (d)). When the source node belongs to any of the other categories, the

difference between the intra- and inter-modular diffusion time-scales is seen to lie between the range seen for these two cases (see Fig. B11, Appendix B). This suggests that the modular character of the mesoscopic organization of the connectome is further shaped by the distribution of roles played by the different nodes in allowing information to spread within a module, as well as, across different modules.

3.3.4 Spatial layout constrains the connectivity but does not determine the modular organization of brain regions

So far we have investigated the modular structure of the network of brain regions exclusively in terms of the connection topology. However, the brain is also a physical system that is embedded in three-dimensional space associated with a distance metric which restricts the possible connections between its constituent regions. Such constraints arise from resource costs related to the spatial volume and transmission time associated with the connections, and the rapid energy consumption during synaptic transmission [49, 200, 201, 202, 203, 204, 205]. Thus, given that the pattern of connections between the regions is a function of the physical distance between them, we can ask to what extent are the modules a consequence of the brain being a spatially embedded network [206]. To investigate the role of spatial constraints on the structure of the brain network, we supplement the network topological information with that of the physical locations and volumes of each of the regions (shown in Fig. 3.1, a-c; for details see Methods). By comparing the distributions of the physical distances d between all possible pairs of regions (connected or not) with that of only the connected pairs [top panel of Fig. 3.3 (a)], we can obtain the dependence of the connection probability between two regions on the distance d between them. As seen from the bottom panel of Fig. 3.3 (a), this probability decays linearly with the reciprocal of the distance, i.e., $P(C|d) \sim 1/d$, explicitly demonstrating the constraint imposed by the spatial layout of the brain regions on their connectivity.

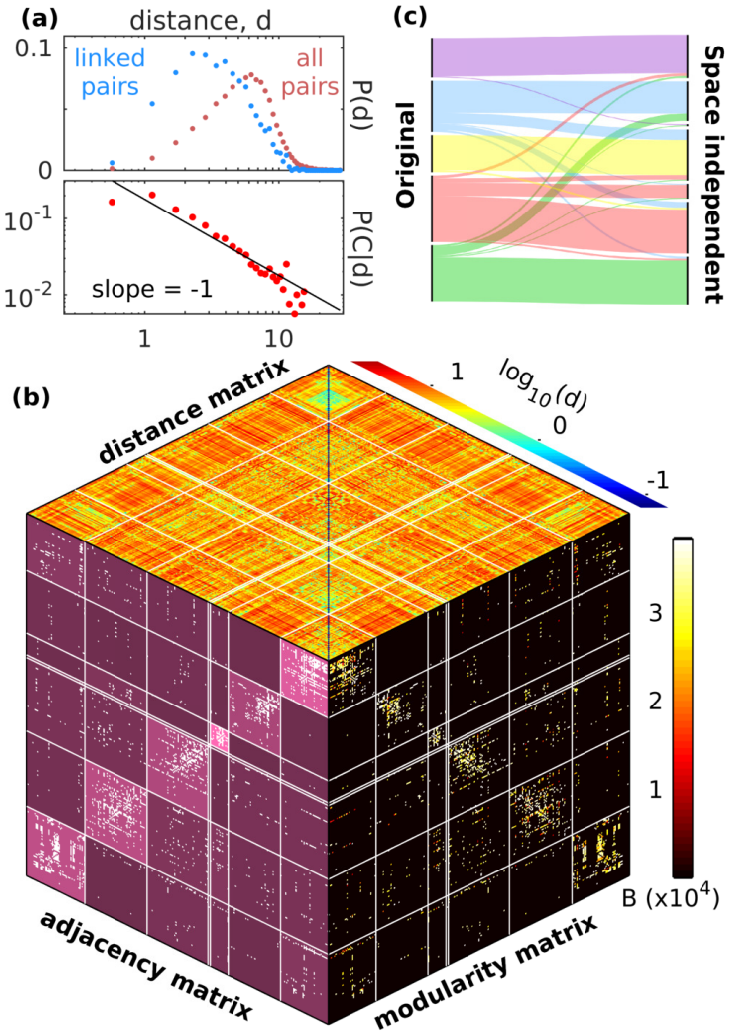


Figure 3.3: Physical distance between brain regions is seen to constrain their connectivity, but the modular organization of the network is independent of their three-dimensional spatial arrangement. (a, top) Probability distribution of the physical distances d between all pairs of nodes (red) contrasted with that of connected pairs (blue). (a, bottom) The variation with physical distance d of the connection probability $P(C|d)$ between a pair of nodes separated by that distance (red). The empirical data is best fit by the relation $P \sim 1/d$ (represented by the solid line). (b) Joint representation of the space-independent modular organization of the network of brain regions showing the matrices indicating adjacency $\{Aij\}$ (left surface), modularity $\{B_{ij}\}$ (normalized by total number of links L, right surface) and physical distance $\{d_{ij}\}$ (top surface) between the different regions. Note that for matrix **A** the background intensity of each block is proportional to the density of connections within that block, and for matrix **B** only the values corresponding to linked pairs of nodes are shown. The nodes are grouped into partitions corresponding to the space-independent modules of the network with the boundaries indicated by solid lines. The relatively large positive values clustered along the diagonal blocks of **B** indicate the occurrence of significantly higher density of connections within each module, compared to that expected from the degrees of, and the distance between, every pair of nodes.

Figure 3.3 ***(previous page)*:** This characteristic signature of modularity is also visible in the adjacency matrix **A** representing the connection topology, suggesting that the mesoscopic structure of the brain network is a consequence of factors beyond the constraints associated with physical distance. Indeed, this is also true for the spatial clustering of nodes in each network module seen in Fig. 3.1 (a-c), as is apparent from the distance matrix showing that the modules comprise many nodes that are spatially proximal even after discounting the effect of distance in identifying the modules. (c) Visual representation of the correspondence between the network modules determined using exclusively information about the connection topology ("Original") and those obtained from *space-independent* partitioning of the network into communities. on the right). This alluvial diagram has been created using the online visualization tool RAW [177].

To see if the restriction on long-range connections implied by the above constraint is responsible for the mesoscopic organization of the network we have reported here, we investigate whether the network can be partitioned into modules even after taking into account the distance dependence of the connection probability in the null model (see Methods for details). Thus, if the modules are exclusively a product of the distance constraint, the deviation of the empirically obtained connection probabilities from those of the null model will be minimal, yielding a single partition comprising the entire network (see Fig. B12-B13 in Appendix B for results on different surrogate networks). In contrast to the above scenario, we find that applying the method on the brain network yields an optimal partitioning comprising seven space-independent modules indicated by the diagonal blocks demarcated by white lines in the adjacency matrix shown in Fig. 3.3 (b) [left surface]. The probability of connections within these modules deviate strongly from the values expected from the null model as shown by the modularity matrix [Fig. 3.3 (b), right surface]. The distance matrix [Fig. 3.3 (b), top surface] also appears to suggest that regions belonging to the same module are, in general, physically closer to each other than those belonging to different modules. However, this physical proximity cannot provide a causal explanation for the modular structure as, even after filtering for spatial effects, the resulting space-independent modules are substantially similar to those reported in the previous subsections [see Fig. 3.3 (c)]. The similarity between the results of these two different modular partitionings is quantitatively indicated by the corresponding normalized mutual information $I_{\mathrm{norm}}(= 0.6)$ [see Methods]. Thus, the spatial layout of the brain regions cannot by

themselves explain the mesoscopic organization of the network, and the existence of the structural modules is a fundamental attribute of the brain.

3.4 Discussion

Despite differences in the details of their organization, the modules that we have identified in the Macaque connectome have common structural features. Most notably, each of them have cortical and thalamic components with the sole exception of module #3, suggesting a distinct functionality of this module. The sizable thalamic contribution to modules #2, #4 and #5 can be understood in terms of the roles that their cortical components play in processing specific sensory modalities. In particular, the information from the corresponding sensory organs arrive at the cortical regions belonging to these modules via relay centers located in the thalamic component of the respective modules. This, however, cannot explain the sizable contribution from thalamic regions to module #1, as the sensory modalities it is associated with, namely, olfaction and gustation, do not involve any thalamic relay. As one of the primary functions of this module is the integration of information processed in different cortical regions (as mentioned earlier), it suggests that regions in the thalamic component of this module serve as relay centers coordinating inter-cortical communication.

The module with which a particular brain region is associated may also alert us to possible functions of this region that have not yet been identified. As an example we consider multi-modal association areas, which integrate and process inputs from different sensory modalities (such as the regions *LIP*, *MIP* and *area 46*). Using information about their modular membership, we can identify which modality or function each of these regions are most strongly associated with. This is illustrated by considering the *LIP*, *VIP*, *AIP* and *MIP* areas of the *Intraparietal Sulcus*. Although they are all multi-modal association areas, *LIP* and *VIP* are part of module #5, whereas areas *AIP* and *MIP* are part of module #2. It is known that *LIP* and *VIP* are involved in visual attention and saccadic eye movements [207, 208, 209], which are predominantly visual processing tasks (consistent with the

broad function of module #5). In contrast, *AIP* and *MIP* coordinate the visual control of reaching and pointing [210, 211, 212], which, although guided by visual information, is primarily a motor function (consonant with the broad function of module #2). Thus, the specific functionalities of these association areas seem to tie in with the modules that they belong to.

The modular nature of the brain has been long recognized, both in terms of function and, more recently, in the topological organization of its structural connections [19]. Considerable attention has been focused on the question of structure-function convergence in the context of brain modules [77]. The hypothesis of "information encapsulation", whereby it is assumed that the information processing related to specific functions are relatively unaffected by those corresponding to other functions, has been suggested as an explanation of how functional modules can arise from the structural organization of the connectome into several communities [213]. Although this may appear intuitive because spreading processes are generally fast within a module and slow down during their passage to a different module [154], we find on the contrary that the specific modular organization of the Macaque connectome allows signals to spread very fast. In fact, the communication of information across the empirical network appears to be even faster than that seen in equivalent networks whose connections are distributed homogeneously. This is surprising as homogeneous networks tend to exhibit the fastest speed of propagation globally, which usually tends to reduce once mesoscopic structural features such as modularity are introduced [154]. We connect this counter-intuitive result to the detailed meso-level attributes of the topological organization, specifically the roles played by different brain regions in terms of their intra- and inter-modular connections. By analyzing these connections we reveal distinctive features of the connectome, namely, the tendency of provincial hubs within a module to connect to each other, and the preference shown by connector hubs to link to peripheral nodes across different modules.

We would like to note that our work parallels the viewpoint proposed in several earlier studies that cognitive processes in the brain necessarily involve integration of information across modalities and functionalities [214, 215, 216]. Thus, behavioral studies have been

used to show that processes such as attention and perception involve interaction and integration among several functional modalities, for example, spatial cognition in humans interacting with semantic cognition of language [217]. In particular, cross-modal integration has been investigated extensively in the context of semantic comprehension [218], where the simultaneous processing of multiple inputs is believed to operate under mutual constraint satisfaction [219], such that probability estimates of each input constrains the estimates of the other inputs. Neurocomputational frameworks such as the 'hub and spoke model' [220, 221, 222] have been used to suggest that modality specific sources of information (spokes) are integrated in a transmodal hub to generate conceptual knowledge. There is also substantial evidence in support of integration across different modalities occurring in non-human primates [223, 224, 225, 226]. Our results show that the mesoscopic organization of the modules, which individually are reminiscent of the information encapsulation perspective, can nevertheless enhance communication globally across the connectome, thereby promoting information integration [214].

While the potential of rapid communication between different regions, made possible by the underlying modular architecture of the network, suggests a plausible explanation for the evolution of the observed mesoscopic organization of the macaque brain, it could also plausibly be a consequence of optimizing for wiring lengths. However, we have explicitly shown that the constraint imposed by the physical distance between the brain regions is insufficient to explain the modular partitions observed by us. Indeed, although the five modules of the connectome that we have identified comprise brain regions that are, for the most part, spatially proximal, module #4 is a prominent exception. It spans two widely separated locations in the brain, one comprising the primary and secondary auditory areas which are in the temporal lobe and the other consisting of association areas located in the prefrontal lobe. While it is well-established that the temporal lobe regions belonging to this module contribute to its associated sensory modality, viz., auditory processing, it is not entirely clear what role the prefrontal regions of this module plays in this context. We note, however, that there are intriguing parallels between these areas and those occupying corresponding locations in the human brain. Specifically, the prefrontal and temporal

parts of module #4 that are known to have a role in social cognition in primates [227, 228] correspond to the Broca's and Wernicke's areas in the human brain, respectively. As is well known, the former is responsible for speech production in humans, while the latter is critical for language comprehension [183]. Although there is no direct counterpart of language in Macaques, non-human primates are known to be capable of communicating through signals such as facial expressions and vocalizations [229]. This correspondence therefore warrants consideration of whether some of the areas in module #4 of the Macaque brain developed from a common evolutionary precursor of the apparatus responsible for facilitating language in humans. Indeed, this view is supported by recent research [230, 231, 232] that have used language-like behavior in non-human primates as models for understanding how speech and language might have evolved in humans [233].

Chapter 4

Invariances and diversity in the human structural connectome : A cohort study

4.1 Introduction

One of the key goals of neuroscience from its very inception has been to unravel the workings of the human brain. Not only is this of great scientific interest and significant from a philosophical perspective, but it is important also for informing clinical and psychiatric practice. Studying the nervous systems of non-human model organisms do allow us to gain an understanding of fundamental aspects of their development, structure and function. Moreover, this has provided us with numerous insights on how a system as complex as the brain could have evolved, and the associated emergence of behavior such as cognition. However, there are limitations to how phenomena observed in relatively simpler nervous systems can be generalized to those with much higher complexity. For instance the ocular dominance columns of the visual cortex that occur in monkey and cat brains are not seen in mice or rat brains. [234]. Similarly, the anatomical and functional organization of a macaque brain is vastly different from that of a human brain [235]. In fact there

are several fundamental aspects of human brain structure and function that are unique to the species [236, 237]. Therefore, in spite of the ethical and technological bottlenecks that hinder the study of the human brain to the level of detail and precision as can be achieved in other mammals and invertebrates, it is crucial that techniques for analyzing and interpreting the structure and function of human brains at multiple length and time scales continue to be developed and refined.

One of the primary approaches that is commonly used to study a brain is to describe its macro-scale connectome [28], i.e., the structure of connectivity between distinct brain regions through axonal tracts. Several studies have pointed to the essential role of such a "wiring diagram" of the nervous system as a foundational model in understanding functional localization at multiple levels, ranging from molecular and cellular up to systems and behavioral levels [31]. As shown in earlier chapters of this thesis, the wiring diagrams of neuronal connectivity can provide insights on the processes that underlie brain development, as well as, the functional implications arising from their structural organization. However, one of the characteristic features of the human structural connectome is its large variability across individuals [28, 238, 239] and over time [240]. This variability in the structural connectome could be a key factor in understanding the generative mechanism underlying the development of brain structure [241]. The diversity in structural connectivity is significantly smaller than that of functional connectivity [242, 243], which is determined by temporal correlations between the electrophysiological activity of different brain regions. These variabilities are important in studying the relationship between structure and function in human brains, as not only is structural connectivity known to affect brain function [244, 23], but function has been shown to influence the structure as well [245, 246]. However, despite the high variability of brain networks in humans, it has been observed that certain structural features are universal [3]. Thus, it is pertinent to ask whether we can describe a typical "representative" structural connectome for a human brain. Such a representative network might not only be useful in studying fundamental aspects of the structure-function relationship in the human brain, but also the deviations from this "basic plan" in a connectome might reveal structural correlates of functional

impairments leading to clinical disorders.

In this work, we study a large ensemble of brain connectivity networks that were obtained from a cohort of 196 healthy human subjects through diffusion tensor imaging for the purpose of characterizing the variability in the structural connectome. We find that there is a correspondence between the diversity of topological connectivities and the variation in the distribution of connection weights, suggesting that the generative mechanisms giving rise to the "wiring" and those determining the weights are related. We further find that the connection strengths of links that are frequently found in the population are described by link-specific Poisson processes, indicating that the generative mechanism of a significant portion of the brain connectivity might involve independent discrete random processes. This allows us to reassign the link weights of structural connectivity, which would represent the discrete Poisson variables instead of original weights and thus obtain the rescaled weight matrices. Using the corresponding resting state functional connectivities obtained from the same cohort, we show that the structural connectome with rescaled weights consistently show better correspondence with the functional connectivity, suggesting that the rescaling process might provide a more informative framework for interpreting the structural connectivity in terms of function. This also provides us with a means for determining the generic "representative" network describing a human connectome. Finally, we show that the representative network is intrinsically resolved into two components, one which is invariant across the population and another that exhibits a much higher degree of variability across individuals.

4.2 Materials and Methods

Connectivity Data.

The human brain structural and functional connectivity dataset analyzed here has been derived from the *Nathan Kline Institute (NKI) / Rockland Sample* [247] - a publicly available repository of diffusion tensor imaging (DTI) and resting state functional magnetic resonance imaging (rs-fMRI) data - which was further processed into connectivity

matrices and made publicly available in the *UCLA multimodal connectivity database* at `http://umcd.humanconnectomeproject.org/` [248].

The data comprises structural connectivity matrices W and functional connectivity matrices C obtained from 196 healthy human subjects: 114 male and 82 female, with ages ranging from 4 to 85 years. Each matrix describes a network comprising 188 nodes that represent 188 brain regions defined by parcellation of the entire gray matter region of the human brain (cerebral cortex, sub-cortical areas, cerebellum, brain stem etc.) using an fMRI based clustering method [249]. It also contains the 3-dimensional coordinates for each of the brain regions in a standardized space. For structural connectivity (SC) matrices W, the connection strengths W_{ij} corresponding to the weighted undirected links (i, j) represent the density of axonal bundles between brain regions i and j as obtained from DTI, while the connection strengths C_{ij} in functional connectivity matrices C represent the Pearson's correlation coefficients between the time-series of dynamical activities in regions i and j, as measured through blood oxygen level dependent (BOLD) imaging using fMRI.

Rescaling to obtain Poisson distributed link weights.

A Poisson distribution of mean λ for a random discrete variable X is given by:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}\,, \tag{4.1}$$

where the parameter characterizing the distribution $\lambda = \langle X \rangle = Var(X)$. Here the link weights W_{ij} for a link between a pair of regions (i, j) are considered to be obtained by rescaling Poisson distributed variables $\mathcal{W}_{ij}$ that have mean λ_{ij}, as $W_{ij} = s_{ij}\mathcal{W}_{ij}$. Such a rescaled Poisson variable has also been described in [250]. The mean and variance of the rescaled Poisson variables W_{ij} are given by:

$$\begin{aligned} \langle W_{ij} \rangle &= s_{ij}\lambda_{ij}\,, \\ Var(W_{ij}) &= s_{ij}^2\lambda_{ij}\,, \end{aligned} \tag{4.2}$$

For a given distribution of weights W_{ij} of a link (i, j), we determine s_{ij} from Eq. (4.2)

$$s_{ij} = \frac{Var(W_{ij})}{\langle W_{ij} \rangle} \tag{4.3}$$

Upon obtaining the rescaling factor s_{ij} for a link (i, j), we rescale the weights W_{ij} across the population to obtain the Poisson distributed rescaled weights:

$$\mathcal{W}_{ij} = \lfloor \frac{W_{ij}}{s_{ij}} + 0.5 \rfloor \tag{4.4}$$

where $\lfloor x + 0.5 \rfloor$ gives nearest integer of x. The rescaled weights are related to the Poisson parameter by $\langle \mathcal{W}_{ij} \rangle = \lambda_{ij}$.

Goodness of fit.

Assuming a rescaled Poisson distribution for all links (i, j), we initially calculate the Poisson parameter λ_{ij}, rescaling factor s_{ij} and the rescaled weights $\mathcal{W}_{ij}$ for each link. We then use a Pearson's Chi-squared test [251] to determine whether the rescaled weights obtained using the method described above fits a theoretically expected Poisson distribution $\mathcal{P}(\lambda_{ij})$ for the corresponding λ_{ij} with significant likelihood. First, we calculate the test statistic χ^2_{ij} for the rescaled weight frequency distribution of each link:

$$\chi^2_{ij} = \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k}\,, \tag{4.5}$$

where n is total number of bins, O_k is number of observations having $\mathcal{W}_{ij} = k$ and E_k is the theoretically expected number of observations, assuming a Poisson distribution for $\mathcal{W}_{ij}$. Here, E_k is obtained from $P(\mathcal{W}_{ij} = k)$ as: $E_k = \lfloor N * P(\mathcal{W}_{ij} = k) + 0.5 \rfloor$, where N is the total number of connectomes analyzed. If a bin has $E_k < 5$, it is merged with the adjacent bins, thus reducing the total number of bins. To ensure the validity of the test, we require that the final number of bins $n \geqslant 3$. If the final number of bins is less than 3, which may arise in the case of links with very low λ_{ij}, we consider the link to be too rare for this statistical test. When the number of bins are sufficient, we compare the statistic χ^2_{ij}

with the Chi-squared critical values for the upper tail one-sided test with significance value $\alpha = 0.01$, obtained from `https://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm`. Links having χ^2_{ij} values less than the corresponding critical values, one cannot reject the possibility that they are from a Poisson distribution.

Partial fitting by excluding outlier data from deviating links.
For each link (i, j) that deviated from a Poisson distribution, we performed an iterative process where at each step the data point with the largest value of $\mathcal{W}_{ij}$ was removed, new values of λ_{ij}, s_{ij} and $\mathcal{W}_{ij}$ were calculated and the Chi-squared test was performed on the reduced dataset. This sequential process is terminated when the distribution of the reduced dataset is found to fit a Poisson distribution, or once as many as 20 data points (10%) have been removed. Through this process, we determine the number of links that are Poisson distributed over at least the bulk ($\geqslant 90\%$) of the population. These links, together with the links fitting Poisson distribution over entire population, are considered to comprise the "representative" structural connectivity for a human brain, with connections weights being λ_{ij}.

Generating surrogate ensemble of finite size populations of brain networks.
In order to quantify the role of finite size effects and the specific distribution of λ_{ij} values in the observed deviation of some of the links from the Poisson distribution, we created a surrogate ensemble of 1000 populations, each population containing the same number (196) of structural connectomes as in the empirical dataset. For each link (i, j) in a surrogate connectome, the link weight was drawn from the Poisson distribution $\mathcal{P}(\lambda_{ij})$ for which we used the `poissrnd` function in *MATLAB Release 2010b*. For each population, we then determined the fraction of links that deviated from the Poisson distribution using the Chi-squared test described above. The distribution of the fraction of deviating links f_{dev} provides a measure of the extent to which apparent deviation of the link weights from Poisson distribution, may arise from finite size effect.

Simulated functional connectivity obtained via dynamical model for neural population activity.

In order to obtain the functional connectome resulting from the dynamics of the structural brain network, we use the Wilson-Cowan (WC) neural mass model [252, 253] to describe the activity in each node of the structural connectome. The temporal evolution of the mean activity of excitatory (u_i) and inhibitory (v_i) subpopulations of node i is given as:

$$\begin{aligned} \tau_u \dot{u}_i &= -u_i + (\kappa_u - r_u u_i)\mathcal{S}_u(u_i^{in}) \,, \\ \tau_v \dot{v}_i &= -v_i + (\kappa_v - r_v v_i)\mathcal{S}_v(v_i^{in}) \,, \end{aligned} \tag{4.6}$$

where $u_i^{in} = c_{uu}u_i - c_{uv}v_i + \sum'(w_{ij}^{uu} - w_{ij}^{uv}) + I_u^{ext}$ and $v_i^{in} = c_{vu}u_i - c_{vv}v_i + \sum'(w_{ij}^{vu} - w_{ij}^{vv}) + I_v^{ext}$ represent the total input to the excitatory and inhibitory subpopulations respectively. Here, $c_{\mu\nu}(\mu, \nu = u, v)$ represents the interaction strengths within and between the subpopulations of a node while $\tau_{u,v}$ and $I_{u,v}^{ext}$ correspond to the time constants and the external stimuli for each of the neural subpopulations. The interaction strength between the subpopulations of different nodes are represented by $w_{ij}^{\mu\nu}(\mu, \nu = u, v)$, which are obtained from the connection weights of the structural connectome of each individual. We assume that all inter-nodal interactions are of equal strength: $w_{ij}^{\mu\nu}(\mu, \nu = u, v) = w_{ij}$. The summation $\sum'$ is performed over all neighbors of the structural network. The sigmoidal response function $\mathcal{S}_\mu(z) = [1 + \exp\{-a_\mu(z - \theta_\mu)\}]^{-1} + \kappa_\mu - 1$ has a maximum value $\kappa_\mu = 1 - [1 + \exp(a_\mu\theta_\mu)]^{-1}$. Parameters are chosen such that the dynamics of isolated nodes ($w_{ij}^{\mu\nu} = 0$) are in the oscillatory regime. The matrices corresponding to the inter-nodal coupling weights w_{ij} are taken to be scalar multiples of individual structural weight matrices W for one set of simulations and individual rescaled weight matrices $\mathcal{W}$ for another set of simulations. We also used other structural connectivity matrices such as the adjacency matrix corresponding to the rescaled weight matrices $\mathcal{A}$ and the two types of representative structural network matrices $\langle W \rangle$ and Λ. The corresponding functional connectivity was obtained by finding Pearson's correlation coefficient between the time series of u_i and u_j for each pair of nodes (i, j). For the sake of comparison, we multiplied all the structural connectivity matrices

(W, $\mathcal{W}$, $\mathcal{A}$, $\langle W \rangle$ and Λ) with corresponding normalizing constants, such that the mean connection strength averaged over each network was always a constant w^{avg}. We fixed the value $w^{avg} = 100$ since at this value of average coupling we obtained temporal activity that is qualitatively very similar to typical empirically observed fMRI time series.

Bimodality coefficient

The bimodal nature of a probability distribution can be characterized by calculating its bimodality coefficient [116]:

$$BC = \frac{m_3^2 + 1}{m_4 + 3 \cdot \frac{(n-1)^2}{(n-2)(n-3)}}, \tag{4.7}$$

where m_3 is the skewness, m_4 is the excess kurtosis and n represents the sample size. A distribution is considered to be bimodal if $BC > BC^*$, where $BC^* = 5/9$. This benchmark value corresponds to a uniform distribution, and if $BC < BC^*$ the distribution is considered unimodal.

Statistics.

The *Kernel smoothened density function* [121] has been used to estimate the probability distribution functions of different quantities (e.g., the joint probability between f_{ij} and W_{ij}). For this purpose we have used the `ksdensity` function in *MATLAB Release 2010b* with a Gaussian kernel.

The *Two-sample Kolmogorov-Smirnov (KS) test* [120] has been used to compare between pairs of samples (e.g., matrix correlations of the functional and structural connectomes) in order to determine whether both of them are drawn from the same continuous distribution (null hypothesis), or if they belong to different distributions. For this purpose we have used the `kstest2` function in *MATLAB Release 2010b*, with the value of the parameter α which determines threshold significance level set to 0.01.

4.3 Results

"Wiring" and "weighting" are not independent processes.

The structural connectivity of a human brain displays large variability across individuals in a population, in terms of both connection topology and weight distributions. However, not all the links in the structural network exhibit the same extent of variability. In our analysis, we consider an ensemble of 196 structural connectivity (SC) matrices, as illustrated in in Fig. 4.1 (a), which allows us to study this variability across a diverse human population (see Methods for details about the connectivity data). Each network contains 188 nodes, and the density of axonal bundles between regions i and j is represented by the connection strength W_{ij}. Fig. 4.1 (b) shows the SC matrix for one of the individuals, where the regions are ordered alphabetically and grouped into three broad regions: brain-stem (BS), left brain and right brain. Not surprisingly, we observe a considerably higher density of ipsilateral connections than contralateral ones. In order to characterize the topological variability of the network across the population, for every link (i, j) we measure the relative frequency of occurrence f_{ij}, which is given by the fraction of the total population in which an axonal tract between regions i and j is observed. For all the ${}^{188}C_2$ node pairs that can in principle have a connection we obtain the f_{ij} values between 0 to 1, where 0 signifies those links that are never observed in any individual and 1 identifies links that are found in every member of the population. The connection topology is largely determined during the course of development through the process of wiring, where a complex cascade of genetic and molecular mechanisms determine the probability of connection between two neurons, as mentioned in Chapter 2. However, there remains uncertainty as to the exact processes that determine the connection weights at the level of the brain regions that are connected through axonal tracts. Note that the connection weights in this case are not equivalent to the weights of synaptic connections, whose strengths are governed by various learning and plasticity mechanisms. Here the connection weights are actually the observed density of axonal bundles, which is partially related to physical thickness of the connections.

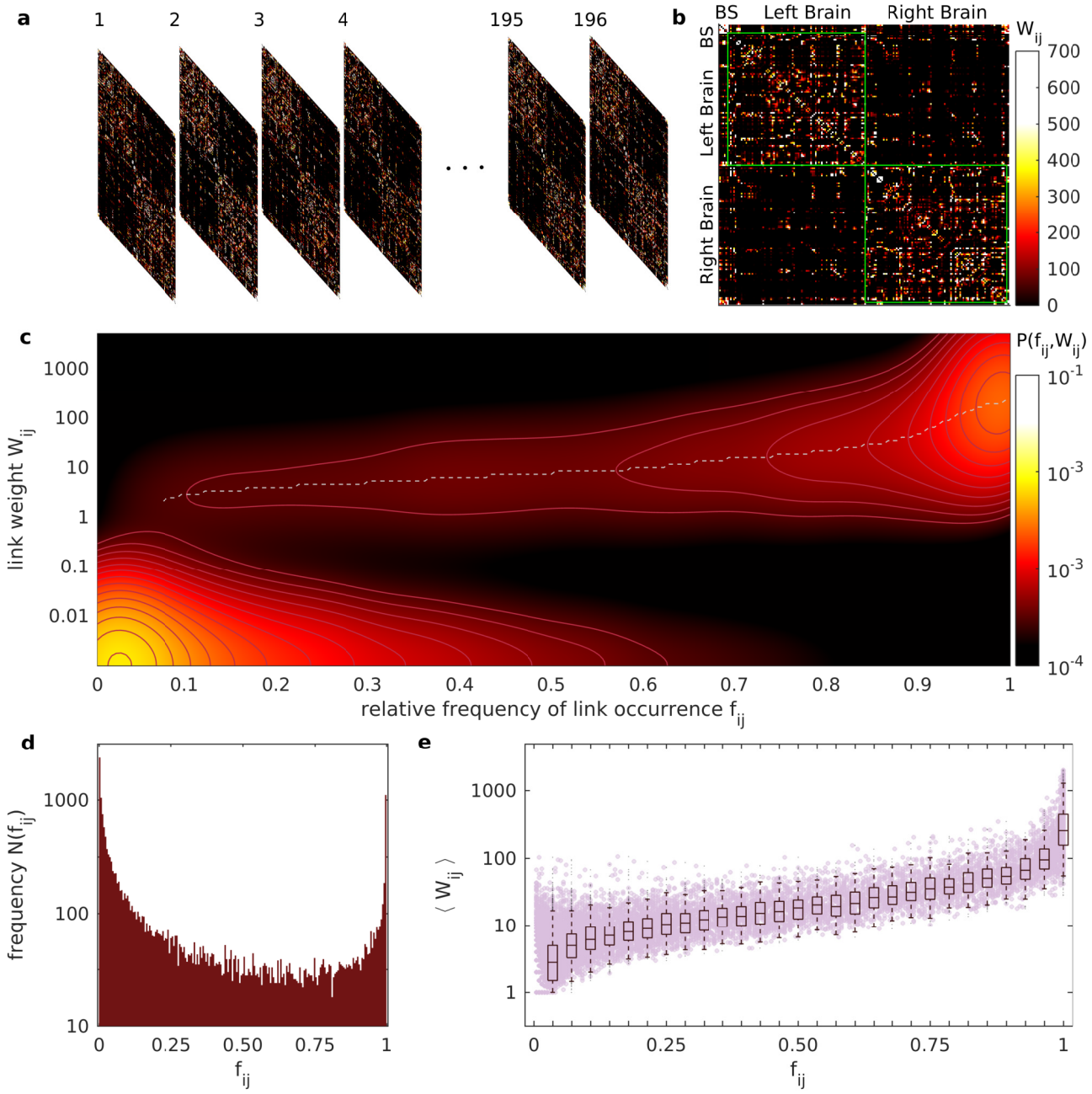


Figure 4.1: An increase in the occurrence of non-ubiquitous connections within a population leads to a steady shift in the corresponding link weight distribution towards higher values. (a) Ensemble of weighted structural connectivity (SC) matrices representing the structural brain networks of 196 human subjects obtained via diffusion tensor imaging (DTI). (b) A sample SC matrix corresponding to the connectivity information obtained from one of the subjects.

Figure 4.1 ***(previous page)*:** Each network comprises 188 nodes that correspond to brain regions connected through white matter tracts, which are represented by weighted undirected links in the SC matrix. For each network the matrix entries W_{ij} represent the density of axonal bundles between nodes i and j. The minimum possible link weight is 1, and matrix entries are set to 0 if the connection does not exist or cannot be detected due to extremely low thickness.The regions are alphabetically arranged and grouped into the brain-stem (BS), the left brain and the right brain. Notice the relatively high density of ipsilateral connections (connections between regions of the same hemisphere, as indicated by diagonal blocks in the matrix) and low density of contralateral connections (connections between regions of opposite hemispheres, as indicated by off-diagonal blocks). (c) Joint probability distribution $P(f_{ij}, W_{ij})$ of the relative frequency f_{ij} that a link between brain regions i and j is seen across individuals in the population and the weight W_{ij} of the link. If a link does not exist we set $W_{ij} = 10^{-3}$ and subsequently use kernel smoothing to obtain the distribution $P(f_{ij}, W_{ij})$. We observe a steady increase in the mode of the distribution of link weights (represented by the broken white curve) as the frequency of occurrence f_{ij} increases, with a much steeper rise after $f_{ij} = 0.9$. Furthermore, the distribution broadens on increasing f_{ij}. (d) Frequency histogram showing the distribution of links over f_{ij} from the set of all possible ${}^{188}C_2 = 17578$ pairs of nodes (i, j), illustrating the variability in connection topology of brain networks across the population. (e) The mean link weights ($\langle W_{ij} \rangle$) for non-ubiquitous links of any given frequency of occurrence f_{ij}, averaged over the sub-population in which they occur, is observed to vary over an order of magnitude with the interval of the range shifting upwards with the increase in occurrence. Each point of the scatter plot represents a link, which provides a link-wise resolution to the distribution of panel (c). The box plots representing the distributions of $\langle W_{ij} \rangle$ over consecutive intervals of f_{ij} clearly illustrate a steady increase of the distribution.

Hence there is no prior reason to expect any correspondence between the ubiquity of a link in the population, as quantified by f_{ij}, and the distribution of its weight across the population.

Fig. 4.1 (c) shows that as the occurrence of links in the population increases, their weights tend to steadily rise, as indicated by the joint probability distribution $P(f_{ij}, W_{ij})$ obtained using kernel smoothing (see Methods). The lower plateau represents the links with weight $W_{ij} = 0$ and the upper plateau, which includes link weights from 1 up to order of 1000, not only displays an increase of $P(f_{ij}, W_{ij})$ with increasing occurrence (which trivially leads to increase in non-zero weights), but also shows a steady increase of W_{ij}. This indicates that those connections which are found more often in a population are likely to have higher connection weights as well. Fig. 4.1 (d) displays the frequency histogram of f_{ij} values, showing the highest concentration of links at 0 and 1. For a species with

rigidly invariant topology over a population, such as *C. elegans*, this histogram would show occupancy only at 0 and 1. Thus a large number of links which occur with frequencies 0 and 1 implies high topological variability in the structural connectivity within a population. Even though the distribution of connection weights in the population shows correspondence with the relative frequency of occurrence f_{ij}, the distribution widens with an increase in the occurrence. This implies a large variability in connection weights, even for links with similar values of occurrence. Assuming that the processes determining weights are still largely independent of the processes determining the formation of the connection itself (wiring), it is meaningful to consider the weight distribution of non-ubiquitous links ($f_{ij} < 1$) over only those individuals from the population in which the link occurs. Fig. 4.1 (e) shows that even when we consider mean weights of each link ($\langle W_{ij} \rangle$) averaged over the corresponding subset of the population in which the link occurs, there is a steady rise in these partially averaged values of the link weights. This further indicates that wiring and weighting processes for the links are not entirely independent, even if they are separate processes.

The variation of the weights of frequently occurring links over the population, as well as their frequency of occurrence, can both be described by a single link-specific Poisson process.

One of the simplest stochastic processes that describes the distribution of the number of recorded events is the Poisson process. It corresponds to a probability distribution of discrete random independent events occurring at a constant rate over time or space. Here we consider the hypothesis that the link weights W_{ij} for link (i, j) are generated by such discrete independent events occurring at a constant rate λ_{ij} for each human subject. In such a case, the link weights W_{ij} would follow a Poisson distribution. As we have already shown above that the processes determining the wiring and the weights of the links appear to be at least mutually dependent, even if they are separate, weights having Poisson distribution would actually mean that a single parameter λ_{ij} would be sufficient to explain the variability of link weights as well as their frequency of occurrence.

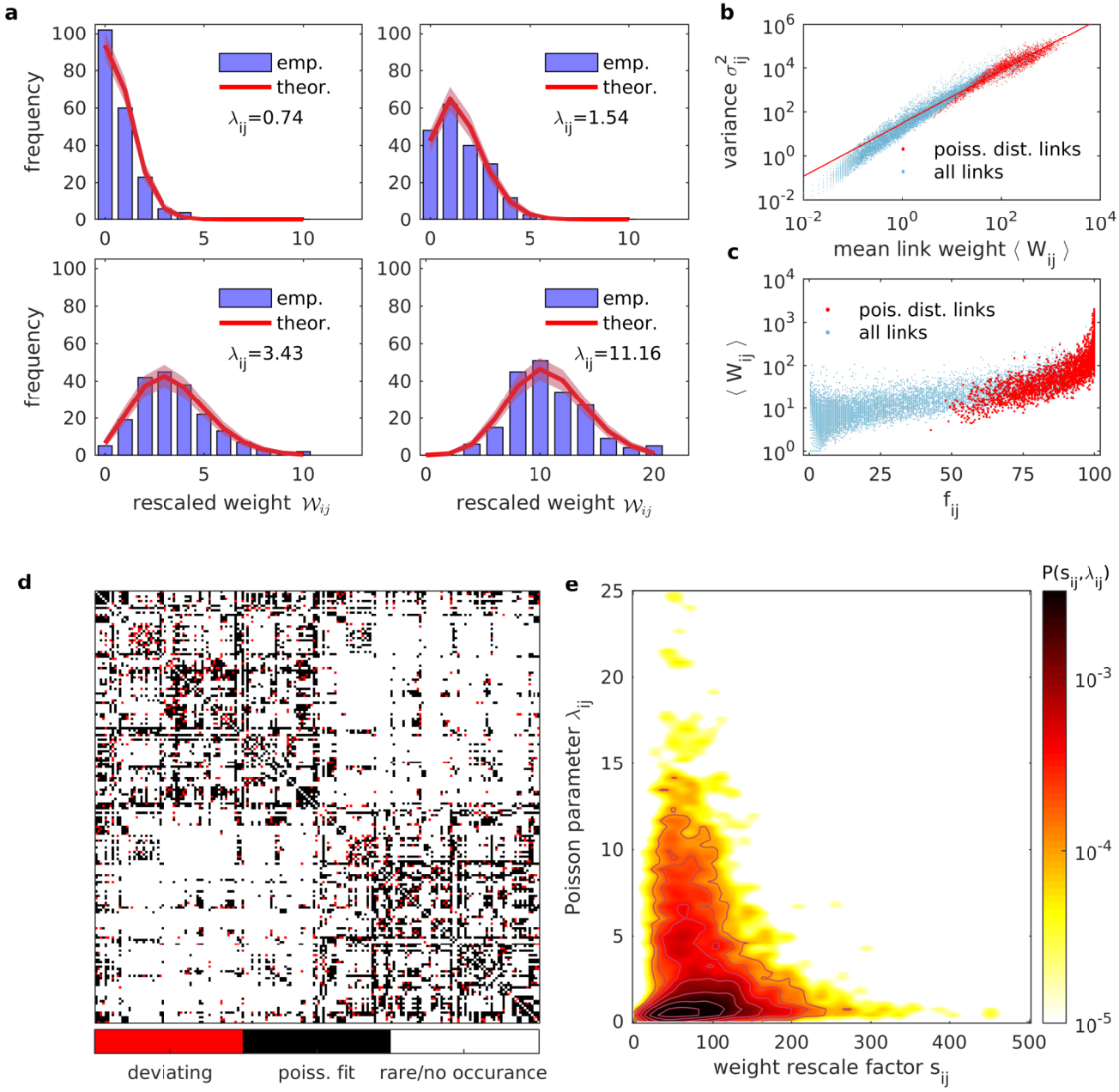


Figure 4.2: The variation of the weights of frequently occurring links over the population, as well as their frequency of occurrence, can both be described by a single link-specific Poisson process. (a) Frequency histograms corresponding to four separate links (i, j) demonstrating that their rescaled link weights $\mathcal{W}_{ij}$, obtained using corresponding rescaling factors s_{ij}, are distributed over the population in a Poisson process, with respective Poisson parameters λ_{ij} displayed in each case.

Figure 4.2 ***(previous page)*:** The solid lines, which represent the theoretical values of the distributions, display close agreement with the histograms corresponding to the observed distributions and the shaded region represents the fluctuations in the frequencies over a large number of randomly drawn samples of size 196 (same as the population size) from the corresponding Poisson distribution. The goodness of fit for the links with the Poisson distribution has been quantitatively tested using Pearson's Chi-square test [251]. (b) Scatter plot of the links, represented in terms of their mean weights across entire population $\langle W_{ij} \rangle$ and the their variances σ_{ij}^2. It can be seen that for most links the weight distributions have $\sigma_{ij}^2 \propto \langle W_{ij} \rangle$, which indicates the possibility of a rescaled Poisson distribution. The links that fit a rescaled Poisson distribution (as shown by Chi-squared goodness of fit test with significance $\alpha = 0.01$), are distinguished by showing them in red color and the linear regression fit for those points (red line) has a slope of 1.2. (c) Links which are shown to fit rescaled Poisson (red dots) have high values for the relative frequency of occurrence (> 0.5) as well as higher values of mean link weights. (d) Adjacency matrix showing the entries corresponding to 3342 links that fit the rescaled Poisson distribution with $\alpha = 0.01$ (black entries) and 577 links whose weight distributions deviate from the corresponding Poisson distributions (red entries). (e) Joint probability distribution $P(s_{ij}, \lambda_{ij})$ of the distribution of rescaling factors s_{ij} and Poisson parameter λ_{ij}. We observe that $P(s_{ij}, \lambda_{ij})$ is relatively unaffected by increasing or decreasing λ, which suggests that both these link properties that together determine the observed weights for a link in the population, might originate from distinct biological factors.

We find that for a large number of frequently occurring links ($f_{ij} > 0.5$), their weights follow a rescaled Poisson distribution (see Methods) with each link having specific value of λ_{ij} and corresponding rescaling factor s_{ij}. The rescaling factor is a link-specific constant scalar value for a link (i, j), which can be applied on its weights W_{ij} across the population to obtain the rescaled weights $\mathcal{W}_{ij}$, which follow a Poisson distribution $\mathcal{P}(\lambda_{ij})$ (see Methods for details). Using Pearson's Chi-squared goodness of fit test (see Methods) to determine whether the rescaled weights for a link fits a Poisson distribution with high significance ($\alpha = 0.01$), we find that out of $15,209$ links that are seen at least once in the population there are 3342 links whose rescaled link weights can be described by Poisson processes and $11,290$ links have too few occurrences for a reliable statistical fitting. Thus there are only 557 links that deviate significantly from the Poisson distribution.

Fig. 4.2 (a) demonstrates that the frequency histogram corresponding to rescaled link weights of four separate links agree with the theoretically expected frequencies from a Poisson distribution having the corresponding λ_{ij}. As both the mean and variance of

Poisson distributions are equal to the same parameter λ, it is expected that the rescaled Poisson distribution followed by W_{ij} has $\langle W_{ij} \rangle \propto \sigma_{ij}^2$, as can be seen in Fig. 4.2 (b). Fig. 4.2 (c) shows that links described by a Poisson process appear to have high significance in the structural connectivity in terms of both the topology (by having higher values of f_{ij}), as well as, the strength of connection (by having higher values of partially averaged mean link weights as described in Fig. 4.1 (e)). The adjacency matrix in Fig. 4.2 (d) illustrates the highly dense network comprising of links described by Poisson processes, while also showing the sparsely distributed 557 links that deviate from the Poisson process. Further, the two factors determining the connection weight for a link, viz. λ_{ij} and s_{ij}, seem to have no dependence on each other, as can be seen in their joint probability distribution in Fig. 4.2 (e). The Pearson's correlation coefficient between respective values of λ_{ij} and s_{ij} is -0.09.

We next examine the 577 links which, in spite of having high frequency of occurrence, do not fit Poisson distribution. We observed that the bulk of these links have a bimodal distribution of rescaled weights, in which the first mode appears to be a Poisson distribution while the other mode (occurring at much higher values) comprises outliers with unexpectedly high weights. Hence the deviation from a purely Poisson process can possibly be attributed to such outliers. The rescaling factor s_{ij} and Poisson parameter λ_{ij} obtained from these links are possibly inaccurate because they are calculated by including the outlier values. Thus, for each of these deviating links, we sequentially remove the data points with the largest weight, calculate the λ_{ij} and s_{ij} for the remaining subset and rescaling the weights until the remaining data agrees with the Poisson process as per Chi-squared goodness of fit with a high significance ($\alpha = 0.01$). Out of 577 deviating links, we find that 520 links fit the rescaled Poisson distribution after excluding less than 10% of the outliers. Fig. 4.3 (a) demonstrates that in four of the deviating links that are shown, the bulk of the distribution ($> 90\%$) agrees with a Poisson distribution, while the outliers ($< 10\%$) deviate significantly from the theoretical distribution. The adjacency matrix in Fig. 4.3 (b) shows that the bulk of the deviating links fit the Poisson distribution by excluding few outliers. We note that out of a total of 3919 links that have sufficient occurrences for statistical

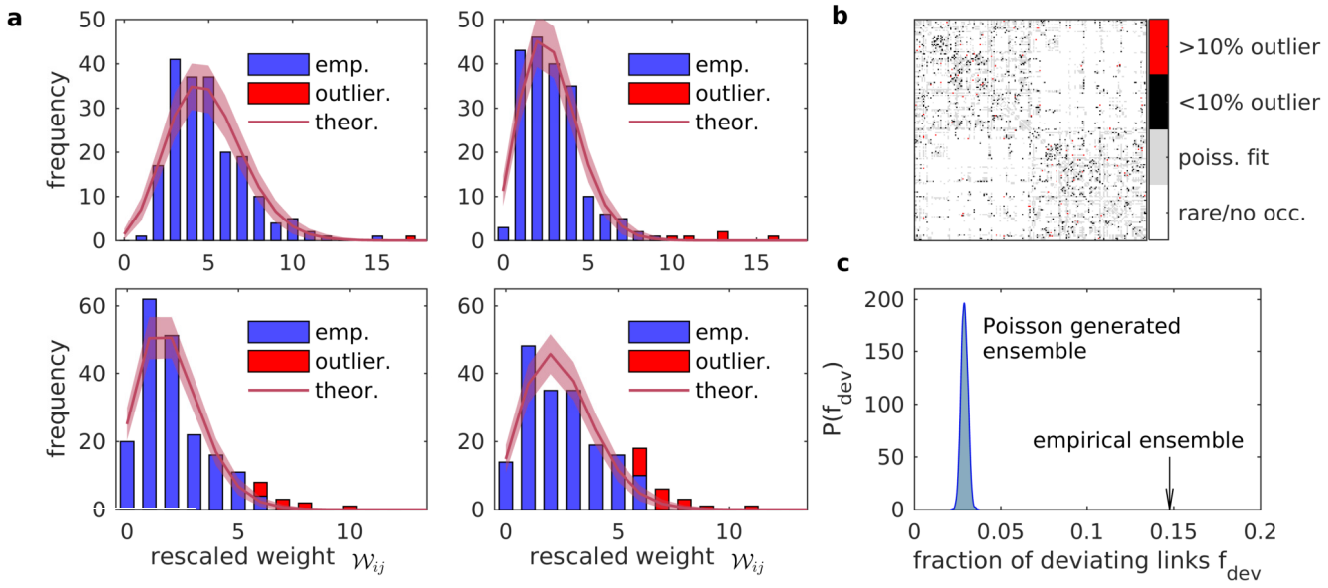


Figure 4.3: Exclusion of a small fraction of outliers can make deviating links fit a Poisson distribution. (a) Frequency histograms corresponding to four separate links (i, j) demonstrating that rescaled link weights $\mathcal{W}_{ij}$ obtained using corresponding rescaling factors s_{ij} are distributed over the bulk of the population in a Poisson process (represented by blue bars), upon excluding a few outliers ($< 10\%$ of the population, represented by red bars) with high values that deviate from the Poisson process. The solid lines which represent the theoretical values of the distributions display close agreement with the histograms corresponding to the observed distributions for the bulk of the population (blue bars) and the shaded region represents the variability in the frequencies across a large number of randomly drawn samples of size 196 (same as the population size) from the corresponding Poisson distribution. (b) Adjacency matrix showing that 520 links (shown as black entries) out of the 577 links that originally deviated from the Poisson distribution over the entire population (as shown in Fig. 4.2 (d)), fit the Poisson distribution on removing only upto 10% outliers from the population, as quantitatively shown using the Chi-squared test. The remaining 57 links do not fit a Poisson distribution even after the removal of 10% outliers (shown as red entries). (c) The empirically observed fraction of links whose weights deviate from the Poisson distribution f_{dev} (out of all the links with sufficient frequency of occurrence, viz.≈ 0.15, as indicated by the arrow), cannot be explained by finite size effect alone. The probability distribution shows the expected values for f_{dev} in a randomly generated network ensemble of sample size 196, where each link weight across all the networks is drawn from a Poisson distribution with link-specific values of λ_{ij} that were obtained from the empirical dataset. Finite size effects contribute a value of $f_{dev} \approx 0.03$ which is much smaller than that observed from empirical data.

tests, 3862 links (98.5%) have their weights distributed according to link specific Poisson processes for more than 90% of the population.

The deviation observed in these 577 links suggests that even though a single Poisson process might be sufficient to describe the wiring and the weights for a large fraction of links for most of the population, there may be other factors in the generative mechanism that lead to a significant fraction of links (≈ 0.15) deviating from a Poisson process, even if that deviation is due to the presence of a few outliers within the population. In order to rule out the possibility that the deviation arises simply because of finite size effects in the specific distribution of λ values, we generated a surrogate ensemble of SC matrix sets, each comprising 196 matrices. The link weights were each drawn from the link-specific Poisson distributions (see Methods for details). We find that in every realization, a small fraction of links does not fit a Poisson distribution, as per the Chi-squared criterion. The distribution of these values is shown in Fig. 4.3 (c). The deviations arising due to such finite size effects are significantly lower than those seen empirically. This suggests that there must be other significant factors in the generative mechanisms apart from the Poisson process that give rise to the structural connectivity in human brains.

Rescaling of link weights using Poisson parameters might provide greater functional interpretability to the structural connectivity.

While we have have thus far interpreted the Poisson parameter λ_{ij} for the links as an abstract representation of the net effect of all biological factors associated with the generative mechanism of the structural connectivity of brain networks, an additional interpretation of the Poisson parameter can be obtained by considering its effect on function. While there have been several attempts to examine the relation between structural connectivity and functional connectivity [244, 23], it has been observed that there is very low correspondence between the two. Unlike synaptic and gap-junction weights which are interpreted as coupling strengths between neuronal activities, the connection weights in the structural macro-connectome elude a functional interpretation (such as for instance being a measure

of dynamical coupling between the activities of different brain regions). We have already observed that rescaled weights appear to be a more fundamental structural property than the original weights, since they are essentially discrete Poisson variables which are known to arise in a wide range of natural phenomena. It is therefore reasonable to ask whether the rescaled weights can have a deeper relation with the functional coupling. For example, could the rescaled weight be directly related to the number of axons in axonal tracts? Here we ask whether the individual structural connectivity of rescaled weights, which are obtained from the analysis of SC matrices across a population, have a stronger correlation with the corresponding functional connectivity as compared to the original structural connectivity.

The *Nathan Kline Institute (NKI) / Rockland Sample* dataset that we consider [247] includes resting state functional connectivities (FC) for each of the 196 individuals whose structural connectivity we have analyzed thus far. Sample functional connectivity matrices are displayed in Fig. 4.4 (a). In contrast to the structural connectivities, the functional connectivities exhibit a large degree of variability across individuals. Examining a single FC matrix (Fig. 4.4 (b)) reveals that the density of strong functional connections is much higher than that of structural connections (Fig. 4.1 (b)), and unlike the structural networks there is a significant number of connections across left and right hemispheres. We first examine the correlation between the structural and functional connection strengths for each link (W_{ij} and C_{ij}, respectively) over the entire population and observe that most of the links have negligible correlation between the variations of structural and functional connectivity strengths over the population (see Fig. D1 in Appendix D). While this supports the previously known observation that there is very low correspondence between structure and function at the level of individual links, we observe that a macroscopic comparison between the FC matrices ($C_{(n)}$) and structural weight matrices ($W_{(n)}$) for each individual n reveals statistically significant correlations, as seen in Fig. 4.4 (c). Furthermore, for every individual in the population we observe that the matrix correlation of $C_{(n)}$ and structural weight matrix $W_{(n)}$ is lower than the matrix correlation of $C_{(n)}$ and the corresponding rescaled structural weight matrix $\mathcal{W}_{(n)}$. Thus, there is a notable correspondence between

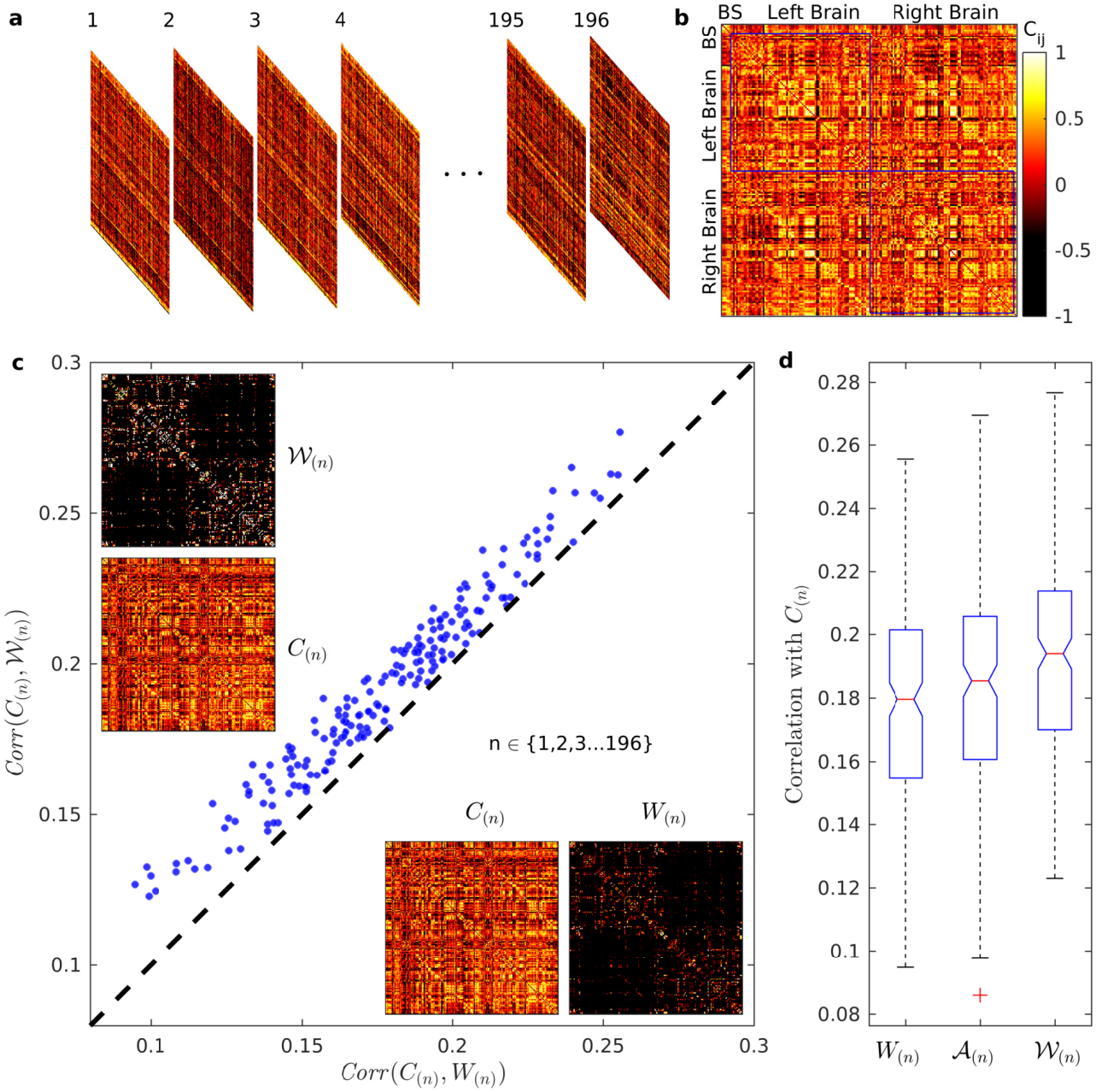


Figure 4.4: Upon rescaling the weights, the structural connectivities consistently show a greater association with the corresponding functional connectivities across the population. (a) Ensemble of functional connectivity (FC) matrices representing the resting state functional brain networks, determined by functional magnetic resonance imaging (fMRI), of the same 196 human subjects whose structural connectivities have been analyzed in this work.

Figure 4.4 ***(previous page)*****:** (b) A sample FC matrix corresponding to one of the subjects. The nodes are arranged in the same way as described in Fig. 4.1 (b), and the matrix entries C_{ij} indicate the correlation between the haemodynamic activities of nodes i and j over a period of time. Notice the striking difference between the SC shown in Fig. 4.1 (b) and the FC shown here. Where the SC have relatively sparse connections and a much higher density of ipsilateral connections than the contralateral ones, the FC have a much higher density of strong connections (both positive and negative) with equal density of functional connectivity between ipsilateral (diagonal blocks) as well as between contralateral regions (off diagonal blocks). (c) A comparison of the dependence between individual functional connectivity matrices ($C_{(n)}$) for an individual n (where $n \in 1, 2, 3 \ldots 196$) and the corresponding structural connectivities described respectively by weight matrices ($W_{(n)}$) and rescaled weight matrices ($\mathcal{W}_{(n)}$). Examples of the three types of matrices are shown as insets. The scatter plot displays the correlation between $C_{(n)}$ and $W_{(n)}$ for each n along the x-axis, and the corresponding correlation between $C_{(n)}$ and $\mathcal{W}_{(n)}$ along the y-axis. Note that while the correlations between functional and structural matrices of both types are very low (< 0.3) (even though statistically significant), for every individual in the population the functional connectivity matrices are more correlated with the corresponding $\mathcal{W}_{(n)}$ than they are with $W_{(n)}$. (d) Box-plots representing the distributions of the correlations of the functional matrix $C_{(n)}$ for each individual n with the corresponding matrices weight $W_{(n)}$ and $\mathcal{W}_{(n)}$ and the adjacency matrix $\mathcal{A}_{(n)}$, generated from the rescaled weight matrix $\mathcal{W}_{(n)}$ by removing information about the link weights. The correlation distribution for $\mathcal{W}_{(n)}$ which is positioned higher than that corresponding to $W_{(n)}$ as expected from the result in panel (c), is higher compared to the correlation distribution for $\mathcal{A}_{(n)}$, which contains information related to the topology of the rescaled matrix but not the link weights. Note that topology of the rescaled weight matrices are different from that of the original weight matrices, since the former has much fewer links. This implies that in addition to the altered topology of rescaled weight matrices, the altered weight distribution also makes $\mathcal{W}_{(n)}$ a better structural correlate of the observed dynamical function of the brain.

structure and function for each individual at the level of the entire brain network, and furthermore, the rescaling of weights in the structural connectivity enhances this correspondence between structure and function in all individuals. A possible explanation for this enhanced correspondence with function might be the alteration in the distribution of weights after rescaling, or it might also be attributed to the change in the connection topology due to the rescaling of weights (as the rescaling process leads to the deletion of links). To examine this, we computed matrix correlations of each $C_{(n)}$ with (unweighted) adjacency matrices having the same connectivity as $\mathcal{W}_{(n)}$, which we refer to as $\mathcal{A}_{(n)}$. The boxplots in Fig. 4.4 (d) shows that the matrix correlations of $C_{(n)}$ with $\mathcal{W}_{(n)}$ are higher than correlations with $\mathcal{A}_{(n)}$, suggesting that the alteration of the weight distribution upon

rescaling is the primary factor in making SC a better correlate of FC. The three distributions of matrix correlation values were found to be significantly distinct from one another using the *Two-sample Kolmogorov-Smirnov (KS) test* (see Methods).

The resting state functional connectivities of the individuals in the dataset are simplified snapshots of a large repertoire of complex dynamical behaviors in the brain, which are associated with various cognitive functions and arise from highly complex non-linear interactions among the constituent brain regions. Therefore a simple correspondence between structural and functional connectivities would not be sufficient to establish that a rescaled weight distribution in structural connectivity would be a better functional correlate than the original weight distribution. In order to understand how the weights in the structural connectivity affect the dynamics, we simulate the complex dynamical activity that would arise when the structural connectivity provides the basis for the non-linear interactions between the brain regions. To describe the dynamical activity of brain regions, we use a well-known neural mass model, viz., the Wilson-Cowan (WC) model [252, 253] (see Methods for details). Fig. 4.5 (a) shows a schematic representation of one WC node which is used to represent the activity in a single brain region. In order to generate the time series of simulated activity of all brain region for each individual, we simulated systems of WC oscillators placed at the nodes of networks with the corresponding structural connectivity matrix (W_{ij}, as well as, $\mathcal{W}_{ij}$), with the associated connection strengths of these matrices taken to be the same as the dynamical coupling strengths w_{ij} between WC nodes. Fig. 4.5 (b) shows samples of such time series. By computing Pearson's correlation coefficient between the time series of activity for each pair of brain regions, we generated the simulated functional connectivity matrices C^W (Fig. 4.5 (c)) and $C^{\mathcal{W}}$ (Fig. 4.5 (d)) for each individual to compare with the corresponding empirical functional matrices C (Fig. 4.5 (e)). We compared the matrix correlations between individual $C_{(n)}$ and $C^{\mathcal{W}}_{(n)}$ with correlations between individual $C_{(n)}$ and $C^{W}_{(n)}$ matrices, as shown in Fig. 4.5 (f). Similar to the result obtained from the comparison of FC matrices with SC matrices in Fig. 4.4 (c), we find that simulated FC generated from rescaled weight matrices $\mathcal{W}_{(n)}$ were consistently better correlated with empirical FC than the ones generated from original

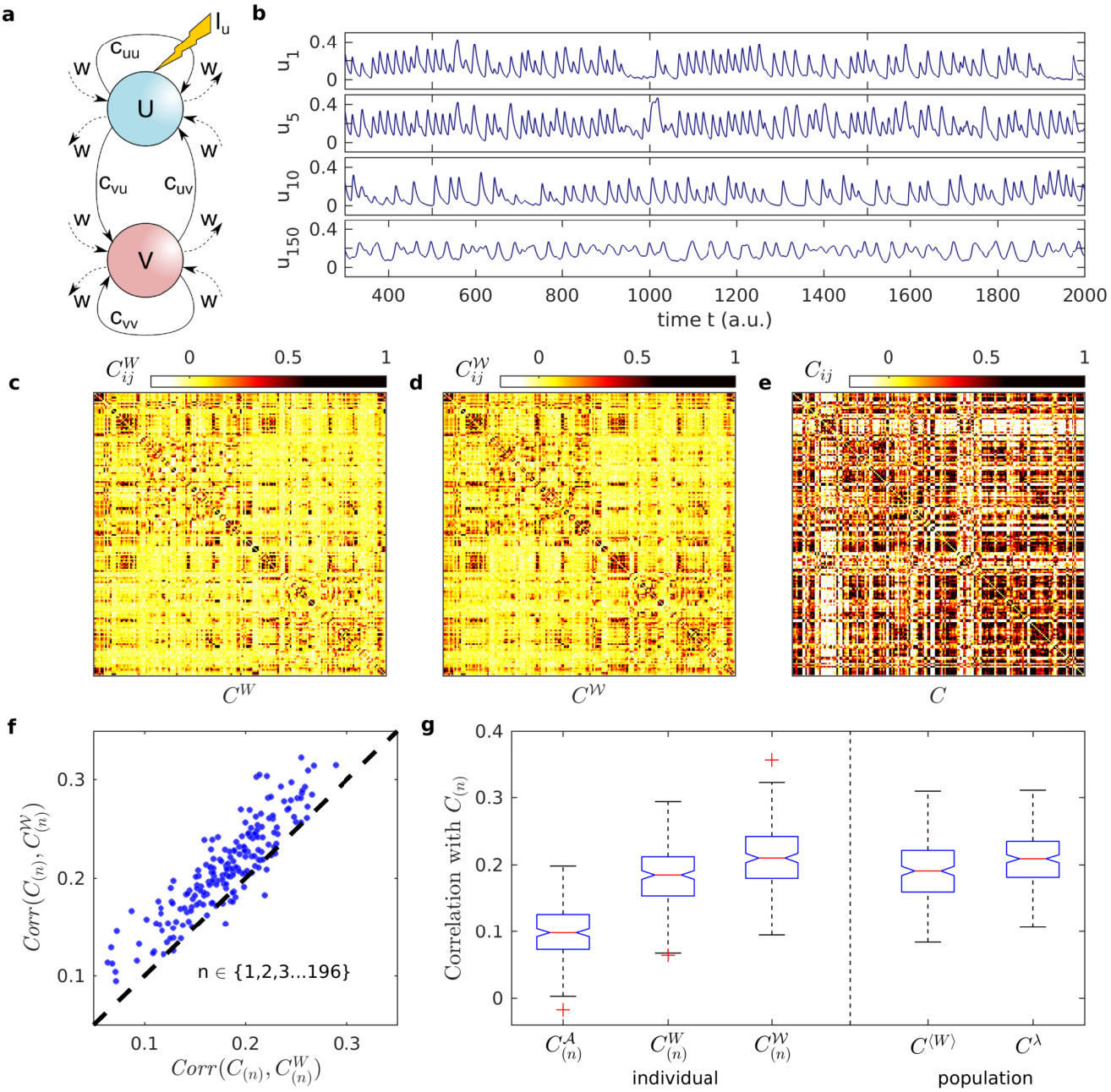


Figure 4.5: Dynamical simulation of whole brain activity using a neuronal population model suggests that structural connectivity with rescaled weights are better structural correlates for brain function than those with original weights. (a) Schematic representation of a single dynamical element (oscillator) of the Wilson-Cowan (WC) model, which simulates the activity within a single region of the brain network, each comprising

Figure 4.5 ***(previous page)*****:** interactions between the excitatory and inhibitory subpopulations (U and V respectively), with strengths denoted by $c_{\mu\nu}$ where $\mu, \nu = \{u, v\}$, and their interactions with the subpopulations belonging to other nodes of the network (with uniform coupling strength w). The lightning bolt represents the external stimulation of strength I_u provided to the excitatory subpopulation. (b) Time evolution of the dynamical activity of the excitatory subpopulations in four out of 188 brain regions of a single individual, obtained from simulations of WC oscillators on the network specified by the structural connectivity matrix associated with the rescaled weights $\mathcal{W}$ as coupling strengths. As the model is dimensionless, here time is displayed in arbitrary units (a.u.). (c-d) Functional connectivity matrices obtained from simulated brain activity using one of the connectomes. The matrices C^W and $C^{\mathcal{W}}$ are respectively obtained from simulations of WC oscillators on (c) the original weight matrix W, and (d) the rescaled weight matrix $\mathcal{W}$. (e) Empirical functional connectivity obtained from the resting state brain activity of the same individual. (f) Scatter plot displaying the correlation between the empirical FC matrices $C_{(n)}$ and the simulated FC matrices obtained from the original weight matrices $C^W_{(n)}$ for each individual n along the abscissa, and the correlation between $C_{(n)}$ and the simulated FC matrix obtained from rescaled weight matrix $C^{\mathcal{W}}_{(n)}$ along the ordinate. It can be observed that $corr(C_{(n)}, C^{\mathcal{W}}_{(n)})$ is consistently higher than $corr(C_{(n)}, C^W_{(n)})$ for the majority of individuals. This further extends the result shown in Fig. 4.4 (c). (g) Box-plot showing that the correlations of the empirical FC of individual brains $C_{(n)}$ with simulated FC from corresponding rescaled weight matrices $C^{\mathcal{W}}_{(n)}$ are comparatively higher than both $corr(C_{(n)}, C^W_{(n)})$ and $corr(C_{(n)}, C^{\mathcal{A}_{(n)}})$, where $C^W_{(n)}$ represents the simulated FC matrices obtained from corresponding weight matrices, and $C^{\mathcal{A}}_{(n)}$ represents the corresponding adjacency matrices of the rescaled weights. Furthermore, as $corr(C_{(n)}, C^{\mathcal{A}_{(n)}})$ is the weakest, it suggests that the dynamical behavior of the brain is governed more strongly by the weight distribution of the structural connectivity than by the network topology. We also obtain simulated FC generated from two alternative representative structural networks of a human brain, viz. $C^{\langle W \rangle}$ which is generated from the matrix $\langle W \rangle$ that we obtain by averaging each link weight W_{ij} over entire population and C^{λ} which is generated from the matrix λ, which comprises the Poisson parameters λ_{ij} for each link that is Poisson distributed over at least 90% of the population. We observe that the empirical FC matrices correlate better with the simulated FC C^{λ} compared to the simulated FC $C^{\langle W \rangle}$, as indicated by the box-plots.

weight matrices $W_{(n)}$. We also generated simulated FC from adjacency matrices $\mathcal{A}_{(n)}$ obtained from $\mathcal{W}_{(n)}$ and found that FC generated from adjacency matrices had the weakest correlation with the empirical matrices out of all three types of simulated FC, as seen in first three box-plots in Fig. 4.5 (g). This strongly suggests that the dynamics in complex non-linear systems such as the brain is governed much more strongly by the weight distributions of the connectivity than the connection topology itself.

Finally, we compare each of the empirical FC with the corresponding matrix generated

by a generic "representative" network that would describe the structural connectivity of the human brain. There are two alternative ways to obtain a "representative" network from the SC ensemble. The widely used approach is to obtain an average network $\langle W \rangle$, which comprises the average weight of each link calculated over entire population. Our results suggest a second approach, which is to consider the λ matrix, comprising the Poisson parameters λ_{ij} for all the Poisson distributed links, as the representative network. By construction, it comprises only those links that follow a Poisson distribution. As we have already observed that rescaled weights of an individual SC might be more relevant in interpreting structural connections, it is more meaningful to consider the Λ matrix (which effectively is the average of all rescaled weight matrices), as the representative structural network. We generate simulated FCs corresponding to $\langle W \rangle$ and Λ, and observe that FC generated from Λ is more strongly correlated with the bulk of the empirical FCs than the one generated form $\langle W \rangle$, as seen in the last two box-plots of Fig. 4.5 (g). All the correlation distributions being considered here have been shown to be significantly distinct from each other using the *Two-sample Kolmogorov-Smirnov (KS) test* (see Methods).

The representative structural connectome can be resolved into two components: "Basal" network and "Superstructure" network.

We have thus far argued that the representative network described by the Λ matrix is significant because: (i) the underlying generative mechanisms for determining the wiring and weighting of links can be described by a Poisson process (Fig. 4.2 and Fig. 4.3), (ii) the constituent links are significant in terms of topology, as well as, connection weights (Fig. 4.2 (c)), and (iii) The Λ matrix is a better structural correlate to observed function in comparison to the average SC matrix. We now return to one of our original questions regarding the extent of variability of the structural network within the population in terms of topology and weight distribution, focusing only on the constituent links of the representative network. On examining how various link-specific properties, such as the coefficient of variation of link weights (CV_{ij}) and Poisson parameters (λ_{ij}) are distributed among the

constituent links of the representative network, we find that the network can be resolved into two distinct classes of links that we refer to as the "basal" and the "superstructure" network, shown in Fig. 4.6 (a). The former comprises links that are seen in all individuals ($f_{ij} = 1$) while the latter contains all the remaining links of the representative network. They can be identified from the clearly observable bimodality in the distributions of CV_{ij} (Fig. 4.6 (b)) and λ_{ij} (Fig. 4.6 (c)). The bimodality in both distributions can be verified by calculating their bimodality coefficients (see Methods). Basal links are distinguished by very low variability in weight across the population but high values of average weights and λ_{ij}, while the superstructure links show highly variable connection weights across the population, but typically low values of average link weights and λ_{ij}. Notably, the distribution of weight rescaling factors s_{ij} does not show any distinction between the basal and superstructure networks (Fig. 4.6 (d)). Planar projections of the basal and superstructure networks on horizontal, sagittal and coronal planes are provided in Fig. 4.7.

Using the representative network for human brain as the basis, we can explore the mesoscopic organization of human brain in the same way as was done for macaque brain in Chapter 3. We have shown the preliminary results from our modular analysis of human brain network in Appendix D.

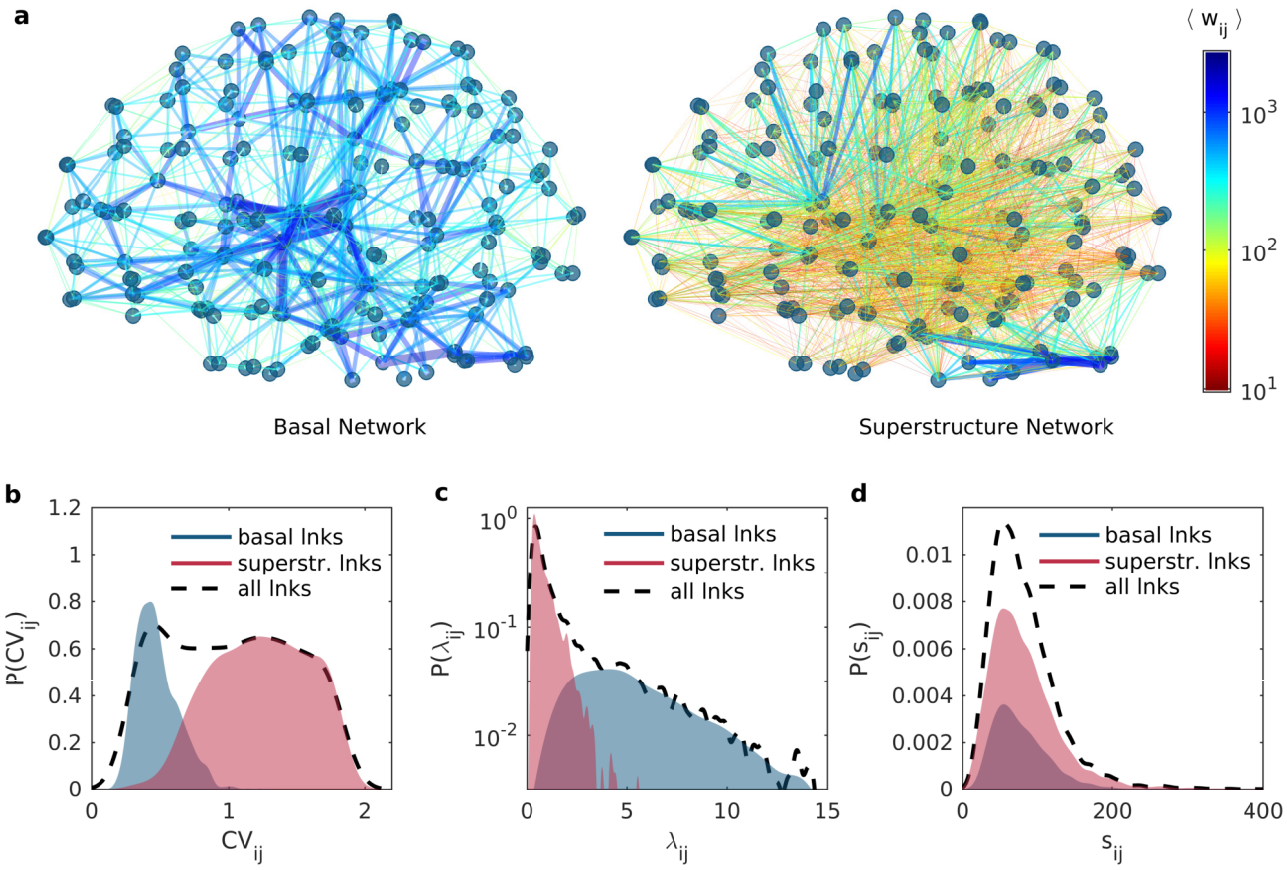


Figure 4.6: The representative structural connectivity of a brain network can be resolved into two components. (a) Sagittal plane projections of the "basal" network (left) and the "superstructure" network (right). The former comprises 1106 ubiquitous links, i.e., those that occur in every individual, and the latter consists of the remaining 2806 links. The thickness and color of each link between a pair of regions (i, j) corresponds to their average weights $\langle w_{ij} \rangle$ across the population (see legend). Note that the average link weights in the basal networks are much higher than those in the superstructure network. (b) The distribution of the coefficients of variation CV_{ij} for the link weights across the population (indicated by broken lines) is observed to be bimodal. The mode corresponding to lower values of CV_{ij} is attributed to links from the basal network (blue shaded region), whereas the links from the superstructure network (red shaded region) primarily contribute to the mode corresponding to higher values of CV_{ij}. This demonstrates that the links of the basal network tend to have higher link weights on an average, and their weights are largely invariant across the population. In contrast, the link weights in the superstructure network vary across individuals.

Figure 4.6 ***(previous page)*:** (c) The distribution of Poisson parameters λ_{ij} (indicated by broken lines) is also bimodal, with each mode corresponding to the basal network links (blue shaded region) and superstructure links (red shaded region) respectively. (d) The distribution of weight rescaling factors s_{ij} (broken lines) is observed to be unimodal, in contrast to λ_{ij} and CV_{ij}, with no distinction between the basal and the superstructure links as indicated by their strongly overlapping distributions (blue and red shaded regions). Note that the separate distributions for basal network links and superstructure network links are normalized according to their relative sizes.

4.4 Discussion

Our results suggest that the expected weight distribution of a link in the structural connectome, as well as its expected probability of occurrence in an individual, can be described by a single parameter. This is indicative of a common generative mechanism that determines both the connection topology of the axonal tract wiring between brain regions, as well as, the anatomical thickness of the tracts, which we refer to as the connection weight. At the neuronal scale connection weights refer to the number of synapses between two neurons, or the synaptic conductivity. These quantities have a direct measurable effect on the complex electrophysiological interactions between the neurons. The plasticity and learning mechanisms that determine and alter the connection weights between neurons are well understood, e.g., spike-time dependent plasticity. However, in the case of the macro-connectome, the role of connection weights, viz., the density of axonal bundles, in the functional interactions of the brain areas is not well understood. By determining a latent Poisson distributed quantity from the observed weight of a connection, which we refer to as the rescaled weight, our results point towards a potential framework for a better functional interpretation of structural connectivity. This might be extremely useful in the development of dynamical models of various cognitive phenomena in the brain. To probe this, we have used a relatively simple neural mass model to show that the rescaled weights consistently give rise to dynamical behavior that have better correspondence with the empirical data. This mode of functional interpretation of structural connectivity can be further enhanced when we also consider information related to whether the synaptic

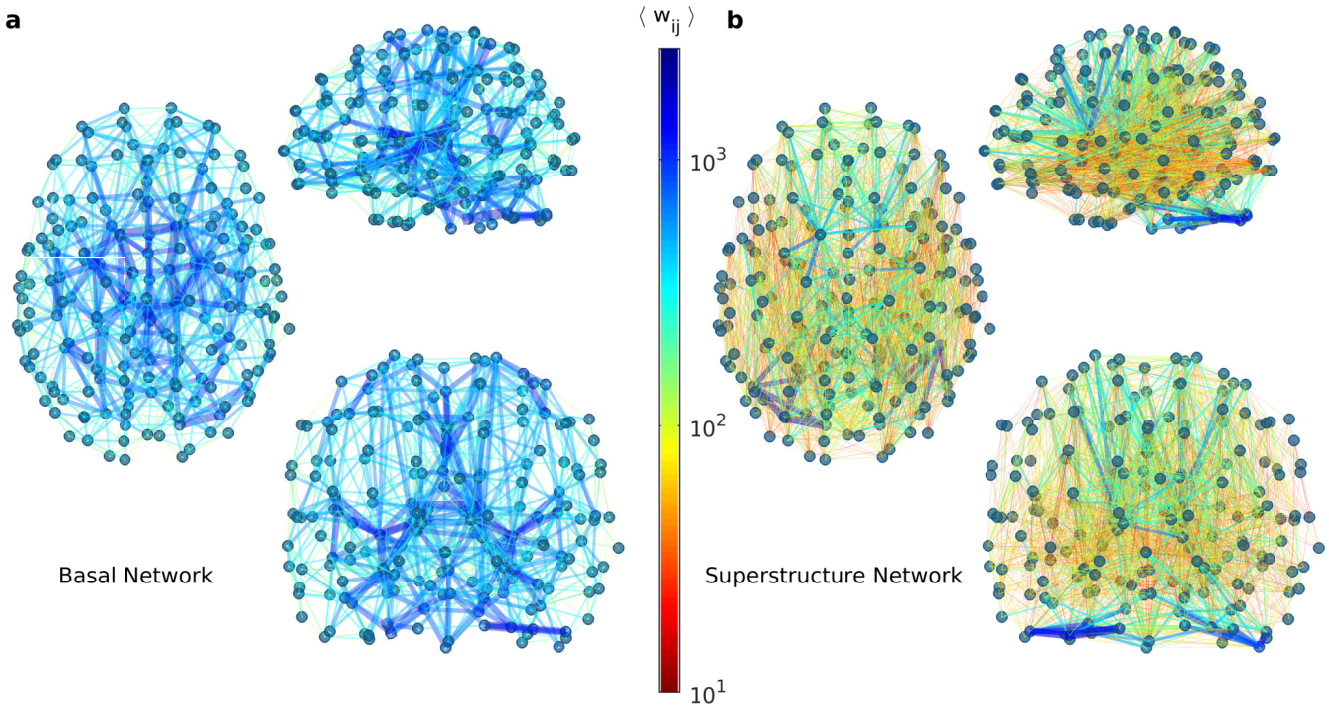


Figure 4.7: "Basal" network and "Superstructure" Network. (a) Horizontal, sagittal and coronal projections of the spatial representations for the "basal" network (left, top right and bottom right respectively). (b) Horizontal, sagittal and coronal projections of the spatial representations for the "superstructure" network (left, top right and bottom right respectively). Note that the average weights $\langle w_{ij} \rangle$ (represented by the thickness and color of the links (i, j), see legend) display a smooth spatial gradient. In addition, the long range connections between spatially distant regions are more frequently observed in the superstructure network, as compared to the basal network whose structure is closer to a lattice in that most of the connections are between spatially adjacent regions.

connections underlying given axonal pathways are excitatory, inhibitory or both. Another component that is missing from the analysis of structural connectome is the directional information about the connections. Although we represent the structural connectome as an undirected weighted network, in reality synaptic connections are always directed. The observation that the rescaled weight distribution showed a widespread enhancement in correspondence between structure and function, even without the associated information about the directionality and the type of connections, underpins the significance of this framework. In that case, the latent Poisson parameter associated with the axonal pathway, i.e., the rescaled weight, might represent functionally relevant factors such as the actual number of axons in the tract or the number of synaptic connections, rather than simply representing the anatomical thickness. Similarly, the corresponding rescaling factors might be representative of peripheral features that do not directly affect the neuronal interactions, e.g., the thickness of myelin sheaths covering the axons.

We further observe that a significant fraction of links showed deviations from a Poisson generative process, and these deviations are far greater than expected by the finite size of the data. This further illuminates the generative mechanism: even though random independent discrete processes might be involved in the wiring of major portion of the brain, as indicated by the occurrence of Poisson distributions, there are other significant effects at play. These might arise from genetic or developmental factors, or may be governed by the specific functioning of an individual brain. For instance, pathways between certain motor regions in the brain of a professional athlete might be exceptionally stronger than that of other individuals due to prolonged specialized usage of certain circuits. Thus the deviations might be indicative of plasticity in the macro-connectome - a hypothesis that requires further exploration.

In this study we have argued that the inclusion of only the Poisson-distributed links in a generic representative network for a human connectome, with the corresponding weights being the link specific Poisson parameters, is a more meaningful approach than simply obtaining an average of all structural connectivity matrices. The representative network based on Poisson parameters at once informs us about the topological significance, the

extent of variability, the generative mechanism and the functional importance of the underlying links, thus making it far more useful for further network-theoretic or dynamical analysis. The two distinct components of the representative network, which we refer as the basal and the superstructure networks respectively, reveal an altogether new organizational aspect of the brain. While the source of this dichotomy within the structural connectome is not clear, one needs to do a more detailed exploration into the developmental and functional implications of the two clearly distinguished components that comprise a generic representative structural connectome of a human brain.

Chapter 5

The Hierarchy Index: A tool for revealing the sequential arrangement of processing levels in brain networks

5.1 Introduction

In order to obtain a systems-level understanding of the functioning of the brain (or a nervous system), several studies have applied a wide range of tools for the analysis of complex networks on numerous structural and functional brain connectivity datasets [3, 254]. However, most studies of such networks at the mesoscopic level have focused on a specific type of mesoscopic organization, viz., modularity [150, 19]. Modularity, or community structure, is a widely-observed property in brain networks across species and length scales where, as described in detail in Chapter 3 of this thesis, the network is partitioned into communities that are distinguished by relatively dense inter-connections between nodes of the same modules and sparse connections across different modules. However, there is another commonly observed feature of brain networks that may not have received as much attention, viz., hierarchical organization. In the literature, the concept of 'hierarchy' in network analysis has been used in diverse contexts to describe different types of network

arrangements. For instance, hierarchy has been used to describe a macroscopic property of complex networks that makes them scale-free and highly clustered [255, 76]. In another context, it has been used to describe the core-periphery structure in the *C. elegans* neuronal network [21]. Another widely-observed network organization that has been described as hierarchical, is the hierarchically modular organization [256, 117, 257, 10], which describes a nested arrangement of modules at several levels, in which smaller modules are nested inside larger modules.

The hierarchical organization we consider here, which is of particular relevance in the context of brain networks, is a specific mode of mesoscopic organization that is distinct from those discussed above. It stems from a commonly observed pattern of connectivity identified by a large number of feed-back and feed-forward pathways between distinct sets of brain regions or neurons. Such an arrangement of reciprocal connectivity is associated with networks involved in highly complex information processing. Here, the network comprises of several hierarchically arranged levels (or layers) where each level processes the information obtained from the preceding level and then passes it onto the successive one. For instance, the macaque visual cortex has a well-defined hierarchical structure [258, 259, 260] in which visual information enters the brain through the Lateral Geniculate Nucleus (LGN), from where the information is relayed to the primary visual cortex V1 and subsequently fed forward to higher areas such as V2, V3, MT and so forth. Here, these areas can be considered as hierarchical levels within the neuronal network, where each level integrates information from its lower levels and successively performs more complex information processing. Similarly, there are top-down feedback connections in the opposite direction, i.e., from the higher to the lower levels. Such hierarchical organization is not confined to the macaque brain, as it is also observed in rat brains [261] and mouse brains [262]. In addition to visual processing, hierarchical organization is also implemented in auditory processing [263]. Furthermore, the hierarchical arrangement of neurons is also important in those neural circuits where a specific temporal sequence of neuronal firing is necessary. The synfire chain is an idealized model of such a neural organization used to explain the synchronous sequential firing of neurons [264]. One concrete example of hierarchically arranged neu-

ronal layers (synaptic chains) giving rise to collective neuronal firing in a precise temporal sequence, is seen in the HVC neurons in zebra finch, one of the most widely studied species of "songbirds" [265].

In all of the above examples of hierarchically organized networks, the distinct levels are identified by their observed functions, such as the temporal sequence of firing in neurons or the complexity of visual processing in brain regions. However, while there are many network analysis tools available for the detection of modular organization in networks, no comparable tools exist for the detection of hierarchical levels within a network of a given connection topology. Recently there has been some work towards this direction [266], where the inherent hierarchy has been identified in mouse cortical connectivity based on the direction of information flow. With the rapid development of techniques for determining the underlying structural connectivities in the brain network, and the resulting abundance of connectivity data, more such approaches are required to identify the inherent structural aspects of brain networks such as hierarchy - a point that been compellingly argued by Sebastian Seung [30].

In this work, we propose a novel approach for identifying the underlying hierarchical levels in a network, which is sufficiently general to be applicable outside the context of neuroscience. We first define the hierarchy index, a new measure for quantifying the extent of hierarchy in a network, and then describe a heuristic simulated annealing routine that maximizes the hierarchy index to yield an optimal partitioning of the network into levels, along with their sequential arrangement. We validate our method using a set of benchmark networks and apply it to the three brain networks studied earlier in this thesis, viz., the network of the *C. elegans* nervous system, the macaque brain connectome and the representative human brain network. The preliminary findings of the hierarchical analysis of these networks that we report here reveal a strongly hierarchical organization with robust levels.

5.2 Methods

Hierarchy index

We consider a network to have a hierarchical organization if it can be partitioned into levels (interchangeably also called layers in this chapter), such that there is high degree of connectivity between nodes of levels that are adjacent to each other in their sequential arrangement and relatively sparse connections between non-adjacent levels. Here, the sequential arrangement of connectivity, arising from the arrangement of the underlying levels, attributes a sense of hierarchy in the chain of interactions across the network from a mesoscopic perspective. Hence, the hierarchical organization of a network is specified by the node membership of each of the partitioned hierarchical levels, along with the particular sequential arrangement of these levels. Note that this is distinct from the concept of hierarchical modular organization, which has been studied extensively [256, 117, 257, 10].

In order to characterize the extent to which a given network exhibits a hierarchical organization, we define the hierarchy index H, which is analogous to the Newman-Girvan measure Q for the degree of modularity in a modular network [115, 165]. For a directed-unweighted network that can be arranged into a sequence of levels, the hierarchy index H is given by:

$$H = \frac{1}{L} \sum_{i,j} \left[A_{ij} - \frac{k_i^{in} \cdot k_j^{out}}{L} \right] \cdot (\delta_{l_i, l_j+1} + \delta_{l_i+1, l_j}) \tag{5.1}$$

where, A_{ij} refers to the element of adjacency matrix of the network which is 1 if there is a directed link from node j to node i and 0 otherwise, k_i^{in} refers to in-degree i.e., the total number of connections received by node i ($k_i^{in} = \sum_j A_{ij}$) and k_j^{out} represents the out-degree i.e., the total number of connections sent out by node j ($k_j^{out} = \sum_i A_{ij}$). Here the total number of connections is represented as L ($= \sum_{ij} A_{ij}$). The Kronecker delta function term $\delta_{l_i, l_j+1} + \delta_{l_i+1, l_j}$ yields 1 if the levels l_i and l_j to which i and j belong are adjacent to each other in the sequential arrangement, and is 0 otherwise.

Maximization of the hierarchy index

In order to find the latent hierarchical structure in a given network, the task is to determine the optimal partitioning of the network into levels, along with their optimal sequence of arrangement that would yield the maximum value for H. A similar problem of finding the optimal modular partitioning for maximizing Q in modular networks is known to be NP-hard [267, 166]. Therefore it is reasonable to assume that finding the most optimal solution for level partitioning and sequencing for a given network is also an NP-hard problem. We address this optimization problem by formulating a heuristic routine of simulated annealing, which attempts to find the global maxima for the hierarchy index H by extensively exploring the configuration space of levels. Here the configuration refers to a specific partitioning of nodes into levels and the associated sequential arrangement of those levels. The routine is described as follows.

We begin with an arbitrary configuration of sequentially arranged levels that is characterized by a hierarchy index H. We then begin an iterative process whereby the level configuration is modified through a single manipulation to the level of a node or level at each step (see below). The hierarchy index of the network upon a single modification is defined as H', and the corresponding change in the hierarchy index is $\Delta H = H' - H$. If $\Delta H > 0$, we accept the modification and consequently assign it to be the "current" configuration, while if $\Delta H < 0$ we accept the modification with a probability P, which is given by $P = e^{-|\Delta H|/T}$. Here, T is a "temperature", which is decreased slowly with each iteration as the routine progresses. If we reject the modification, the current configuration is retained for the next iteration. When continued for a sufficient number of iterations until $T \to 0$, the algorithm is guaranteed to converge to a local maximum in the hierarchy index. For each network, we run many realizations of the simulated annealing routine to obtain an ensemble of optimal level partitionings of the network.

The modifications that are performed at each iteration could be one of two types: (i) single node modification and (ii) level modification. For a single node modification, we randomly select a node from the network and either shift it from its current level to

another randomly chosen level, or place it in a new level by itself at the end of the existing level sequence. This modification set consists of a total of $(N(k-1)+N =)Nk$ moves at each step, where N is number of nodes and k is the number of levels. For the level modifications, we perform one of three actions: (a) merge a pair of randomly chosen levels into a single level, (b) swap the sequential positions of a pair of randomly chosen levels, or (c) split a randomly chosen level into two adjacent levels. This modification set comprises $({}^{k}C_2 \times 2 + k =)\ k^2$ moves at each step. The splitting of a randomly chosen level can in principle be done using a random cut, but for faster convergence we cut the level into two levels in such a way the number of links between the two is maximized. For this we adopted the MINCUT algorithm [268] which is conventionally used to find the minimum cut weight in a graph, i.e., partition a network into two subnetworks with a minimum number of links between them. With a simple modification to the algorithm, we performed the opposite, viz., partition a network into two subnetworks with maximum number of links between them.

Thus, starting from an arbitrary level configuration, at each iterative step we chose one out of all possible $Nk+k^2$ modification moves described above with a uniform probability, and each modification is accepted or rejected according to the above described criterion. The initial temperature T_0 was set as 10 and we decreased this as function of time (measured in iteration steps) t, as $T = T_0 \cdot e^{-\lambda t}$, where $\lambda = 2 \times 10^{-6}$. The routine is run for a sufficiently long time ($t = 2{\times}10^7$) to allow the system to converge to an optimal hierarchical configuration. If the system is stuck in a local maximum for too long, i.e., if there are no accepted modifications for a large number of time steps t_{thresh}, the routine is terminated. As the value of t_{thresh} must be sufficiently larger than the total number of possible moves, we set $t_{thresh} = 5 \times (Nk + k^2)$. All these parameters are tuned for the sizes of networks that are of interest to us, viz., $N \approx 300$. The schematic representation summarizing the above described procedure is shown in Fig. 5.1.

Given the inherent degeneracy of the method, namely that the solutions obtained for optimal hierarchical level partitionings via the simulated annealing method are not unique, we require a sufficiently large ensemble of solutions for any given network. For a network

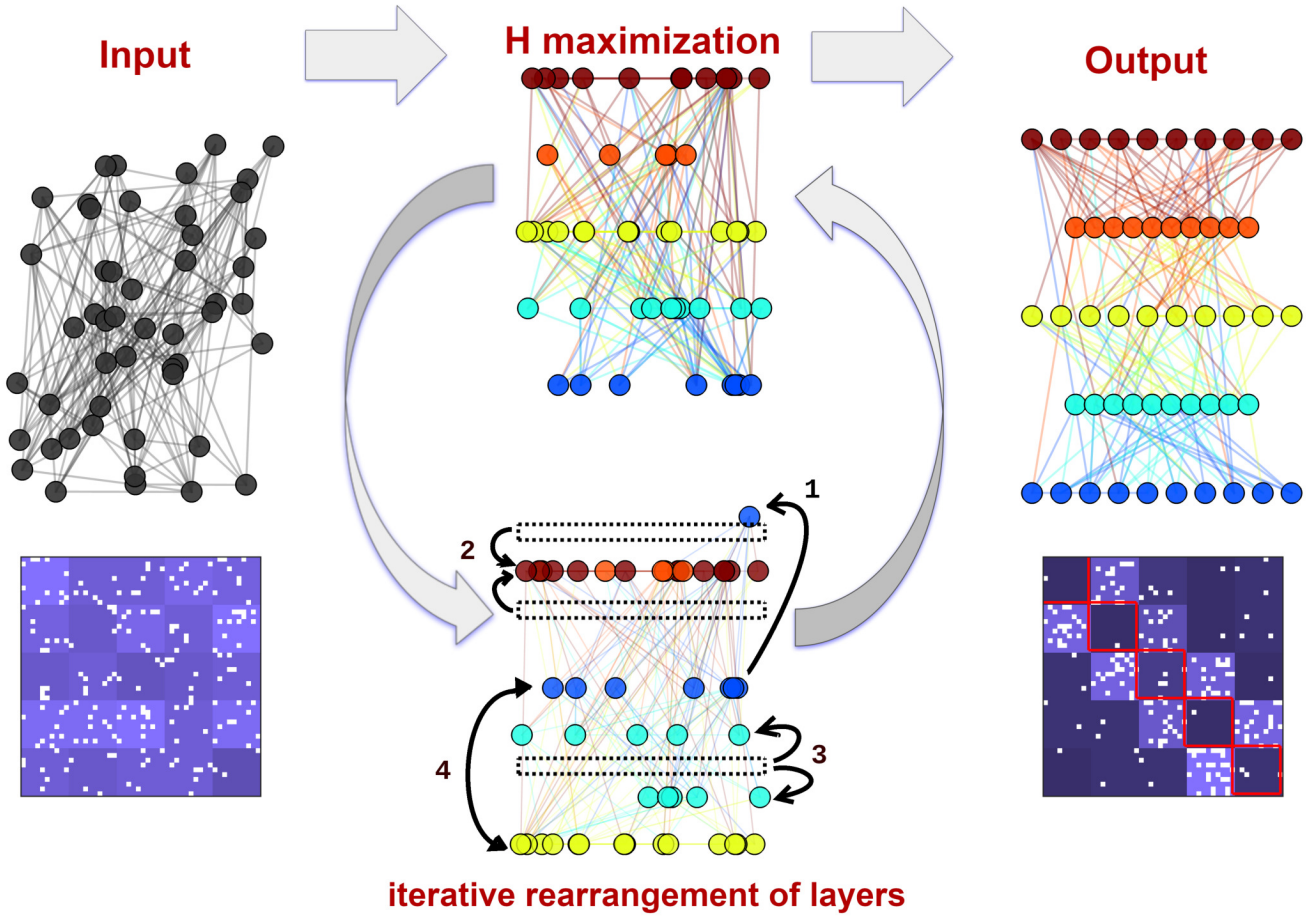


Figure 5.1: Schematic representation of hierarchy detection method. The input (shown on the left panels, top and bottom) consists of a network with latent hierarchical arrangement of levels (or layers). The adjacency matrix shown in the bottom left panel, characterized by a random ordering of nodes, does not display any inherent structure. The iterative maximization routine for the hierarchy index H is represented in the middle panels, where at each step we perform one of the four types of level rearrangements that are illustrated in the bottom middle panel, viz: (1) Moving a single randomly chosen node from its current level to either another randomly chosen existing level or a new level, (2) merging two randomly chosen levels into a single level, (3) splitting a randomly chosen level into two adjacent levels using a modified MINCUT algorithm, and (4) swapping the sequence of two randomly chosen levels. The output (shown in the right panels, top and bottom) contains the optimal ordering of nodes into sequentially arranged levels with relatively denser connections between consecutive levels being clearly apparent. The adjacency matrix shown in bottom right has the nodes ordered according to the levels obtained from the H maximization routine, and shows a clear hierarchical structure (note the dense connections within the off-diagonal blocks, corresponding to relatively high connectivity between adjacent levels).

with a robust hierarchical structure, all the solutions obtained must be consistent with each other to a large degree.

Random benchmark networks to test the performance of the method

In order to test the performance of the method, we generated random directed, unweighted networks with embedded hierarchical levels. The extent of hierarchy in the networks were parameterized by the ratio ρ/ρ_{con}, where ρ represents density of connections between nodes of non-consecutive levels (or the same level) whereas ρ_{con} is the density of connection between any two consecutive levels. For completely hierarchical levels we have $\rho/\rho_{con} = 0$, while for a completely homogeneous network lacking any mesoscopic structure we have $\rho/\rho_{con} = 1$. We generated 20 benchmark networks of size $N = 250$ with varying degrees of hierarchy from values of $\rho/\rho_{con} = 0.05$ up to $\rho/\rho_{con} = 1$. Each network had 5 levels with 50 nodes in every level. We randomly rearranged the nodes of the adjacency matrix for each network and ran our simulated annealing routine on those networks to identify the inherent levels and their sequential arrangement. For each network, we carried 50 realizations to obtain an ensemble of optimal solutions. The accuracy of the method was measured by the similarity between the level partitioning obtained from the method and the original embedded level partitioning used for generating the network.

We quantify the similarity between two level partitionings $\{l_i^A\}_{i=1}^{L_A}$ and $\{l_j^B\}_{j=1}^{L_B}$ resulting from different partitionings A and B of a network (that comprise levels L_A and L_B, respectively) in terms of the normalized mutual information [171]:

$$I_{\text{norm}}\ (A,B) = \frac{2\sum_i \sum_j P(l_i^A, l_j^B) \ln[P(l_i^A, l_j^B)/P(l_i^A)P(l_j^B)]}{-\sum_i P(l_i^A)\ln P(l_i^A) - \sum_j P(l_j^B)\ln P(l_j^B)}, \tag{5.2}$$

where $P(l_i^A)$ is the probability that a randomly chosen node lies in level l_i^A in partition A, $P(l_j^B)$ is the probability that a randomly chosen node lies in level l_j^B in partition B, and $P(l_i^A, l_j^B)$ is the joint probability that a randomly chosen node belongs to level l_i^A in partition A, as well as to level l_j^B in partition B ($i = 1, \ldots, L_A$, and $j = 1, \ldots, L_B$). Each

of the probabilities can be estimated from the ratio of the level sizes to the size of the entire network.

Data for neuronal/brain networks

Caenorhabditis elegans

Connectivity. We have used the information about the connectivity between 279 connected neurons of the *C. elegans* somatic nervous system from the latest dataset, which is published in Ref. [97].

Functional type. The information about the functional type of the neuron (viz. sensory, inter- and motor neurons) has been obtained from the database provided in Ref. [102].

Spatial information. Information about the locations of the neurons was obtained from the database reported in Ref. [100], accessible online from `https://www.dynamic-connectome.org/`. The location information provides coordinates of each neuronal cell body projected on a two-dimensional plane defined by the anterior-posterior axis and the dorsal-ventral axis.

Macaque

Connectivity. We have used a reconstructed macaque structural connectome comprising 266 cortical and subcortical brain regions as described in Chapter 3, which was based upon the compilation in Ref. [159]. The compilation had used several hundred tract tracing studies obtained from CoCoMac - a comprehensive neuroinformatics database [160, 161, 162].

Spatial Information. We have obtained the stereotaxic coordinates of each brain region in our connectome from several sources. Information about 134 of the 266 regions included in the connectome has been obtained from the website [163] associated with the Paxinos Rhesus Monkey Atlas [164]. For the remaining regions, we manually curated the requisite data from the relevant research literature. The position of a region is identified with the approximate location of its center, obtained from the online three-dimensional visualization platform in the website mentioned above. The volume spanned by a particular region was

estimated by approximating the cross-sectional area occupied by the region in each of the coronal sections of the brain in which it appears and obtaining the sum of these areas weighted by the thickness of the sections measured along the rostral-caudal axis.

Human

The Human brain structural connectivity analyzed in this study constitutes the "basal network" component of the representative human brain network (see Chapter 4 for details). It comprises 188 brain regions spanning the gray matter region (cerebral cortex, sub-cortical areas, cerebellum, brain stem, etc) connected to each other through axonal tracts. The "basal" network was determined through the cohort analysis of structural brain networks of human subjects, as obtained from the *Nathan Kline Institute (NKI) / Rockland Sample* [247] - a publicly available repository of diffusion tensor imaging (DTI) data - which was further processed into connectivity matrices and made publicly available by the *UCLA multimodal connectivity database* at `http://umcd.humanconnectomeproject.org/` [248]. As described in Chapter 4, the "basal" network comprises those connections which are seen in each of the 196 individual brain networks, and are distinguished by very low variability in connection weights across the population. The database also contains the 3-dimensional coordinates for each of the brain regions in a standardized space.

In this study, all the networks analyzed were unweighted. We carried 200 realizations of simulated annealing routines for each of the empirical networks.

5.3 Results

Prior to applying the detection method to uncover the inherent hierarchical structures on real-world networks, where the existence of underlying sequentially interconnected levels is not known *a priori*, we test the reliability and the efficacy of the results obtained using this method. To this end, we generate a set of "synthetic" random networks with an embedded hierarchical structure embedded in them, that we refer to as benchmark networks (see Methods). These networks are obtained by specifying the level partitioning of nodes and their sequential arrangement, and randomly connecting the nodes while maintaining a

relatively high density of connections between consecutive level ρ_{con}, in comparison to the connection density over remaining network ρ. As the hierarchy maximizing algorithm does not utilize any prior information about the levels, it is expected to identify the embedded levels accurately even if the nodes are rearranged randomly.

As can be seen from Fig. 5.2 (a-b), not only is the method able to recover the embedded levels in the benchmark networks for strongly hierarchical networks with extremely high accuracy (Fig. 5.2 (a)), but it is remarkably accurate in identifying the embedded levels in networks with much weaker hierarchical structure as well (Fig. 5.2 (b)). In Fig. 5.2 (c) we show the variation in the average performance of the method over many realizations of H maximization, with the degree of hierarchy that is embedded in the network. We observe that the method performs with remarkably high accuracy ($I_{norm} \geqslant 0.8$) over a broad range of values for the extent of hierarchy (up to $\rho/\rho_{con} \approx 0.4$). This performance level is consistent across the distinct realizations of hierarchy detection for a single network, as indicated by the narrow shaded region representing the variation of I_{norm} across many realizations for each network. The optimality of the identified levels from the method becomes further evident by comparing the values of hierarchy index H for each network, as obtained with respect to the originally embedded level arrangement, with those that were obtained with respect to the level arrangement determined from the method. Fig. 5.2 (d) shows that the H values obtained with respect to the determined levels almost coincide with those corresponding to the originally embedded levels up to $\rho/\rho_{con} \approx 0.4$.

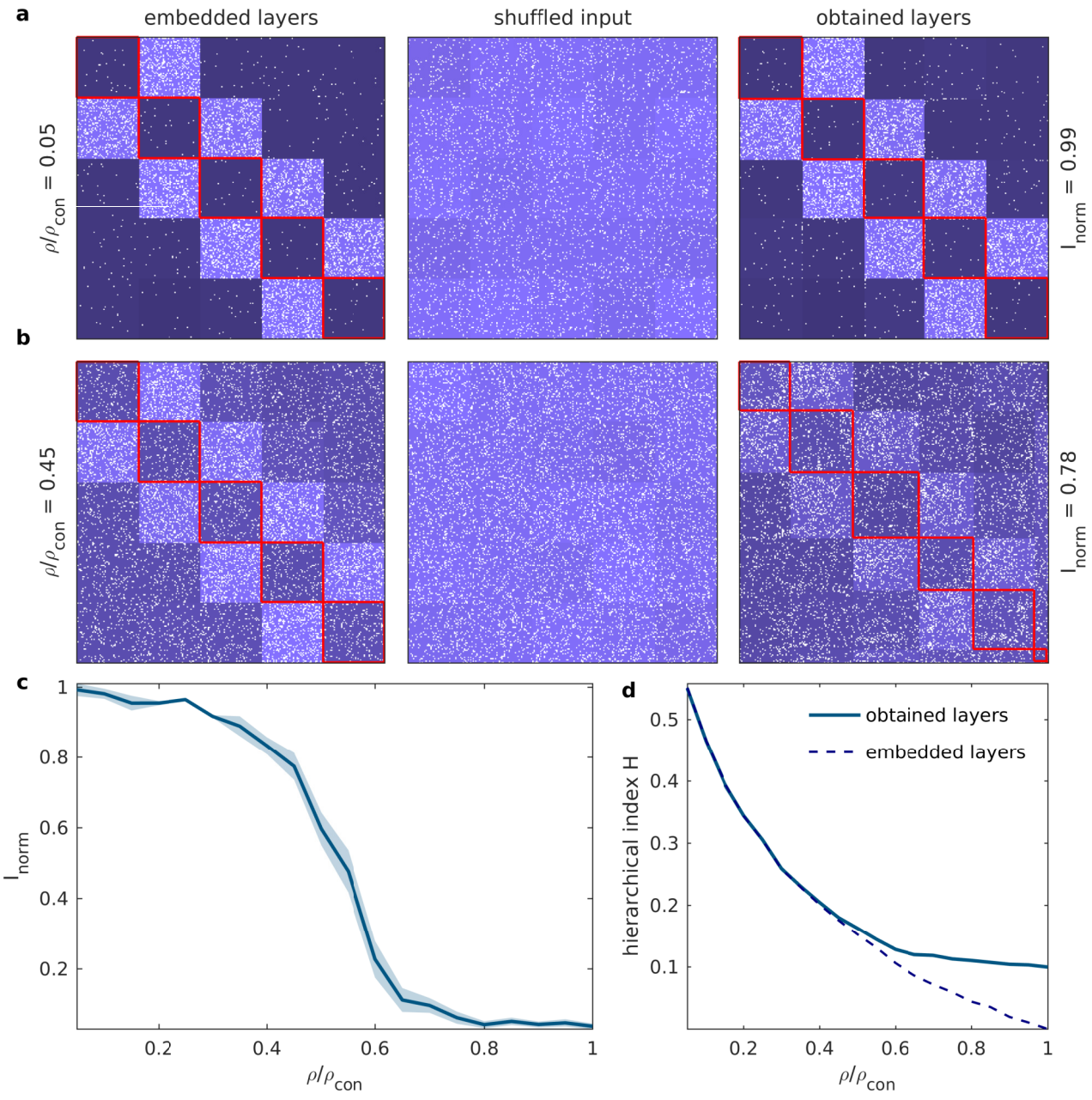


Figure 5.2: Investigating the accuracy of the method in detecting hierarchical levels in random benchmark networks with embedded hierarchy. (a-b) Application of the hierarchy detection method to random directed-unweighted benchmark networks with embedded hierarchical levels.

Figure 5.2 ***(previous page)*****:** The extent of hierarchy in each benchmark network is indicated by the ratio of the density of connections between nodes of non-consecutive levels ρ with that of the density of connections between nodes of consecutive levels ρ_{con}, viz. (a) $\rho/\rho_{con} = 0.05$ (strongly hierarchical) and (b) $\rho/\rho_{con} = 0.45$ (weakly hierarchical). The three adjacency matrices in each row represent the same benchmark network, but differ in terms of the ordering of the nodes. The white entries in the matrices correspond to links, while the connection density between nodes belonging to any two distinct levels, as well as within-level connectivity for each level is indicated by the brightness of the corresponding block in the matrix. The matrices at the left represent the original arrangement of nodes, in which the embedded level structure (indicated by red lines) is apparent, as only the off diagonal blocks adjacent to the diagonal blocks show a high density of connections. The matrices in the middle are obtained by randomly reordering the nodes, while preserving the connection topology. This matrix serves as the input for the hierarchy detection method. The matrices on the right are obtained by rearranging the nodes according to the levels that were obtained from the method (indicated by red lines). It can be observed that the obtained levels closely match the original embedded levels in each case. The accuracy of the method is quantified by the degree of similarity between the initial network partitioning according to the embedded levels and the final network partitioning according to the obtained levels, by calculating the normalized mutual information I_{norm} (see Methods) between the two partitionings. For two completely identical partitionings we have $I_{norm} = 1$ and for maximally distinct partitionings we have $I_{norm} = 0$. It can be observed that even for a weakly hierarchical network (panel (b)), the method detects the underlying levels with remarkably high accuracy ($I_{norm} = 0.78$). (c) Performance of the hierarchy detection method, as indicated by I_{norm}, tested on benchmark networks with values of ρ/ρ_{con} varying between 0.05 to 1, over a sufficiently large number of realizations (50 for each benchmark network). Note the remarkably high accuracy of the method up to $\rho/\rho_{con} = 0.4$, which corresponds to networks that are weakly hierarchical. The solid curve corresponds to the mean of I_{norm} calculated over the ensemble, while the shaded region represents the standard deviation. (d) The average value of hierarchy index H over all realizations corresponding to each value of ρ/ρ_{con} (solid line), compared with the H values corresponding to the original embedded level partitionings (dashed line). For higher values of ρ/ρ_{con} (> 0.5), the H values for the obtained levels tend to saturate at 0.1, whereas those corresponding to the original levels tend to 0.

The benchmark analysis provides a strong validation for our hierarchy detection method. In order to investigate its applicability in uncovering the latent hierarchical structure of real-world networks, particularly in brain networks, we consider three neuronal systems that have been studied in the previous chapters of this thesis. In this chapter we report findings from the preliminary analysis of: (i) *Caenorhabditis elegans* somatic nervous system, (ii) structural connectome of the macaque brain, and (iii) structural connectome of the human brain (see Methods for the details of the network data).

The hierarchical arrangement of levels obtained from one of the 200 realizations of

H-maximization for the case of the *C. elegans* nervous system is shown in Fig. 5.3 (a-b). The spatial representation of the neurons (colored according to their levels) and their connections across the anterior-posterior axis (head to tail) shows the spatial configuration of the hierarchical levels. The network is arranged into 14 levels, with the initial levels (1 to 4) mostly confined around the tail region, and which appear to predominantly consist of interneurons. Intriguingly, the following sequence of levels (5 to 10) are located around the very opposite end of the nervous system, viz., the head region. The magnified view of the head region in Fig. 5.3 (b) shows a large diversity in the type of cells in these levels but sensory neurons are relatively predominant. The subsequent terminal levels (11 to 14) are spread across the ventral nerve chord, which is known to mostly consist of motor neurons that help the organism in locomotion. The hierarchical levels vary considerably in size (within an order of magnitude) as shown in Fig. 5.3 (c), upper panel, which displays the average number of neurons in each level, as well as its variation, over all 200 realizations. These quantities are also shown separately for each neuron type (viz. sensory, inter- and motor neurons). We observe that the predominance of motor neurons in the terminal levels is consistent across all realizations. A more illuminating scenario is revealed from the composition of each level in terms of the types of constituent neurons, as shown in Fig. 5.3 (c), lower panel. The average share of the neurons of a particular type within each level (expressed in fractions) serves as a functional characterization for the levels. The composition of the levels appear to be largely consistent over all the realizations, as indicated by the variation of the fractions across realizations (shaded region). Surprisingly, in all of the realizations of optimal level partitionings obtained from the method we find that the initial levels are predominantly composed of interneurons, while the following intermediate levels are dominated by sensory neurons, which in turn are followed by levels that are mostly motor neurons. This is consistent with observations from Fig. 5.3 (a-b), which corresponds to just one out of 200 realizations. The adjacency matrix for *C. elegans* neuronal connectivity shown in Fig. 5.3 (d) illustrates the dense inter-level connectivity between consecutive levels, which is characteristic of a strongly hierarchical organization. Furthermore, the robustness of hierarchical description of the network with respect to

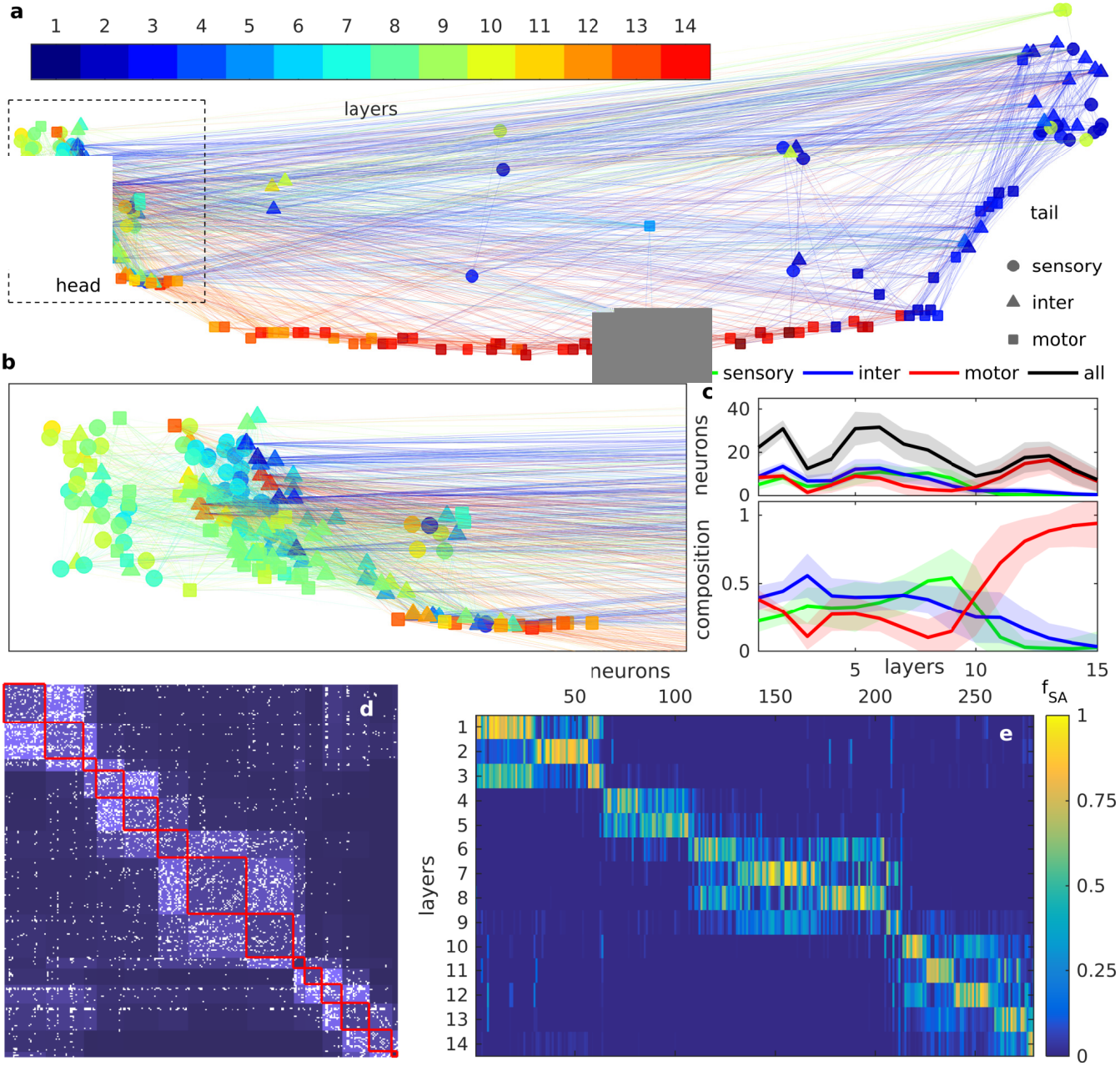


Figure 5.3: Hierarchical structure of the *C. elegans* somatic nervous system. (a) Planar projection of the spatial representation of the *Caenorhabditis elegans* neuronal network underlying the somatic nervous system along the anterior-posterior axis (head and tail are indicated), where each node represents a neuron and each link represents a synapse or a gap-junction. The shape of the nodes indicates the neuron type, viz. sensory neurons (circles), interneurons (triangles) or motor neurons (squares).

Figure 5.3 ***(previous page)*:** The node color represents the hierarchical level of the corresponding neuron, as obtained from one of the realizations of hierarchy detection method (see the legend above). The links are colored in accordance with the corresponding source node. Note that the initial levels are found near the tail region (nodes colored in blue shades), the mid-levels are located in the head region (nodes colored in green shades), and the terminal levels are spread across the ventral nerve cord, comprising mostly motor neurons (nodes colored in red shades). (b) A magnified view of the head region that is enclosed by the broken lines in panel (a). Note the high density of connections between neurons of consecutive levels (mostly levels 7 and 8). (c) Variation in the sizes of levels, and their composition, in terms of neuronal type, determined using the hierarchy detection method over an ensemble of 200 realizations. The upper panel shows the absolute number of neurons in each level, where the solid curve represents the mean and the translucent band represents the standard deviation across all realizations. The number of sensory, inter and mirror neurons are represented by green, blue and red curves respectively, while the black curve represents the total number of neurons in each level. The composition of each level in terms of the constituent neurons is represented by fraction of each level occupied by each type of neuron, as shown in the lower panel. Note that the initial levels are dominated by interneurons, followed by levels with high dominance of sensory neurons, and the terminal levels are almost entirely occupied by motor neurons. (d) The adjacency matrix representing the *C. elegans* somatic nervous system, where nodes are rearranged according to the levels obtained from one of the realizations using the hierarchy detection method. The white entries represent unweighted directed links and the connection density between neurons belonging to any two distinct levels, as well as within-level connectivity for each level, is indicated by the brightness of the corresponding block in the matrix. Here, the high density of connections between consecutive levels are indicative of the strongly hierarchical organization of the underlying network. (e) Variability in the hierarchical organization obtained across all the realizations of hierarchy detection. The fraction of simulated annealing realizations f_{SA} in which a particular neuron (abscissa) is placed in a given level (ordinate) is indicated by the color coding provided in the legend. The neurons are arranged according to the levels obtained from one of the realizations using the hierarchy detection method, and is used as a reference to compare all other realizations. We observe that there is a significant number of neurons whose membership in their respective levels are highly consistent across different realizations, as indicated by the high values of f_{SA}. This indicates a robust hierarchical organization with an underlying sequence of neuronal connectivity at the mesoscopic level (see the diagonal arrangement of yellow-shaded blocks across the levels).

all realizations is shown in Fig. 5.3 (e). The frequency with which a particular neuron appears in a given level over different realizations (expressed in terms of the fraction of realizations f_{SA}), shows that a significant number of neurons consistently appear in the same level. This implies that even though the distinct partitionings obtained from all realizations might differ marginally at the microscopic level of few nodes, those partitionings are largely in agreement at the mesoscopic level. This strongly suggests that there is an

inherent hierarchy in the neuronal connectivity of the *C. elegans* nervous system at the mesoscopic level, collectively represented by all the degenerate solutions obtained from the many realizations of our method.

The hierarchical analysis of the macaque brain network is shown in Fig. 5.4. In comparison to the *C. elegans* nervous system, the spatial configuration of the 17 hierarchical levels in the macaque connectome exhibits a much more spatially contiguous arrangement in that the sequentially adjacent levels also appear to be spatially adjacent (as seen in Fig. 5.4 (a)). We observe a contiguous progression of levels from the anterior to posterior region (see the sagittal section in top right panel), starting from the prefrontal cortex (nodes shown in blue shades) across the parietal lobe, and up to the occipital lobe. From the occipital lobe the subsequent levels progress in the reverse direction across the temporal lobe to finally terminate in the prefrontal cortex (nodes shown in red). Intriguingly, both the initial and the terminal levels are located in the frontal lobe. The adjacency matrix showing the dense inter-level connections illustrate the strongly hierarchical organization in the macaque connectome, which is visually even more apparent than that in the *C. elegans* neuronal network (Fig. 5.3 (d)). Finally in Fig. 5.4 (c), the frequency of occurrence for each of the brain regions in different levels that are obtained across all the realizations shows an extremely robust partitioning that is largely invariant across realizations. Almost all brain regions show a high frequency of occurrence in their respective levels. Comparing this with the scenario in Fig. 5.3 (e) suggests that the hierarchy in the macaque connectome is far more robust than that of the *C. elegans* nervous system.

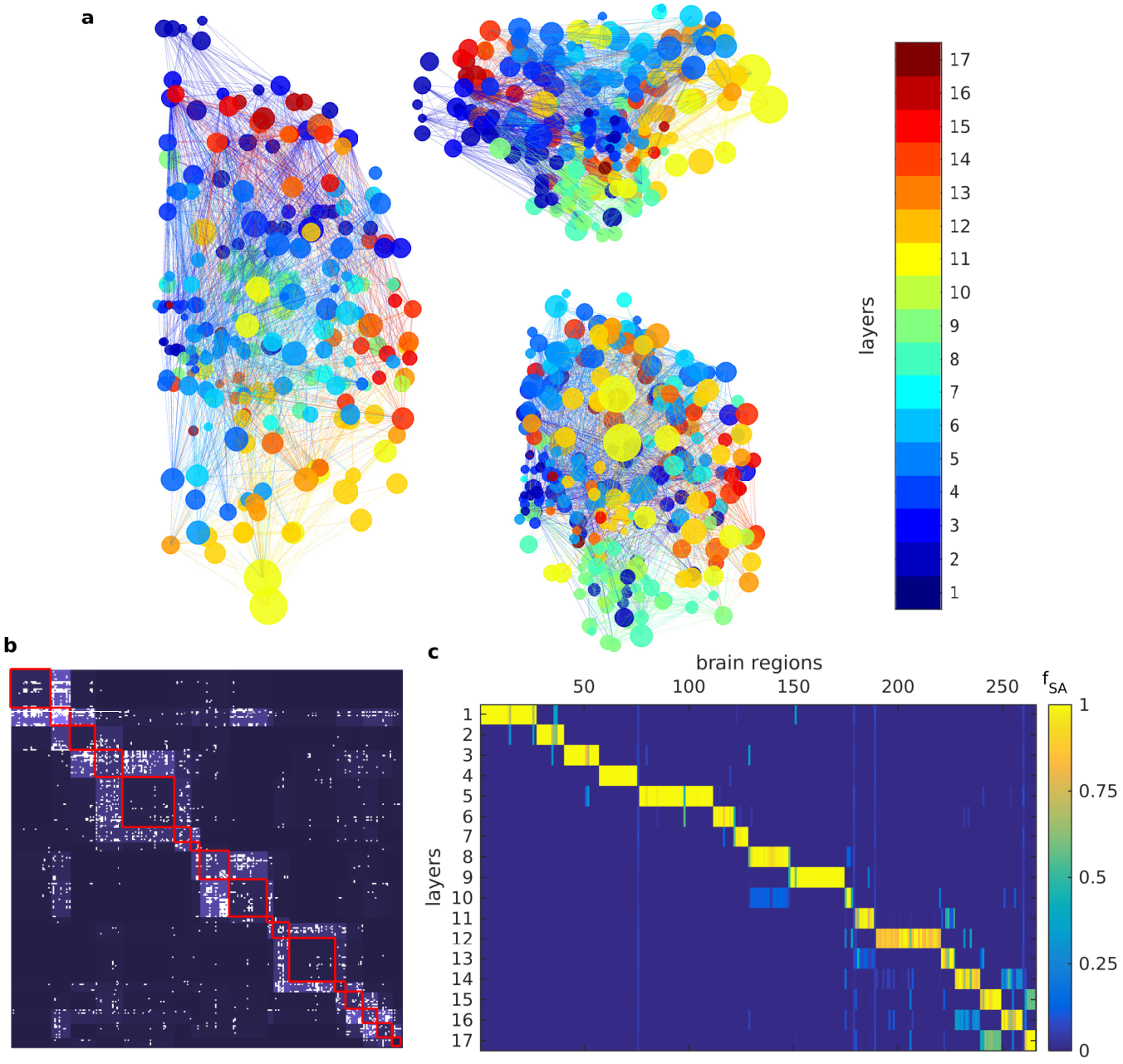


Figure 5.4: Hierarchical structure of the macaque brain network. (a) Planar projections of the spatial representation for the structural connectivity of the macaque brain hemisphere (as described in Chapter 3) on the horizontal (left panel), sagittal (top right panel) and coronal (bottom right) planes. Each node represents a brain region and links represent directed axonal tracts between the brain regions.

Figure 5.4 ***(previous page)*:** The size of the nodes indicate the spatial volumes occupied by corresponding brain regions and color of the nodes represent their respective hierarchical levels obtained from one of the realizations of hierarchy detection method (see legend). The links are colored in accordance with the corresponding source node. (b) The adjacency matrix representing the macaque connectome, where nodes (brain regions) are rearranged according to the levels obtained from one of the realizations using the hierarchy detection method. The white entries represent unweighted directed links and the connection density between brain regions belonging to any two distinct levels, as well as within-level connectivity for each level, is indicated by the brightness of the corresponding block in the matrix. Here, the high density of connections seen between consecutive levels are indicative of the strongly hierarchical organization in the underlying network. (c) Variability in the hierarchical organization obtained across all the realizations of hierarchy detection. The fraction of simulated annealing realizations f_{SA} in which a particular brain region (abscissa) is placed in a given level (ordinate) is indicated by the color coding provided in the legend. The brain regions are arranged according to the levels obtained from one of the realizations using the hierarchy detection method, and is used as a reference to compare all other realizations. We observe that almost all brain regions display a highly consistent membership in the respective levels across different realizations, as indicated by high values of f_{SA}. This indicates a highly robust hierarchical organization with an underlying sequence of brain connectivity at the mesoscopic level (see the diagonal arrangement of yellow-shaded blocks across the levels).

The final network whose hierarchical organization we analyzed in this study is the "basal" component of the representative human brain network, as described in Chapter 4 of this thesis. The basal network, as determined from the cohort study in Chapter 4, comprises the ubiquitous structural connections, i.e., connections that are seen to occur in all the individuals of the cohort, and which are distinguished by extremely low variability of connection weights across individuals (see Methods). The analysis of the hierarchy of the basal network is hence expected to be of more general applicability as compared to, for instance, analyzing a connectome obtained from a single individual. Fig. 5.5 (a) shows the spatial configuration of the hierarchical levels. The first striking observation is that the spatial arrangement of levels across the left and the right hemispheres are qualitatively distinct from each other, such that the directions of level progression in the two hemispheres are orthogonal to each other. Although the levels are spatially contiguous in both the hemispheres, the levels in left hemisphere progress across the anterior-posterior direction, i.e. from the frontal lobe to the left temporal lobe eventually to the occipital lobe (see sagittal section, top right panel), while in the right hemisphere, the levels progress

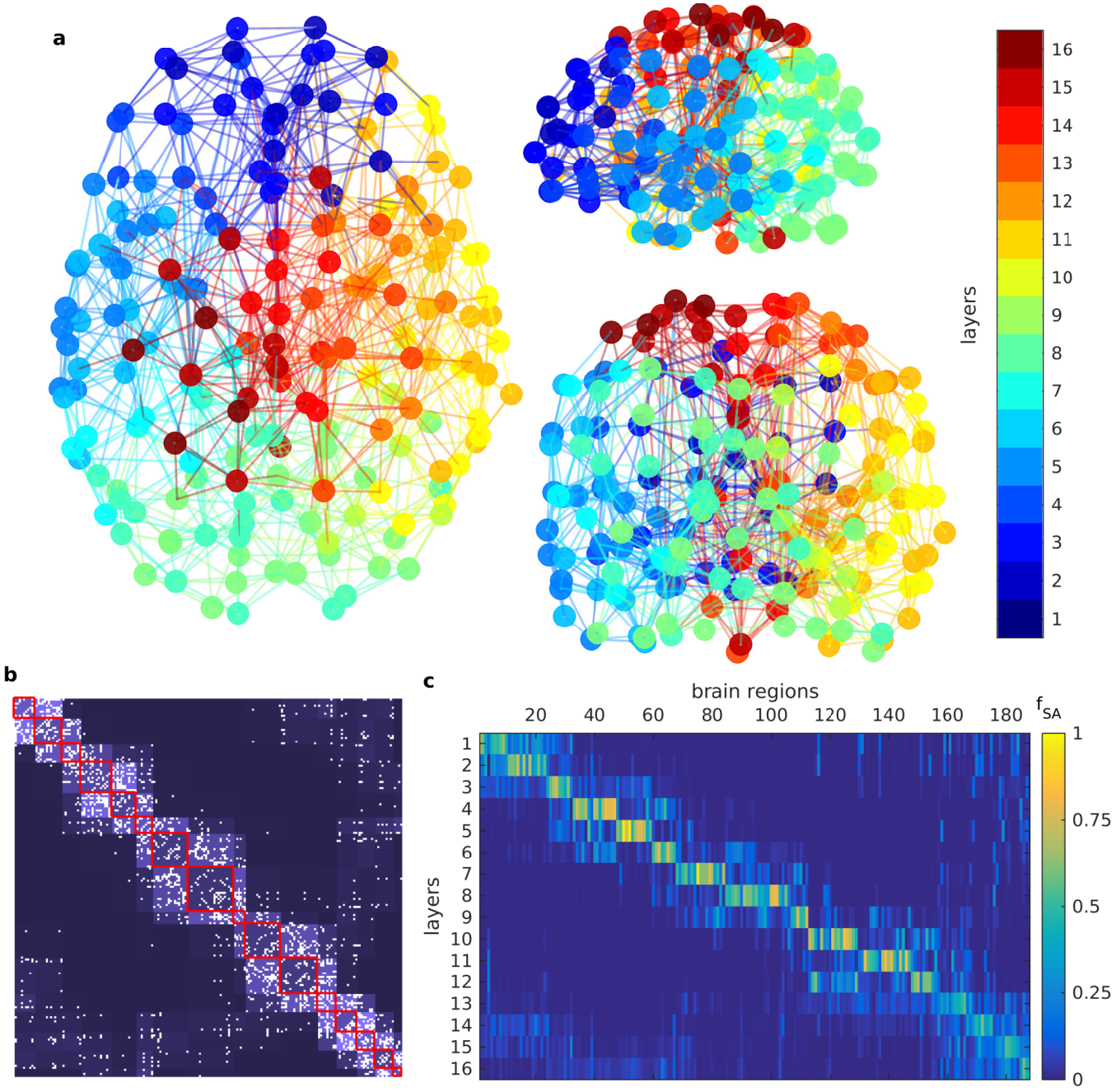


Figure 5.5: Hierarchical structure of the human brain network. (a) Planar projections of spatial representations for the structural connectivity of the "basal" network from the human connectome (as described in Chapter 4) on the horizontal (left panel), sagittal (top right panel) and coronal (bottom right) planes. Each node represents a brain region and links represent axonal tracts between the brain regions.

Figure 5.5 ***(previous page)*:** The color of the nodes represent their respective hierarchical levels obtained from one of the realizations of the hierarchy detection method (see legend). The links are colored in accordance with the corresponding source node. (b) The adjacency matrix representing the basal network of the human connectome where nodes (brain regions) are rearranged according to the levels obtained from one of the realizations using the hierarchy detection method. The white entries represent unweighted directed links and the connection density between brain regions belonging to any two distinct levels, as well as within-level connectivity for each level, is indicated by the brightness of the corresponding block in the matrix. The high density of connections seen between consecutive levels are indicative of the strongly hierarchical organization of the underlying network. (c) Variability in the hierarchical organization obtained across all the realizations of hierarchy detection. The fraction of simulated annealing realizations f_{SA} in which a particular brain region (abscissa) is placed in a given level (ordinate) is indicated by the color coding provided in the legend. The brain regions are arranged according to the levels obtained in one of the realizations of the hierarchy detection method, and is used as a reference to compare all other realizations. We observe that a significant number of brain regions display a highly consistent membership in the respective levels across different realizations, as indicated by high values of f_{SA}. This indicates a highly robust hierarchical organization with an underlying sequence of brain connectivity at the mesoscopic level (see the diagonal arrangement of yellow-shaded blocks across the levels).

across the dorsal-ventral direction, i.e., from right temporal lobe to the parietal lobe (see coronal section, bottom right panel). This lateral asymmetry in the hierarchical arrangement is particularly intriguing. The strength and robustness of the hierarchy in the human brain network is shown in Fig. 5.5 (b-c). As in the previously examined cases, we observe a dense inter-level connectivity between consecutive levels (Fig. 5.5 (b)), and a significant number of brain regions occurring consistently in their respective levels across many realizations (Fig. 5.5 (c)). The robustness however, is far lower than that of the macaque brain network (see Fig. 5.4 (c)).

In Table 5.1, we display a comparison between the three connectomes analyzed above, viz., *C. elegans*, macaque and human, in terms of the three main parameters characterizing their hierarchical organization: the hierarchy index H, the ratio ρ/ρ_{con} and the total number of levels N_{levels}. The mean and standard deviation of these quantities, computed over 200 realizations, are displayed in each case. The results suggest that the hierarchy in the *C. elegans* connectome appears to be relatively weaker than the other two, as indicated by lower values of H and higher values of ρ/ρ_{con}.

Table 5.1: Quantitative comparison between the hierarchical structures of neuronal/brain networks of *C. elegans*, macaque and human

	C. elegans	**Macaque**	**Human**
H	0.332 ± 0.003	0.423 ± 0.001	0.422 ± 0.005
ρ/ρ_{con}	0.173 ± 0.013	0.133 ± 0.010	0.111 ± 0.010
N_{levels}	14.90 ± 0.95	16.86 ± 0.86	17.74 ± 1.86

5.4 Discussion

In this chapter we propose a novel paradigm for studying the mesoscopic organization of complex networks, viz., their hierarchical organization. This notion of hierarchy is distinct from the hierarchy manifest in the hierarchical modular organization of networks [256, 117, 257, 10]. Our method follows an approach broadly similar to modularity detection algorithms in that it aims to identify mesoscopic sub-components (viz., hierarchical levels in our case). However, the connectivity among the levels of a hierarchical network strictly adhere to their sequential arrangement, where only the adjacent levels are densely connected to each other. Hence, while modularity represents a structural compartmentalization within a network, the hierarchy represents a "sequentialization" of interactions between the network nodes. In order to quantify the extent of hierarchy in a given network, we have defined a new measure, viz., the hierarchy index. Using the this index as the quantity to be optimized, we have developed a heuristic simulated annealing routine that can identify the optimal layer partitioning and their corresponding sequential arrangement of levels which would maximize the hierarchy index.

Along with a detailed description of our hierarchy detection method, in this chapter we report results from our preliminary analysis of some of the brain networks in terms of this still-developing paradigm. Although preliminary, the results shown here, demonstrate the potential for opening up new directions of inquiry regarding the structural organization of brain networks. In addition to demonstrating that the three brain networks analyzed here are strongly hierarchical, having well defined sequentially arranged levels, we also showed that these hierarchical levels are robust and largely invariant across realizations. This suggests that the observed levels are an inherent feature of the network, as opposed

to being an artifact of the specific method that we implement here.

Our findings from the analysis of the *C. elegans* neuronal network is particularly intriguing. We observed a counter-intuitive trend in the composition of the underlying hierarchical levels in terms of sensory, inter- and motor neurons. It is conventionally known that sensory neurons act as the 'doorway' of the nervous system to the external stimuli, interneurons behave as the intermediate processors of the sensory stimuli, and the motor neurons function as the actuators of the organisms response to the stimuli. Given these broad functional descriptions, one would expect the initial levels of the hierarchical arrangement to be composed primarily of sensory neurons, the subsequent levels to comprise mainly of inter-neurons, which in turn would be followed by levels consisting of motor neurons. However, in contrast to expectations, we observe the predominance of sensory neurons in the intermediate levels instead of the initial levels, while the interneurons dominated the levels that preceded the sensory-dominated levels in the hierarchy. This reveals a very novel structural aspect of the *C. elegans* nervous system, especially from the perspective of information processing.

The spatially contiguous progression of levels seen in the macaque and human brain networks are equally intriguing for their own specific reasons. In the macaque brain network (comprising a single hemisphere), we observe potentially two separate streams of information processing that progress in parallel along the anterior-posterior axis. One stream extends from the frontal lobe to the occipital lobe via the parietal lobe, whereas the other extends from the occipital lobe to frontal lobe via the temporal lobe. The functional implications of this architecture needs to be investigated thoroughly, along the same lines as our modular analysis of the macaque brain in Chapter 3. Similarly, the spatial progression of levels in the human brain (which comprises both the hemispheres) reveals a mutually orthogonal arrangement in the two hemispheres. In the left hemisphere, the levels progress along the anterior-posterior axis from the frontal lobe to occipital lobe, whereas in the right hemisphere the levels progress along the dorsal ventral axis from the right temporal lobe to the parietal lobe. In future work, it would be intriguing to see how this orthogonal arrangement of hierarchically connected levels in the left and right hemi-

spheres may be related to the specific functional specializations of the two hemispheres. One instance of such functional specialization is speech and language processing, which is primarily executed in the left hemisphere. One could potentially investigate whether the left-right asymmetry in the hierarchical organization in the human brain manifests in such functional specialization.

Chapter 6

Conclusions

"I am a brain, Watson. The rest of me is a mere appendix. Therefore, it is the brain I must consider."

–Arthur Conan Doyle, *"The Adventure of the Mazarin Stone"*

The work described in this thesis has as its primary focus the investigation of different aspects of structural connectivity in three diverse neuronal systems, viz., the somatic nervous system of the nematode *C. elegans* and the macaque and human brains. The central underlying idea is that as the wiring between neurons and brain areas provides the anatomical basis for behavior and cognition, by uncovering patterns in the connectome we can gain insights about how brains function, and about the factors that shape the organization of the network. In the following subsections, we summarize the key results that are reported in the thesis, and conclude by providing a glimpse of possible future investigations that are suggested by our work.

6.1 Summary of main results

Developmental trajectory of *Caenorhabditis elegans* nervous system governs its structural organization

A central problem of neuroscience involves uncovering the principles governing the organization of nervous systems which ensure robustness in brain development. The nematode *Caenorhabditis elegans* provides us with a model organism for studying this question. In this chapter, we focus on the invariant connection structure and spatial arrangement of the neurons comprising the somatic neuronal network of this organism to understand the key developmental constraints underlying its design. We observe that neurons with certain shared characteristics - such as, neural process lengths, birth time cohort, lineage and bilateral symmetry - exhibit a preference for connecting to each other. Recognizing the existence of such homophily and their relative degree of importance in determining connection probability within neurons (for example, in synapses, symmetric pairing is the most dominant factor followed by birth time cohort, process length and lineage) helps in connecting specific neuronal attributes to the topological organization of the network. Further, the functional identities of neurons appear to dictate the temporal hierarchy of their appearance during the course of development. Providing crucial insights into principles that may be common across many organisms, our study shows how the trajectory in the developmental landscape constrains the structural organization of a nervous system.

Mesoscopic architecture enhances communication across the Macaque connectome revealing structure-function correspondence in the brain

Analyzing the brain in terms of organizational structures at intermediate scales provides an approach to negotiate the complexity arising from the interactions between its large number of components. Focusing on a wiring diagram that spans the cortex, basal ganglia and

thalamus of the Macaque brain, we provide a more insightful description of the topological architecture at the mesoscopic level of one of the most well-studied of mammalian connectomes. The robust modules we identify each comprise densely inter-connected cortical and sub-cortical areas that complement each other in carrying out specific cognitive functions. While the areas in each module are broadly spatially contiguous, we find that this physical proximity cannot by itself explain the modular organization, as a similar mesoscopic structure can be obtained even after factoring out the effect of distance constraints on the connectivity. We observe that the distribution profile of brain areas, classified in terms of their intra- and inter-modular connectivity, is conserved across the principal cortical subdivisions, as well as, sub-cortical structures. In particular provincial hubs, which have significantly higher number of connections with other members of their module, but relatively less well-connected to other modules, are the only class that exhibits homophily, i.e., a discernible preference to connect to each other. By considering a process of diffusive propagation between areas, we demonstrate that this architecture facilitates rapid communication across the connectome. Our work, which supplements the topological information about the Macaque connectome with that of physical locations, volumes and functions of the constituent areas collated by us, suggests a counter-intuitive role played by the modular architecture of the brain in promoting global interaction instead of aiding in information encapsulation as might be assumed.

Invariances and diversity in the human structural connectome : A cohort study

In order to understand the complex cognitive functions of the human brain, it is essential to study the structural macro-connectome, i.e., the wiring of different brain regions to each other through axonal pathways, that has been revealed by the ongoing progress in non-invasive imaging techniques such as diffusion tensor imaging (DTI). However, relating structure to function is complicated by the high degree of plasticity and cross-population variability that are seen in human brains, motivating a search for invariant patterns in the

connectivity. At the same time, intra-species variability can provide information about the generative mechanisms at work. In this chapter we consider an ensemble of human structural connectomes obtained from 196 subjects and analyze the connection topology and link-weight distribution for each of these. By demonstrating a correspondence between the occurrence frequency of individual links and their average weight across the population, we show that the "wiring" process in the human brain is not independent of the "weighting" process. Furthermore, using the specific distribution of connection weights associated with each link over the entire population, we show that a single parameter that is specific to a link can account for its frequency of occurrence, as well as, the variation in its weight across different subjects. This parameter provides a basis for "rescaling" the link weights in each connectome, allowing us to eventually obtain a generic "representative" network for the human brain, that is more informative than simply averaging over all connectomes. By implementing a well known dynamic model for neural population activity on each of the vertices in the structural connectome to simulate the functional connectome, and comparing this with the empirical functional network, it is seen that the "rescaling" procedure yields a closer structure-function correspondence in the human brain. We also show that the representative network is divided into two components, a "basal" network that is stable across the population, and a highly variable "superstructure" network.

The Hierarchy Index: A tool for revealing the sequential arrangement of processing levels in brain networks

The mesoscopic organization of complex networks crucially affects their function. One such feature shown by many networks is hierarchy, involving dense feed-forward and feed-back connectivity between nodes occurring in sequentially arranged levels. Such an organization is associated with circuits in the brain that process information, e.g., in visual pathways, and has been suggested to occur in circuits responsible for executing collective neuronal firing in a precise temporal sequence, e.g., HVC neurons in songbird. However, there are no well-accepted methods for identifying such hierarchical structures from the network

topology alone. In this chapter, we propose a novel approach to determine the hierarchy in any given network, revealing the underlying levels and their sequential arrangement. We define a Hierarchy Index to quantify the degree of hierarchy present in a network. We also propose an optimization procedure using simulated annealing to obtain the optimal hierarchical configuration which would yield the maximal value of the Hierarchy Index. After validating the accuracy of the proposed method using synthetic benchmark networks, we use it to determine the hierarchical arrangement embedded within the *C. elegans* somatic nervous system, the macaque connectome and the human connectome. We find that all three networks are strongly hierarchical having robust levels (consistent across many realizations of the optimization routine), the hierarchy of the Macaque connectome being the most robust among the three. Our results provide a basis for possible future investigations into the functional implications of the specific hierarchical organization of a brain network.

6.2 Outlook

The systems analyzed in this thesis, viz., *C. elegans*, macaque and human structural connectomes, vary greatly in terms of their length scale, complexity, as well as, resolution. While the *C. elegans* connectivity is described at the neuronal level and encompasses the entire somatic nervous system with a description of all the chemical synapses and electrical gap junctions, the macaque brain is described at a lower resolution, comprising brain areas belonging to an entire hemisphere that are connected by directed, unweighted links. For the human brain, we analyze an ensemble of structural connectomes where each network comprises areas across both the hemispheres that are connected by links whose weights, as well as, connection topology is available. This diversity allows us to address distinct scientific questions using each of the systems.

The intra-species invariability of neuronal connectivity in *C. elegans* renders it suitable for addressing the "wiring problem", i.e., how the genetic program and developmental processes of an organism governs the structural layout of its neuronal network. We uncover a set of key strategies that may connect the level at which the molecular mechanisms of

gene expression operate with that at which the resulting neuronal network in the adult organism becomes manifest. We have shown that genetic relatedness plays a significant role in deciding connectivity. A natural extension of this investigation would be to consider the dynamics of gene regulation within the cells in order to bridge the gap between the machinery of gene expression and the wiring of the neuronal network. A similar attempt has been made for brain networks of rodents [269, 270]. This bridging of networks that occur at two distinct scales can be expected to be more fruitful in the case of *C. elegans* as we not only have complete information about the wiring diagram, but we also know the entire lineage tree for all the cells of the organism. The lineage relation between neurons would be a key link that connects the gene expression within neurons and the resultant neuronal connectivity.

Our analysis of the macaque brain reveals a very significant property of the connectome that can be attributed to the specific mesoscopic arrangement of connectivity within and between different modules, viz., the facilitation of rapid communication at both the local and global scales. This suggests a novel evolutionary purpose for the emergence of modularity in brain networks, apart from roles which have traditionally been attributed to it, such as information encapsulation or robustness. This raises the question as to whether this property may be conserved across species? For this purpose, one would need to examine brain networks of other organisms in terms of their intra- and inter-modular connectivity. For example, if one finds that networks as different as that of the *C. elegans* somatic nervous system and the human brain have similar mesoscopic arrangement that enables rapid communication across the network compared to randomized surrogate ensembles, it would establish more strongly that the modular organization of nervous systems that is observed across species is specifically designed for fast transmission of information. Further, as we discovered that the structural modules seen in the macaque brain are not a consequence of spatial constraints, it would be of interest to know whether the modules in the human structural connectome are also spatially independent. The representative network for human brain obtained from our study would be a suitable starting point for such an investigation.

Apart from looking for the spatial independence of modules in the representative network of human brain, one can undertake a study of the modular organization in human brains across the population. We have only scratched the surface by analyzing the variability in connection weights of the underlying links. The next step would be to analyze the cross-population variation in the modular structures in the human connectome. Since the networks in this case would be weighted, it would add an extra layer of complexity to the mesoscopic analysis and determination of structure-function correspondences. As the human brain connectome comprises both the hemispheres, another possible aspect to study in the human connectome could be the lateralization of the network across left and right hemispheres. This could be done by analyzing the degree of left-right asymmetry within an individual brain network and also across the population. While functional lateralization has been extensively studied, i.e., how left and right hemispheres are specialized for separate functions, there are as yet no extensive and systematic studies of structural lateralization.

The most pressing question however, which calls for a separate investigation, is: how are the structural links correlated to each other across the individual brains of human subjects ? We have already shown that the Poisson distribution of connection weights suggests that the generative process involves discrete independent random events. In this, we have not considered that the links can influence each other. As we know that various cognitive functions are executed through brain circuits involving a subset of brain regions and their connections, it may be reasonable to expect that the link weights of connections belonging to a specific circuit are correlated across the population. Thus, if link (i, j) and link (k, l) belong to the same circuit corresponding to a particular function, we may expect their weights across different individuals to be significantly correlated. Conversely, observation of significant correlations among a pair of links might indicate that these links belong to a specific circuit. Using data analysis tools such as Principal Component Analysis one can search for such correlations. However, a major bottleneck for such a study would be the paucity of the number of connectomes (≈ 200) compared to the number of links (≈ 4000) which would impair the statistical robustness of the findings.

Finally, using our proposed method for determining the hierarchical organization of networks, a large number of systems can be investigated. A detailed study of the functional implications of the observed hierarchical levels in the three systems we have analyzed can be a separate study in its own right. Thus, the results presented here could be the harbinger of a much more detailed understanding of the implications of the hierarchical organization in nervous systems.

Appendix A

The following tables and figures comprise the supporting information for the analysis described in Chapter 2.

Table A1: Process length homophily among neurons segregated into groups comprising cells with long, medium and short processes. The extent of homophily is quantified by the modularity measure Q computed over the different classes of neurons (which are considered to be the communities or modules for the purpose of calculation of Q). The empirical Q values are compared with those calculated from two different types of randomized surrogate ensembles. Members of one of these ensembles are constructed by randomly shuffling the process length categories of the empirical network keeping the network connections invariant, while those belonging to the other ensemble are obtained by degree preserved randomization of the empirical network with the cell positions and process lengths kept unchanged (top and center, respectively). Note that the Q values are significantly higher than that expected by chance (as seen for the surrogate ensembles) for the entire network, as well as, individually within each of the different process length categories, suggesting process length homophily in both synaptic and gap-junction connections between neurons. (Bottom) Note that significantly higher values of class-specific Q (compared to the randomized surrogate) occur only when we consider pairs of neurons belonging to the same class, further underlining the process length homophily.

Comparison of Q value for empirical network and ensemble of surrogate networks obtained by randomly shuffling process lengths of neurons keeping network unchanged

	Synaptic		**Gap junction**	
Process	Q (empirical)	Q (randomized)	Q (empirical)	Q (randomized)
Long	0.042	-0.001 ± 0.003	0.073	-0.001 ± 0.004
Medium	0.016	-0.001 ± 0.003	0.022	-0.001 ± 0.005
Short	0.067	-0.001 ± 0.005	0.085	-0.002 ± 0.008
All	0.125	-0.003 ± 0.008	0.18	-0.003 ± 0.013

Table A1 continued from previous page

Comparison of Q value for empirical network and ensemble of surrogate networks obtained by randomly shuffling connections keeping process length, positions of cell bodies and node degree unchanged

	Synaptic		**Gap junction**	
Process	Q (empirical)	Q (randomized)	Q (empirical)	Q (randomized)
Long	0.042	0.000 ± 0.003	0.073	-0.012 ± 0.007
Medium	0.016	0.001 ± 0.002	0.022	0.002 ± 0.005
Short	0.067	0.003 ± 0.005	0.085	-0.006 ± 0.008
All	0.125	-0.004 ± 0.008	0.18	-0.016 ± 0.014

Z-scores of class specific Q values obtained by comparing empirical values with surrogate ensemble of randomized networks obtained by randomly shuffling connections keeping process length, positions of cell bodies and node degree unchanged

Synaptic

	Short	Medium	Long
Short	**14.77**	-6.07	-14.73
Medium	-6.43	**5.57**	2.13
Long	-13.93	1.67	**15.39**

Table A1 continued from previous page

Gap junction

	Short	Medium	Long
Short	**11.16**	-3.34	-13.97
Medium	-3.34	**4.55**	-1.5
Long	-13.97	-1.5	**16.27**

Table A2: Birth cohort homophily among neurons segregated into groups comprising cells differentiating before and after hatching (early and late-born, respectively). The extent of homophily is quantified by modularity Q computed over the different classes of neurons (which are considered to be the communities or modules for the purpose of calculating Q). (Top) The empirical values are compared with those calculated from the corresponding surrogate ensemble obtained by degree preserved randomization with the cell positions and process lengths of neurons kept unchanged. Note that the Q values are significantly higher than that expected by chance (as seen from the surrogate ensemble) for the entire network, as well as, individually for the early-born and late-born cohorts. This is indicative of birth cohort homophily for synaptic, as well as, gap-junctional connections between neurons. (Bottom) The z-scores for class-specific Q values computed with respect to the randomized surrogate ensemble. Significantly higher values of Q occur only when we consider pairs of neurons belonging to the same category (early- or late-born) and not for pairs where the constituent neurons belong to different categories, further underlining the homophily.

Comparison of Q value for empirical network and ensemble of surrogate networks obtained by randomly shuffling connections keeping process length, positions of cell bodies and node degree unchanged

	Synaptic		**Gap junction**	
Birth cohort	Q (empirical)	Q (randomized)	Q (empirical)	Q (randomized)
Early	0.044	0.010 ± 0.003	0.035	0.013 ± 0.005
Late	0.042	0.009 ± 0.003	0.035	0.002 ± 0.005
Total	0.087	0.02 ± 0.005	0.07	0.026 ± 0.011

Table A2 continued from previous page

Z-scores of class specific Q values obtained by comparing empirical values with surrogate ensemble of randomized networks obtained by randomly shuffling connections keeping process length, positions of cell bodies and node degree unchanged

	Synaptic	
	Early	Late
Early	**13.8**	-13.57
Late	-13.4	**13.24**

	Gap junction	
	Short	Medium
Early	**4.2**	-4.19
Medium	-4.19	**4.18**

Table A3: For synaptically connected neurons, process length of the pre-synaptic neuron primarily decides the average distance between the cell bodies. Statistically significant deviation (measured in terms of z-score) between the average distance $\langle d \rangle$ of cell bodies in pairs of connected neurons (having short, medium or long processes) and the average distance $\langle D \rangle$ between any pair of neurons randomly sampled from the same process length categories. The latter average is calculated over a set having the same number of pairs as for the set of connected pairs. Note that except for two cases (pre-synaptic long process to post-synaptic short process and pre-synaptic medium process to post-synaptic long process, shown in bold font), connections between cells in all other process length categories tend to be much shorter than that expected by chance, as indicated by $z < 0$.

Synapse

	short	medium	long
short	-19.71	-4.54	**1**
medium	-12.03	-8.11	-0.69
long	-7.84	**1.41**	-1.5

Table A3 continued from previous page

Gap Jn.

	short	medium	long
short	-12.7	-7.65	-2.41
medium	-7.65	-7.72	-4.1
long	-2.41	-4.1	-1.61

Table A4: Neurons belonging to the somatic nervous system of *C. elegans* segregated into those which are born in the early (embryonic) and those born in the late (post-embryonic) developmental bursts. The respective lineage information and functional description are also provided. Motor neurons are highlighted. We note that motor neurons that appear early mostly innervate dorsal muscles, whereas, motor neurons that appear late primarily innervate ventral muscles. All information shown here is obtained from WormAtlas [98].

Early Born Neurons		
Neurons	**Lineage**	**Description**
ADAL	AB plapaaaapp	Ring interneuron
ADAR	AB prapaaaapp	Ring interneuron
ADEL	AB plapaaaapa	Anterior deirid, sensory neuron
ADER	AB prapaaaapa	Anterior deirid, sensory neuron
ADFL	AB alpppppaa	Amphid neuron
ADFR	AB praaappaa	Amphid neuron
ADLL	AB alppppaad	Amphid neuron
ADLR	AB praaapaad	Amphid neuron
AFDL	AB alpppapav	Amphid finger cell
AFDR	AB praaaapav	Amphid finger cell
AIAL	AB plppaappa	Amphid interneuron
AIAR	AB prppaappa	Amphid interneuron
AIBL	AB plaapappa	Amphid interneuron
AIBR	AB praapappa	Amphid interneuron
AIML	AB plpaapppa	Ring interneuron
AIMR	AB prpaapppa	Ring interneuron
AINL	AB alaaaalal	Ring interneuron
AINR	AB alaapaaar	Ring interneuron
AIYL	AB plpapaaap	Amphid interneuron
AIYR	AB prpapaaap	Amphid interneuron
AIZL	AB plapaaapav	Amphid interneuron
AIZR	AB prapaaapav	Amphid interneuron
ALA	AB alapppaaa	Neuron, sends processes laterally and along dorsal cord
ALML	AB arppaappa	Anterior lateral microtubule cell
ALMR	AB arpppappa	Anterior lateral microtubule cell

Table A4 continued

ALNL	AB plapappppap	Neuron associated with ALM
ALNR	AB prapappppap	Neuron associated with ALM
ASEL	AB alpppppppaa	Amphid neurons, single ciliated endings
ASER	AB praaapppaa	Amphid neurons, single ciliated endings
ASGL	AB plaapapap	Amphid neurons, single ciliated endings
ASGR	AB praapapap	Amphid neurons, single ciliated endings
ASHL	AB plpaappaa	Amphid neurons, single ciliated endings
ASHR	AB prpaappaa	Amphid neurons, single ciliated endings
ASIL	AB plaapapppa	Amphid neurons, single ciliated endings
ASIR	AB praapapppa	Amphid neurons, single ciliated endings
ASJL	AB alpppppppa	Amphid neurons, single ciliated endings
ASJR	AB praaappppa	Amphid neurons, single ciliated endings
ASKL	AB alpppapppa	Amphid neurons, single ciliated endings
ASKR	AB praaaapppa	Amphid neurons, single ciliated endings
AUAL	AB alpppppppp	Neuron, process runs with amphid processes but lacks ciliated ending
AUAR	AB praaappppp	Neuron, process runs with amphid processes but lacks ciliated ending
AVAL	AB alppaaapa	Ventral cord interneuron
AVAR	AB alaappapa	Ventral cord interneuron
AVBL	AB plpaapaap	Ventral cord interneuron
AVBR	AB prpaapaap	Ventral cord interneuron
AVDL	AB alaaapalr	Ventral cord interneuron
AVDR	AB alaaapprl	Ventral cord interneuron
AVEL	AB alpppaaaa	Ventral cord interneuron, like AVD but outputs restricted to anterior cord
AVER	AB praaaaaaa	Ventral cord interneuron, like AVD but outputs restricted to anterior cord
AVG	AB prpapppap	Ventral cord interneuron
AVHL	AB alapaaaaa	Neuron, mainly postsynaptic in ventral cord and presynaptic in the ring
AVHR	AB alappapaa	Neuron, mainly postsynaptic in ventral cord and presynaptic in the ring
AVJL	AB alapapppa	Neuron, synapses like AVHL/R
AVJR	AB alappppppa	Neuron, synapses like AVHL/R
AVKL	AB plpapapap	Ring and ventral cord interneuron
AVKR	AB prpapapap	Ring and ventral cord interneuron

Table A4 continued

AVL	AB prpappaap	Ring and ventral cord interneuron and an excitatory GABAergic motor neuron for rectal muscles. Few synapses
AWAL	AB plaapapaa	Amphid wing cells, neurons having ciliated sheet-like sensory endings closely associated with amphid sheath
AWAR	AB praapapaa	Amphid wing cells, neurons having ciliated sheet-like sensory endings closely associated with amphid sheath
AWBL	AB alpppppap	Amphid wing cells, neurons having ciliated sheet-like sensory endings closely associated with amphid sheath
AWBR	AB praaappap	Amphid wing cells, neurons having ciliated sheet-like sensory endings closely associated with amphid sheath
AWCL	AB plpaaaaap	Amphid wing cells, neurons having ciliated sheet-like sensory endings closely associated with amphid sheath
AWCR	AB prpaaaaap	Amphid wing cells, neurons having ciliated sheet-like sensory endings closely associated with amphid sheath
BAGL	AB alppappap	Neuron, ciliated ending in head, no supporting cells, associated with ILso
BAGR	AB arappppap	Neuron, ciliated ending in head, no supporting cells, associated with ILso
BDUL	AB arppaappp	Neuron, process runs along excretory canal and into ring, unique darkly staining synaptic vesicles
BDUR	AB arpppappp	Neuron, process runs along excretory canal and into ring, unique darkly staining synaptic vesicles
CEPDL	AB plaaaaappa	Cephalic neurons, contain dopamine
CEPDR	AB arpapaappa	Cephalic neurons, contain dopamine
CEPVL	AB plpaappppa	Cephalic neurons, contain dopamine
CEPVR	AB prpaappppa	Cephalic neurons, contain dopamine
DA1	AB prppapaap	Ventral cord motor neurons, innervate dorsal muscles
DA2	AB plppapapa	Ventral cord motor neurons, innervate dorsal muscles
DA3	AB prppapapa	Ventral cord motor neurons, innervate dorsal muscles
DA4	AB plppapapp	Ventral cord motor neurons, innervate dorsal muscles
DA5	AB prppapapp	Ventral cord motor neurons, innervate dorsal muscles
DA6	AB plpppaaap	Ventral cord motor neurons, innervate dorsal muscles
DA7	AB prpppaaap	Ventral cord motor neurons, innervate dorsal muscles

Table A4 continued

DA8	AB prpapappp	Ventral cord motor neurons, innervate dorsal muscles
DA9	AB plpppaaaa	Ventral cord motor neurons, innervate dorsal muscles
DB1	AB plpaaaapp	Ventral cord motor neurons, innervate dorsal muscles, reciprocal inhibitor
DB2	AB arappappa	Ventral cord motor neurons, innervate dorsal muscles, reciprocal inhibitor
DB3	AB prpaaaapp	Ventral cord motor neurons, innervate dorsal muscles, reciprocal inhibitor
DB4	AB prpappapp	Ventral cord motor neurons, innervate dorsal muscles, reciprocal inhibitor
DB5	AB plpapappp	Ventral cord motor neurons, innervate dorsal muscles, reciprocal inhibitor
DB6	AB plppaappp	Ventral cord motor neurons, innervate dorsal muscles, reciprocal inhibitor
DB7	AB prppaappp	Ventral cord motor neurons, innervate dorsal muscles, reciprocal inhibitor
DD1	AB plppappap	Ventral cord motor neurons, reciprocal inhibitors, change synaptic pattern during L1
DD2	AB prppappap	Ventral cord motor neurons, reciprocal inhibitors, change synaptic pattern during L1
DD3	AB plppapppa	Ventral cord motor neurons, reciprocal inhibitors, change synaptic pattern during L1
DD4	AB prppapppa	Ventral cord motor neurons, reciprocal inhibitors, change synaptic pattern during L1
DD5	AB plppapppp	Ventral cord motor neurons, reciprocal inhibitors, change synaptic pattern during L1
DD6	AB prppapppp	Ventral cord motor neurons, reciprocal inhibitors, change synaptic pattern during L1
DVA	AB prpppppapp	Ring interneurons, cell bodies in dorsorectal ganglion
DVC	C aapaa	Ring interneurons, cell bodies in dorsorectal ganglion
FLPL	AB plapaaapad	Neuron, ciliated ending in head, no supporting cells, associated with ILso
FLPR	AB prapaaapad	Neuron, ciliated ending in head, no supporting cells, associated with ILso
HSNL	AB plapppappa	Hermaphrodite specific motor neurons (die in male embryo), innervate vulval muscles, serotonergic
HSNR	AB prapppappa	Hermaphrodite specific motor neurons (die in male embryo), innervate vulval muscles, serotonergic
IL1DL	AB alapappaaa	Inner labial neuron
IL1DR	AB alappppaaa	Inner labial neuron
IL1L	AB alapaappaa	Inner labial neuron
IL1R	AB alaappppaa	Inner labial neuron
IL1VL	AB alppapppaa	Inner labial neuron
IL1VR	AB arappppppaa	Inner labial neuron
IL2DL	AB alapappap	Inner labial neuron
IL2DR	AB alappppap	Inner labial neuron

Table A4 continued

IL2L	AB alapaappp	Inner labial neuron
IL2R	AB alaappppp	Inner labial neuron
IL2VL	AB alppappppp	Inner labial neuron
IL2VR	AB arappppppp	Inner labial neuron
LUAL	AB plpppaapap	Interneuron, short process in post ventral cord
LUAR	AB prpppaapap	Interneuron, short process in post ventral cord
OLLL	AB alpppapaa	Lateral outer labial neurons
OLLR	AB praaapapaa	Lateral outer labial neurons
OLQDL	AB alapapapaa	Quadrant outer labial neuron
OLQDR	AB alapppapaa	Quadrant outer labial neuron
OLQVL	AB plpaaappaa	Quadrant outer labial neuron
OLQVR	AB prpaaappaa	Quadrant outer labial neuron
PDA	AB prpppaaaa	Motor neuron, process in dorsal cord, same as Y cell in hermaphrodite, Y.a in male
PHAL	AB plpppaapp	Phasmid neurons, chemosensory
PHAR	AB prpppaapp	Phasmid neurons, chemosensory
PHBL	AB plapppappp	Phasmid neurons, chemosensory
PHBR	AB prapppappp	Phasmid neurons, chemosensory
PLML	AB plapappppaa	Posterior lateral microtubule cells, touch receptor neurons
PLMR	AB prapappppaa	Posterior lateral microtubule cells, touch receptor neurons
PVCL	AB plpppaapaa	Ventral cord interneuron, cell body in lumbar ganglion, synapses onto VB and DB motor neurons, formerly called delta
PVCR	AB prpppaapaa	Ventral cord interneuron, cell body in lumbar ganglion, synapses onto VB and DB motor neurons, formerly called delta
PVPL	AB plpppaaa	Interneuron, cell body in preanal ganglion, projects along ventral cord to nerve ring
PVPR	AB prpppaaa	Interneuron, cell body in preanal ganglion, projects along ventral cord to nerve ring
PVQL	AB plapppaaa	Interneuron, projects along ventral cord to ring
PVQR	AB prapppaaa	Interneuron, projects along ventral cord to ring
PVR	C aappa	Interneuron, projects along ventral cord to ring
PVT	AB plpappppa	Interneuron, projects along ventral cord to ring
RIAL	AB alapaapaa	Ring interneuron, many synapses
RIAR	AB alaapppaa	Ring interneuron, many synapses

Table A4 continued

RIBL	AB plpaappap	Ring interneuron
RIBR	AB prpaappap	Ring interneuron
RICL	AB plppaaaapp	Ring interneuron
RICR	AB prppaaaapp	Ring interneuron
RID	AB alappaapa	Ring motor neuron/interneuron, projects along dorsal cord
RIFL	AB plppapaaap	Ring interneuron
RIFR	AB prppapaaap	Ring interneuron
RIGL	AB plppappaa	Ring interneuron
RIGR	AB prppappaa	Ring interneuron
RIH	AB prpappaaa	Ring interneuron
RIML	AB plppaapap	Ring motor neuron
RIMR	AB prppaapap	Ring motor neuron
RIPL	AB alpapaaaa	Ring/pharynx interneuron, only direct connection between pharynx and ring
RIPR	AB arappaaaa	Ring/pharynx interneuron, only direct connection between pharynx and ring
RIR	AB prpapppaa	Ring interneuron
RIS	AB prpappapa	Ring interneuron
RIVL	AB plpaapaaa	Ring interneuron
RIVR	AB prpaapaaa	Ring interneuron
RMDDL	AB alpapapaa	Ring motor neuron/interneuron, many synapses
RMDDR	AB arappapaa	Ring motor neuron/interneuron, many synapses
RMDL	AB alpppapad	Ring motor neuron/interneuron, many synapses
RMDR	AB praaaapad	Ring motor neuron/interneuron, many synapses
RMDVL	AB alppapaaa	Ring motor neuron/interneuron, many synapses
RMDVR	AB arapppaaa	Ring motor neuron/interneuron, many synapses
RMED	AB alapppaap	Ring motor neuron
RMEL	AB alaaaarlp	Ring motor neuron
RMER	AB alaaaarrp	Ring motor neuron
RMEV	AB plpappaaa	Ring motor neuron
RMGL	AB plapaaapp	Ring motor neuron/interneuron
RMGR	AB prapaaapp	Ring motor neuron/interneuron
SAADL	AB alppapapa	Ring interneuron, anteriorly projecting process that runs sublaterally

Table A4 continued

SAADR	AB arapppapa	Ring interneuron, anteriorly projecting process that runs sublaterally
SAAVL	AB plpaaaaaa	Ring interneuron, anteriorly projecting process that runs sublaterally
SAAVR	AB prpaaaaaa	Ring interneuron, anteriorly projecting process that runs sublaterally
SABD	AB plppapaap	Ring interneuron, anteriorly projecting process that runs sublaterally, synapses to anterior body muscles in L1
SABVL	AB plppapaaaa	Ring interneuron, anteriorly projecting process that runs sublaterally, synapses to anterior body muscles in L1
SABVR	AB prppapaaaa	Ring interneuron, anteriorly projecting process that runs sublaterally, synapses to anterior body muscles in L1
SIADL	AB plpapaapa	Receive a few synapses in the ring, have posteriorly directed processes that run sublaterally
SIADR	AB prpapaapa	Receive a few synapses in the ring, have posteriorly directed processes that run sublaterally
SIAVL	AB plpapappa	Receive a few synapses in the ring, have posteriorly directed processes that run sublaterally
SIAVR	AB prpapappa	Receive a few synapses in the ring, have posteriorly directed processes that run sublaterally
SIBDL	AB plppaaaaa	Similar to SIA
SIBDR	AB prppaaaaa	Similar to SIA
SIBVL	AB plpapaapp	Similar to SIA
SIBVR	AB prpapaapp	Similar to SIA
SMBDL	AB alpapapapp	Ring motor neuron/interneuron, has a posteriorly directed process that runs sublaterally
SMBDR	AB arappapapp	Ring motor neuron/interneuron, has a posteriorly directed process that runs sublaterally
SMBVL	AB alpapappp	Ring motor neuron/interneuron, has a posteriorly directed process that runs sublaterally
SMBVR	AB arappappp	Ring motor neuron/interneuron, has a posteriorly directed process that runs sublaterally
SMDDL	AB plpapaaaa	Ring motor neuron/interneuron, has a posteriorly directed process that runs sublaterally
SMDDR	AB prpapaaaa	Ring motor neuron/interneuron, has a posteriorly directed process that runs sublaterally
SMDVL	AB alppappaa	Ring motor neuron/interneuron, has a posteriorly directed process that runs sublaterally
SMDVR	AB arappppaa	Ring motor neuron/interneuron, has a posteriorly directed process that runs sublaterally
URADL	AB plaaaaaaa	Ring motor neuron
URADR	AB arpapaaaa	Ring motor neuron
URAVL	AB plpaaapaa	Ring motor neuron
URAVR	AB prpaaapaa	Ring motor neuron
URBL	AB plaapaapa	Neuron, presynaptic in ring, ending in head
URBR	AB praapaapa	Neuron, presynaptic in ring, ending in head

Table A4 continued

URXL	AB plaaaaappp	Ring interneuron
URXR	AB arpapaappp	Ring interneuron
URYDL	AB alapapapp	Neuron, presynaptic in ring, ending in head
URYDR	AB alapppapp	Neuron, presynaptic in ring, ending in head
URYVL	AB plpaaappp	Neuron, presynaptic in ring, ending in head
URYVR	AB prpaaappp	Neuron, presynaptic in ring, ending in head

Late Born Neurons		
Neurons	**Lineage**	**Description**
AQR	QR.ap	Neuron, basal body. not part of a sensillum, projects into ring
AS1	P1.apa	Ventral cord motor neuron, innervates dorsal muscles, no ventral counterpart
AS2	P2.apa	Ventral cord motor neuron, innervates dorsal muscles, no ventral counterpart
AS3	P3.apa	Ventral cord motor neuron, innervates dorsal muscles, no ventral counterpart
AS4	P4.apa	Ventral cord motor neuron, innervates dorsal muscles, no ventral counterpart
AS5	P5.apa	Ventral cord motor neuron, innervates dorsal muscles, no ventral counterpart
AS6	P6.apa	Ventral cord motor neuron, innervates dorsal muscles, no ventral counterpart
AS7	P7.apa	Ventral cord motor neuron, innervates dorsal muscles, no ventral counterpart
AS8	P8.apa	Ventral cord motor neuron, innervates dorsal muscles, no ventral counterpart
AS9	P9.apa	Ventral cord motor neuron, innervates dorsal muscles, no ventral counterpart
AS10	P10.apa	Ventral cord motor neuron, innervates dorsal muscles, no ventral counterpart
AS11	P11.apa	Ventral cord motor neuron, innervates dorsal muscles, no ventral counterpart
AVFL	P1.aaaa/ W.aaa	Interneuron
AVFR	P1.aaaa/ W.aaa	Interneuron
AVM	QR.paa	Anterior ventral microtubule cell, touch receptor
DVB	K.p	An excitatory GABAergic motor neuron/interneuron located in dorso-rectal ganglion. Innervates rectal muscles
PDB	P12.apa	Motor neuron, process in dorsal cord, cell body in pre-anal ganglion
PDEL	V5L.paaa	Neuron, dopaminergic of postderid sensillum
PDER	V5R.paaa	Neuron, dopaminergic of postderid sensillum
PHCL	TL.pppaa	Neuron, striated rootlet in male, possibly sensory in tail spike

Table A4 continued

PHCR	TR.pppaa	Neuron, striated rootlet in male, possibly sensory in tail spike
PLNL	TL.pppap	Interneuron, associated with PLM
PLNR	TR.pppap	Interneuron, associated with PLM
PQR	QL.ap	Neuron, basal body, not part of a sensillum, projects into preanal ganglion
PVDL	V5L.paapa	Neuron, lateral process adjacent to excretory canal
PVDR	V5R.paapa	Neuron, lateral process adjacent to excretory canal
PVM	QL.paa	Posterior ventral microtubule cell, touch receptor
PVNL	TL.appp	Interneuron/motor neuron, post. vent. cord, few synapses
PVNR	TR.appp	Interneuron/motor neuron, post. vent. cord, few synapses
PVWL	TL.ppa	Interneuron, posterior ventral cord, few synapses
PVWR	TR.ppa	Interneuron, posterior ventral cord, few synapses
RMFL	G2.al	Ring motor neuron/interneuron
RMFR	G2.ar	Ring motor neuron/interneuron
RMHL	G1.l	Ring motor neuron/interneuron
RMHR	G1.r	Ring motor neuron/interneuron
SDQL	QL.pap	Post. lateral interneuron, process projects into ring
SDQR	QR.pap	Ant. lateral interneuron, process projects into ring
VA1	W.pa	Ventral cord motor neuron, innervates vent. body muscles
VA2	P2.aaaa	Ventral cord motor neuron, innervates vent. body muscles
VA3	P3.aaaa	Ventral cord motor neuron, innervates vent. body muscles
VA4	P4.aaaa	Ventral cord motor neuron, innervates vent. body muscles
VA5	P5.aaaa	Ventral cord motor neuron, innervates vent. body muscles
VA6	P6.aaaa	Ventral cord motor neuron, innervates vent. body muscles
VA7	P7.aaaa	Ventral cord motor neuron, innervates vent. body muscles
VA8	P8.aaaa	Ventral cord motor neuron, innervates vent. body muscles
VA9	P9.aaaa	Ventral cord motor neuron, innervates vent. body muscles
VA10	P10.aaaa	Ventral cord motor neuron, innervates vent. body muscles
VA11	P11.aaaa	Ventral cord motor neuron, innervates vent. body muscles
VA12	P12.aaaa	Ventral cord motor neuron, innervates vent. body muscles, but also interneuron in preanal ganglion

Table A4 continued

VB1	P1.aaap	Ventral cord motor neuron, innervates vent. body muscles, also interneuron in ring
VB2	W.aap	Ventral cord motor neuron, innervates vent. body muscles
VB3	P2.aaap	Ventral cord motor neuron, innervates vent. body muscles
VB4	P3.aaap	Ventral cord motor neuron, innervates vent. body muscles
VB5	P4.aaap	Ventral cord motor neuron, innervates vent. body muscles
VB6	P5.aaap	Ventral cord motor neuron, innervates vent. body muscles
VB7	P6.aaap	Ventral cord motor neuron, innervates vent. body muscles
VB8	P7.aaap	Ventral cord motor neuron, innervates vent. body muscles
VB9	P8.aaap	Ventral cord motor neuron, innervates vent. body muscles
VB10	P9.aaap	Ventral cord motor neuron, innervates vent. body muscles
VB11	P10.aaap	Ventral cord motor neuron, innervates vent. body muscles
VC1	P3.aap	Hermaphrodite specific ventral cord motor neuron innervates vulval muscles and ventral body muscles
VC2	P4.aap	Hermaphrodite specific ventral cord motor neuron innervates vulval muscles and ventral body muscles
VC3	P5.aap	Hermaphrodite specific ventral cord motor neuron innervates vulval muscles and ventral body muscles
VC4	P6.aap	Hermaphrodite specific ventral cord motor neuron innervates vulval muscles and ventral body muscles
VC5	P7.aap	Hermaphrodite specific ventral cord motor neuron innervates vulval muscles and ventral body muscles
VD1	W.pp	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor
VD2	P1.app	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor
VD3	P2.app	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor
VD4	P3.app	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor
VD5	P4.app	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor
VD6	P5.app	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor
VD7	P6.app	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor
VD8	P7.app	Ventralcord motor neuron, innervates vent body muscles, reciprocal inhibitor
VD9	P8.app	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor

Table A4 continued

VD10	P9.app	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor
VD11	P10.app	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor
VD12	P11.app	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor
VD13	P12.app	Ventral cord motor neuron, innervates vent body muscles, reciprocal inhibitor

Table A5: Time of appearance of neurons in the course of development of the *C. elegans* nervous system. The data has been manually transcribed from lineage charts provided in references [66, 67]

Neuron	Birth time (min.)	Neuron	Birth time (min.)
ADAL	420	ALML	420
ADAR	420	ALMR	420
ADEL	420	ALNL	480
ADER	420	ALNR	480
ADFL	280	AQR	1140
ADFR	280	AS01	1440
ADLL	300	AS02	1500
ADLR	300	AS03	1500
AFDL	300	AS04	1500
AFDR	300	AS05	1500
AIAL	300	AS06	1500
AIAR	300	AS07	1620
AIBL	280	AS08	1620
AIBR	280	AS09	1620
AIML	280	AS10	1620
AIMR	300	AS11	1620
AINL	320	ASEL	340
AINR	320	ASER	340
AIYL	300	ASGL	300
AIYR	300	ASGR	300
AIZL	420	ASHL	300
AIZR	420	ASHR	300
ALA	300	ASIL	400

Table A5 continued from previous page

ASIR	400	AVM	1260
ASJL	340	AWAL	300
ASJR	340	AWAR	300
ASKL	380	AWBL	280
ASKR	380	AWBR	280
AUAL	340	AWCL	300
AUAR	340	AWCR	300
AVAL	300	BAGL	300
AVAR	300	BAGR	300
AVBL	300	BDUL	420
AVBR	300	BDUR	420
AVDL	280	CEPDL	380
AVDR	280	CEPDR	380
AVEL	300	CEPVL	400
AVER	300	CEPVR	400
AVFL	1560	DA01	300
AVFR	1500	DA02	320
AVG	300	DA03	300
AVHL	300	DA04	320
AVHR	300	DA05	300
AVJL	300	DA06	300
AVJR	300	DA07	300
AVKL	280	DA08	300
AVKR	280	DA09	300
AVL	280	DB01	280

Table A5 continued from previous page

DB02	300	IL2DL	280
DB03	300	IL2DR	280
DB04	280	IL2L	300
DB05	300	IL2R	300
DB06	300	IL2VL	300
DB07	300	IL2VR	300
DD01	300	LUAL	420
DD02	300	LUAR	420
DD03	320	OLLL	400
DD04	320	OLLR	400
DD05	320	OLQDL	400
DD06	320	OLQDR	400
DVA	300	OLQVL	400
DVB	1560	OLQVR	400
DVC	320	PDA	300
FLPL	420	PDB	1620
FLPR	420	PDEL	1920
HSNL	400	PDER	1920
HSNR	400	PHAL	300
IL1DL	400	PHAR	300
IL1DR	400	PHBL	400
IL1L	400	PHBR	400
IL1R	400	PHCL	1560
IL1VL	400	PHCR	1560
IL1VR	400	PLML	480

Table A5 continued from previous page

PLMR	480	RID	320
PLNL	1560	RIFL	420
PLNR	1560	RIFR	410
PQR	1260	RIGL	300
PVCL	420	RIGR	300
PVCR	420	RIH	280
PVDL	2100	RIML	300
PVDR	2100	RIMR	300
PVM	1260	RIPL	280
PVNL	2040	RIPR	300
PVNR	2040	RIR	300
PVPL	300	RIS	280
PVPR	300	RIVL	300
PVQL	300	RIVR	300
PVQR	300	RMDDL	280
PVR	220	RMDDR	300
PVT	300	RMDL	300
PVWL	1320	RMDR	300
PVWR	1320	RMDVL	300
RIAL	320	RMDVR	300
RIAR	320	RMED	300
RIBL	300	RMEL	300
RIBR	300	RMER	300
RICL	400	RMEV	260
RICR	410	RMFL	2220

Table A5 continued from previous page

RMFR	2220	SMBVR	300
RMGL	300	SMDDL	300
RMGR	300	SMDDR	300
RMHL	1500	SMDVL	300
RMHR	1500	SMDVR	300
SAADL	300	URADL	320
SAADR	300	URADR	320
SAAVL	300	URAVL	300
SAAVR	300	URAVR	300
SABD	300	URBL	300
SABVL	420	URBR	280
SABVR	410	URXL	380
SDQL	1260	URXR	380
SDQR	1260	URYDL	280
SIADL	300	URYDR	280
SIADR	300	URYVL	300
SIAVL	300	URYVR	300
SIAVR	300	VA01	1320
SIBDL	300	VA02	1560
SIBDR	300	VA03	1620
SIBVL	300	VA04	1620
SIBVR	300	VA05	1680
SMBDL	400	VA06	1680
SMBDR	400	VA07	1680
SMBVL	300	VA08	1680

Table A5 continued from previous page

VA09	1680	VD06	1500
VA10	1680	VD07	1500
VA11	1800	VD08	1620
VA12	1800	VD09	1620
VB01	1560	VD10	1620
VB02	1500	VD11	1620
VB03	1560	VD12	1620
VB04	1620	VD13	1620
VB05	1620		
VB06	1680		
VB07	1680		
VB08	1680		
VB09	1680		
VB10	1680		
VB11	1680		
VC01	1500		
VC02	1500		
VC03	1500		
VC04	1500		
VC05	1620		
VD01	1320		
VD02	1440		
VD03	1500		
VD04	1500		
VD05	1500		

Table A6: Classification of *C. elegans* neurons according to their role in the mesoscopic organization.

The classification of neurons in the C. elegans somatic nervous system, according to their role in intra- and inter-modular connectivity

R1	**R2**	**R3**	**R5**	**R6**
Ultraperipheral	**Peripheral**	**Nonhub connector**	**Provincial**	**Connector**
nodes	**nodes**	**nodes**	**hubs**	**hubs**
ALNL	ADEL/R	ADAL	AIAL	ADFR
AS06	ADFL	ADAR	AIAR	AIZL
DD03	ADLL	ADLR	ASEL	ASHR
IL1L	AFDL/R	AIBR	ASER	AVAL
IL1R	AIBL	AIML	AWCL	AVAR
OLLL	AINL/R	AIMR	AWCR	AVBL
OLQDL	AIYL/R	ALA	CEPDL	AVBR
PDA	AIZR	AQR	DD02	AVDR
PDB	ALML/R	AVDL	DD06	AVER
RMEL	ALNR	AVEL	OLLR	AVG
RMEL	AS01	AVFR	PVNL	AVHL
VA07	AS02	BDUL	RMDDL	AVHR
VB07	AS03	DVC	RMDL	AVJL
	AS04	FLPL	RMDR	AVKR
	AS05	FLPR	RMDVL	DA02
	AS07	HSNR	RMDVR	DD01
	AS08	PVQR	RMEV	DVA
	AS09	RID	SMDDL	HSNL
	AS10	RIFL	SMDDR	PVCL
	AS11	RIFR	SMDVL	PVCR
	ASGL/R	SABD	VD12	PVNR

Table A6 continued from previous page

	ASHL	SABVL		PVPL
	ASIL/R	SABVR		PVPR
	ASJL/R	URXL		PVR
	ASKL/R	VB01		PVT
	AUAL/R			RIAL
	AVFL			RIAR
	AVJR			RIBL
	AVKL			RIBR
	AVL			RIH
	AVM			RIML
	AWAL/R			RIMR
	AWBL/R			RIR
	BAGL/R			RIS
	BDUR			RMGL
	CEPDR			SMDVR
	CEPVL/R			
	DA01			
	DA03			
	DA04			
	DA05			
	DA06			
	DA07			
	DA08			
	DA09			
	DB01			
	DB02			
	DB03			

Table A6 continued from previous page

	DB04			
	DB05			
	DB06			
	DB07			
	DD04			
	DD05			
	DVB			
	IL1DL/R			
	IL1VL/R			
	IL2DL/R			
	IL2L/R			
	IL2VL/R			
	LUAL/R			
	OLQDR			
	OLQVL/R			
	PDEL/R			
	PHAL/R			
	PHBL/R			
	PHCL/R			
	PLML/R			
	PLNL/R			
	PQR			
	PVDL/R			
	PVM			
	PVQL			
	PVWL/R			
	RICL/R			

Table A6 continued from previous page

	RIGL/R			
	RIPL/R			
	RIVL/R			
	RMDDR			
	RMED			
	RMFL/R			
	RMGR			
	RMHL/R			
	SAADL/R			
	SAAVL/R			
	SDQL/R			
	SIADL/R			
	SIAVL/R			
	SIBDL/R			
	SIBVL/R			
	SMBDL/R			
	SMBVL/R			
	URADL/R			
	URAVL/R			
	URBL/R			
	URXR			
	URYDL/R			
	URYVL/R			
	VA01			
	VA02			
	VA03			
	VA04			

Table A6 continued from previous page

	VA05			
	VA06			
	VA08			
	VA09			
	VA10			
	VA11			
	VA12			
	VB02			
	VB03			
	VB04			
	VB05			
	VB06			
	VB08			
	VB09			
	VB10			
	VB11			
	VC01			
	VC02			
	VC03			
	VC04			
	VC05			
	VD01			
	VD02			
	VD03			
	VD04			
	VD05			
	VD06			

Table A6 continued from previous page

	VD07			
	VD08			
	VD09			
	VD10			
	VD11			
	VD13			

Table A7: The neuronal composition of different functional circuits in the *C. elegans* somatic nervous system.

F1	**F2**	**F3**	**F4**	**F5**
Mechanosensation	**Egg laying**	**Thermotaxis**	**Chemosensation**	**Feeding**
CEPVL/R	ASKL/R	AFDL/R	ADFL/R	ADLL/R
CEPDL/R	AVJL/R	ASIL/R	ADLL/R	AIBL/R
ASHL/R	AWCL/R	AWCL/R	AIBL/R	AIZL/R
AVAL/R	AIAL/R	FLPL/R	AIZL/R	ASKL/R
AVBL/R	AVFL/R	PVDL/R	ASEL/R	AWAL/R
AVDL/R	HSNL/R	PHCL/R	ASGL/R	AWCL/R
AVEL/R	VC1	AIAL/R	ASHL/R	RIBL/R
ADEL/R	VC2	AIYL/R	ASIL/R	
ALML/R	VC3	AIZL/R	ASJL/R	
AVM	VC4	AIBL/R	ASKL/R	
PDEL/R	VC5	RIAL/R	AVAL/R	
PVDL/R	VC6	RIBL/R	AVBL/R	
PVM		RIML/R	AVEL/R	
PLML/R			AWAL/R	
PVCL/R			AWBL/R	

Table A7 continued from previous page

			AWCL/R	
			RIBL/R	
			AIAL/R	
			AIYL/R	

F6	**F7**	**F8**	**F9**	
Exploration	**Tap withdrawal**	**O2 sensation**	**CO2 sensation**	
ASIL/R	AVAL/R	AQR	ASEL/R	
RIML/R	AVBL/R	ASHL/R	PQR	
AIYL/R	AVDL/R	URXL/R	AFDL/R	
	ALML/R	ALNL/R	BAGL/R	
	AVM	PQR	URXL/R	
	PVDL/R	PLNL/R	AQR	
	DVA	ADLL/R	AIYL/R	
	PLML/R	RMGL/R		
	PVCL/R	SDQL/R		
		BAGL/R		

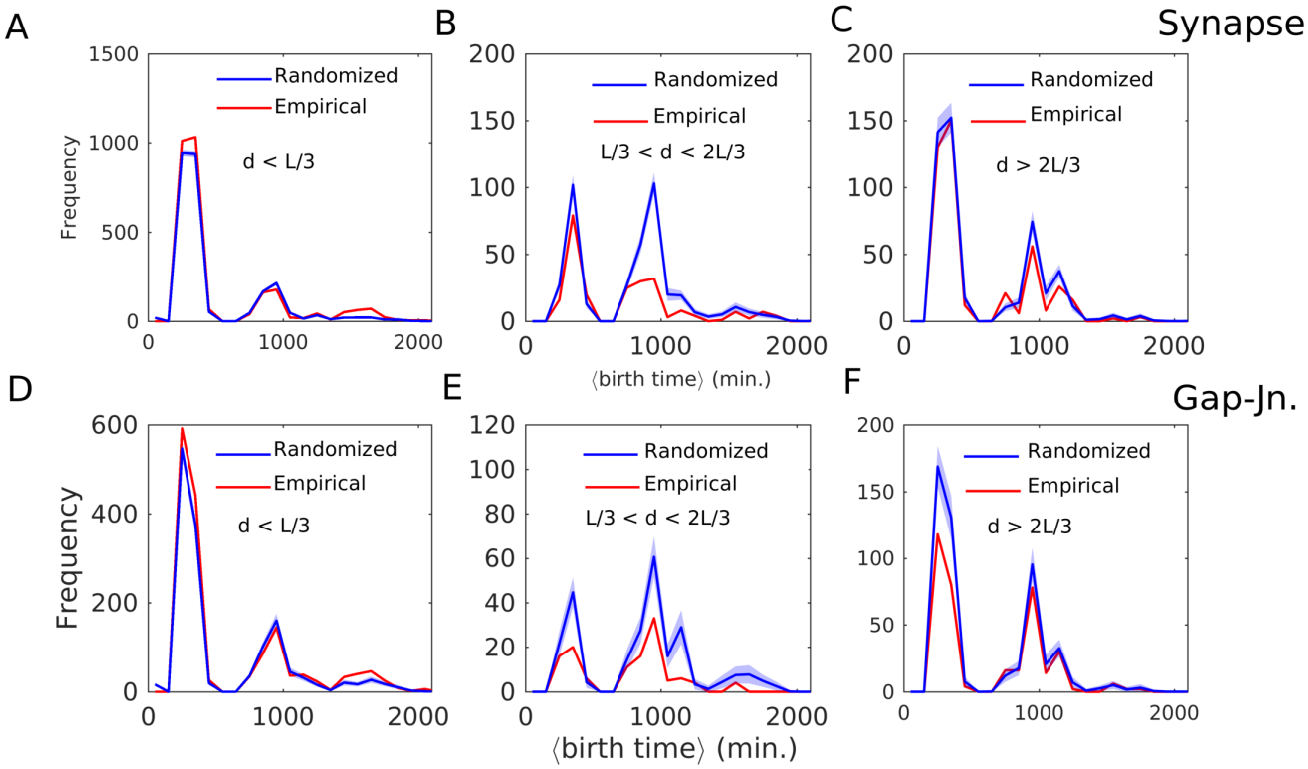


Figure A1: Birth cohort homophily is seen specifically for connections between neurons whose cell bodies are in close physical proximity. Frequency distributions of the mean birth time for all pairs that are connected via synapses (A-C) or gap-junctions (D-F). The distributions for the empirical network (shown in red) are compared with distributions obtained from surrogate ensembles of randomized networks (blue curve shows the average over 100 realizations, the dispersion being indicated by the shaded area). The latter are constructed from the empirical network by randomly rewiring the connections while keeping the total number of connections (degree) for each neuron, the spatial location of its cell body and its process length unchanged. In addition, to allow only physically possible connections between neurons, we have imposed process-length constraint which disallow linking two cells if the distance between their cell bodies is greater than the sum of their individual process lengths. The different panels correspond to connections between neurons whose cell bodies are separated by distance d which is short ($d < L/3$: A, D), medium ($L/3 < d < 2L/3$: B, E) or long ($d > 2L/3$: C, F) relative to the total body length of the worm L. As in Fig. 2.4 in Chapter 2, the trimodal nature of these distributions arise from three classes of connected neuronal pairs, viz., (i) where both cells are born early (i.e., in the embryonic stage), (ii) where one is born early while the other late (i.e., in the post-embryonic stage) and (iii) where both are born late. Birth cohort homophily is indicated when the peaks of the empirical frequency distribution, corresponding to connections between neurons that are either both born early or both born late, have significantly higher values than the randomized distribution (the latter corresponding to a null model where connections between cells can occur independent of the time of their birth). This is seen only in panels (A) and (D), i.e., for connections between neurons whose cell bodies are located relatively close to each other.

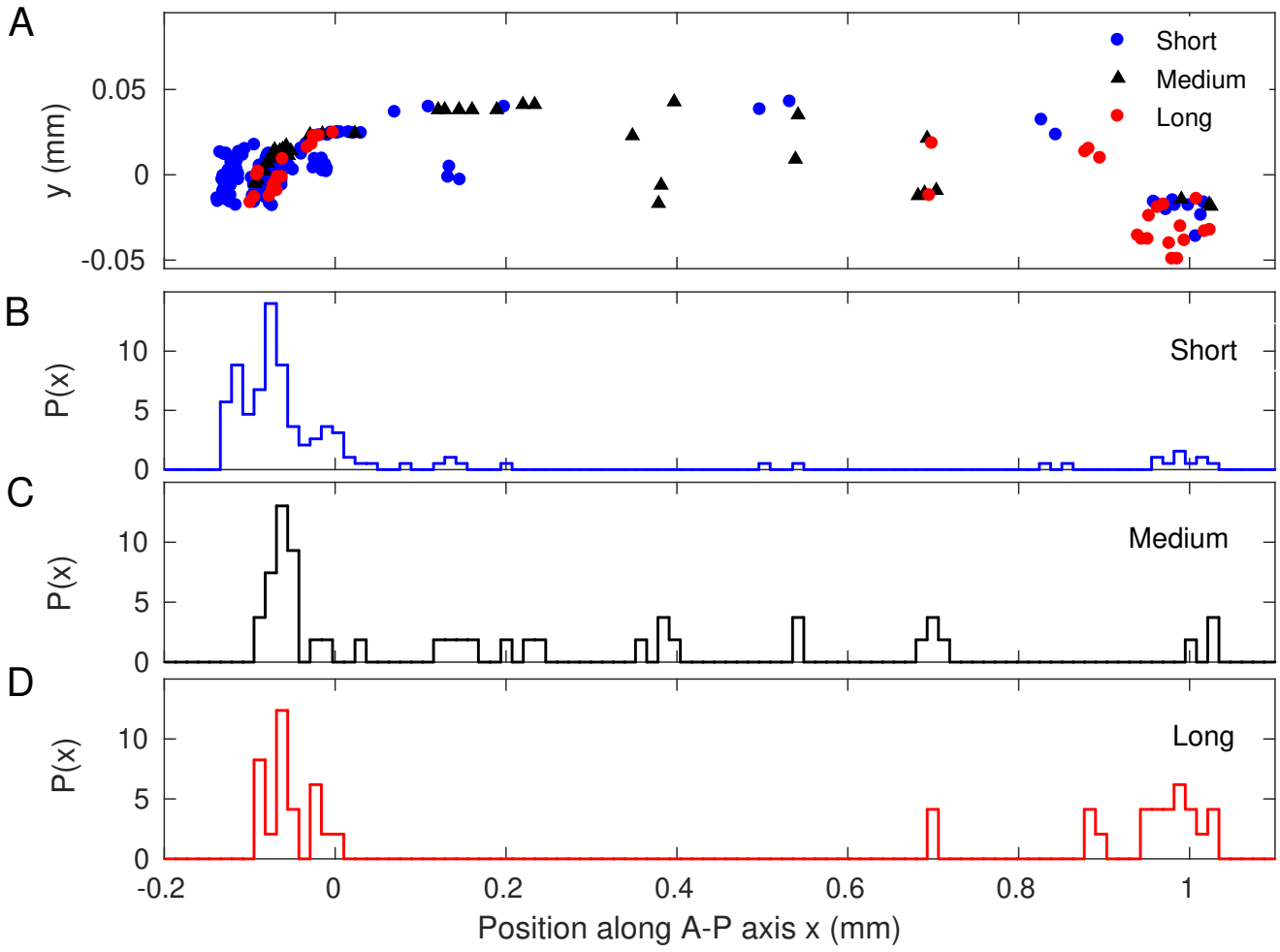


Figure A2: Spatial distribution of cell bodies of the neurons belonging to the somatic nervous system of *Caenorhabditis elegans*. (A) Projection of the physical locations of the neuronal cell bodies on the two-dimensional plane formed by the anterior-posterior (AP) axis (x, along the horizontal) and the ventral-dorsal axis (y, along the vertical). Cells having short, medium and long processes are indicated using different symbols. The animal is oriented such that its head is located near the left end and tail near the right end of the plane. (B-D) Probability distributions of the location of the cell bodies along the AP axis (x, measured in mm) for neurons having (B) short, (C) medium and (D) long processes. We note that the distributions for neurons having short and long processes, both have an approximately bimodal nature. It suggests that most cells of these two types are localized near either the head or the tail regions, while neurons with medium length processes are distributed across the body of the worm in a relatively more homogeneous manner.

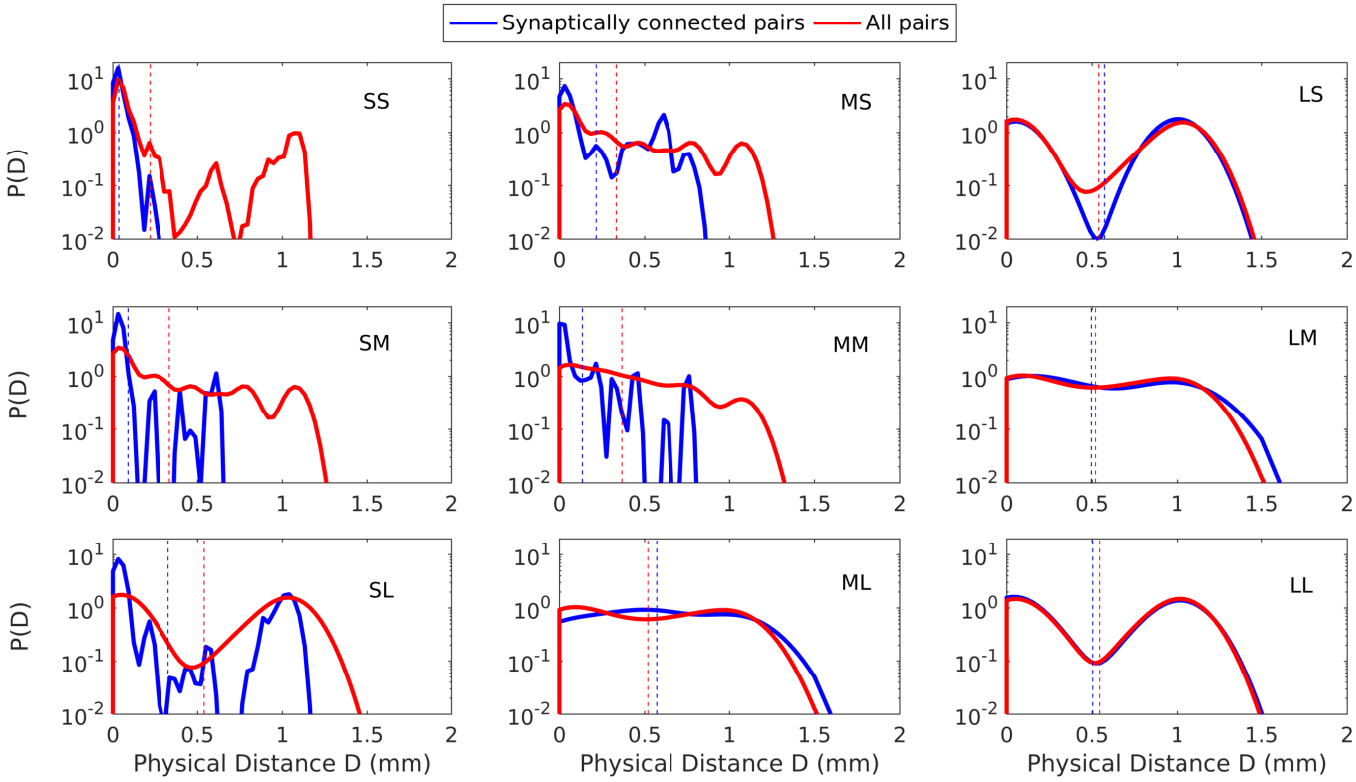


Figure A3: Distribution of distances between cell bodies of synaptically connected pairs of neurons differs from that of all pairs. Comparison of the distributions of distances between synaptically connected pairs (blue) and all pairs (red) of neurons distinguished in terms of their respective process lengths (S: short, M: medium, L: long). The mean value for each of the distributions (distinguished by their color) is marked by broken lines. Note that for all pairs of process length categories (with the exception of LS and ML), the average distance for synaptically connected pairs is less than the average calculated over all pairs of neurons, consistent with the results shown in Table A3.

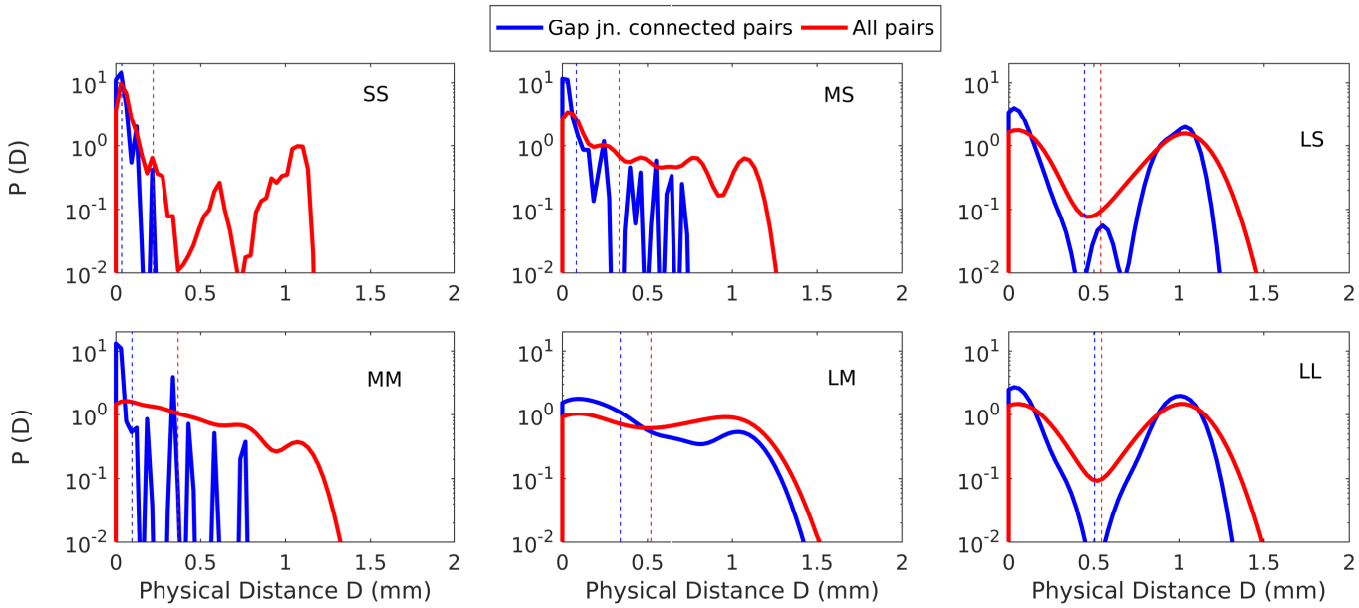


Figure A4: Distribution of distances between cell bodies of pairs of neurons connected by gap-junctions differs from that of all pairs. Comparison of the distributions of distances between pairs connected by gap-junctions (blue) and all pairs (red) of neurons distinguished in terms of their respective process lengths (S: short, M: medium, L: long). The mean value for each of the distributions (distinguished by their color) is marked by broken lines. Note that for all pairs of process length categories, the average distance for pairs connected by gap-junctions is less than the average calculated over all pairs of neurons, consistent with the results shown in Table A3.

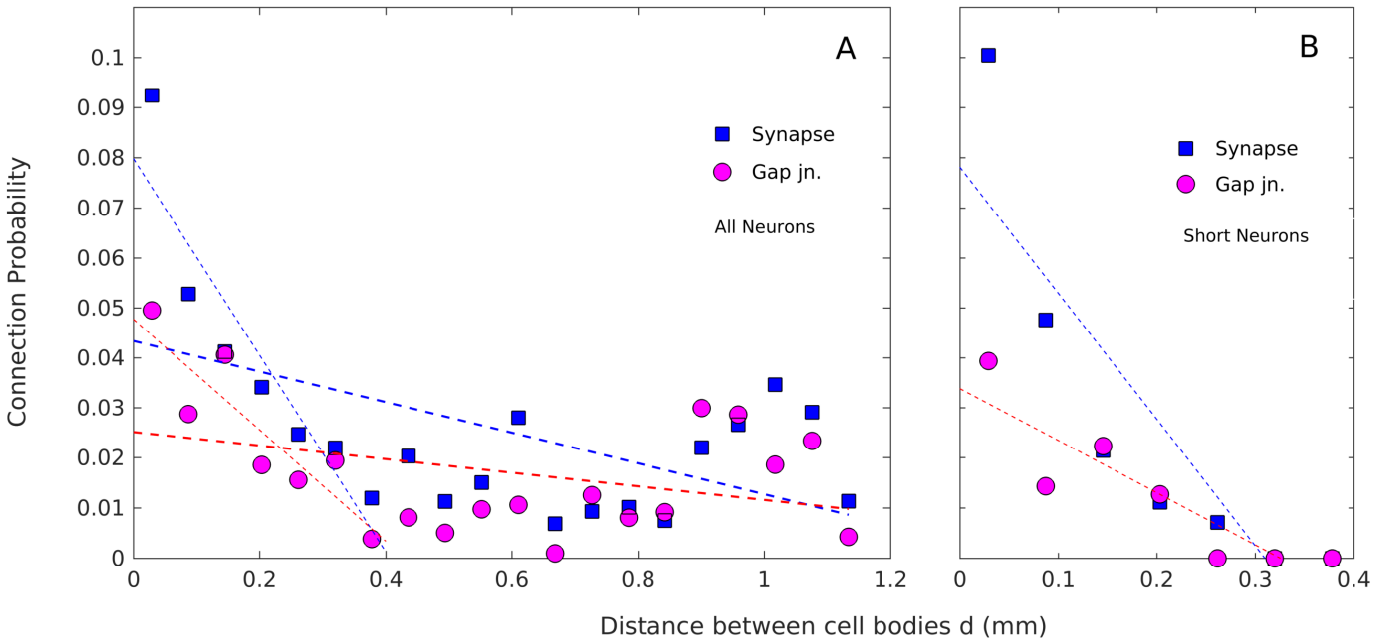


Figure A5: Dependence of the probability of connection between two neurons on the physical distance between their cell bodies. (A) The variation of the probability of connection between two neurons by either synapse (squares) or gap-junction (circles) as a function of the physical distance between their cell bodies d (measured in mm). Linear fitting of the functions show a decay with d overall (thick broken lines), but the relation is much weaker compared to that seen between probability of synaptic connection between two cells and their lineage distance l [see Fig. 2.2 (D) in Chapter 2]. In particular, the correlation is diluted by the relatively high probability for synapses to form between neurons whose cell bodies are located at the opposite ends of the worm (corresponding to the peak around $x = 1$ mm). However, when we focus only on connections between cell bodies that are in close physical proximity ($d < 0.4$ mm), the dependence on d appears to be much more prominent (thin broken lines). This stronger correlation between connection probability and d at short distances is not necessarily an outcome of constraints imposed by the process lengths of the neurons. This is suggested by panel (B), where we focus exclusively on neurons with short processes. (B) The relation between connection probability between neurons, both of which have short processes, and the distance between their cell bodies, d, is seen to be not more prominent than that already seen for all neurons [in panel (A)].

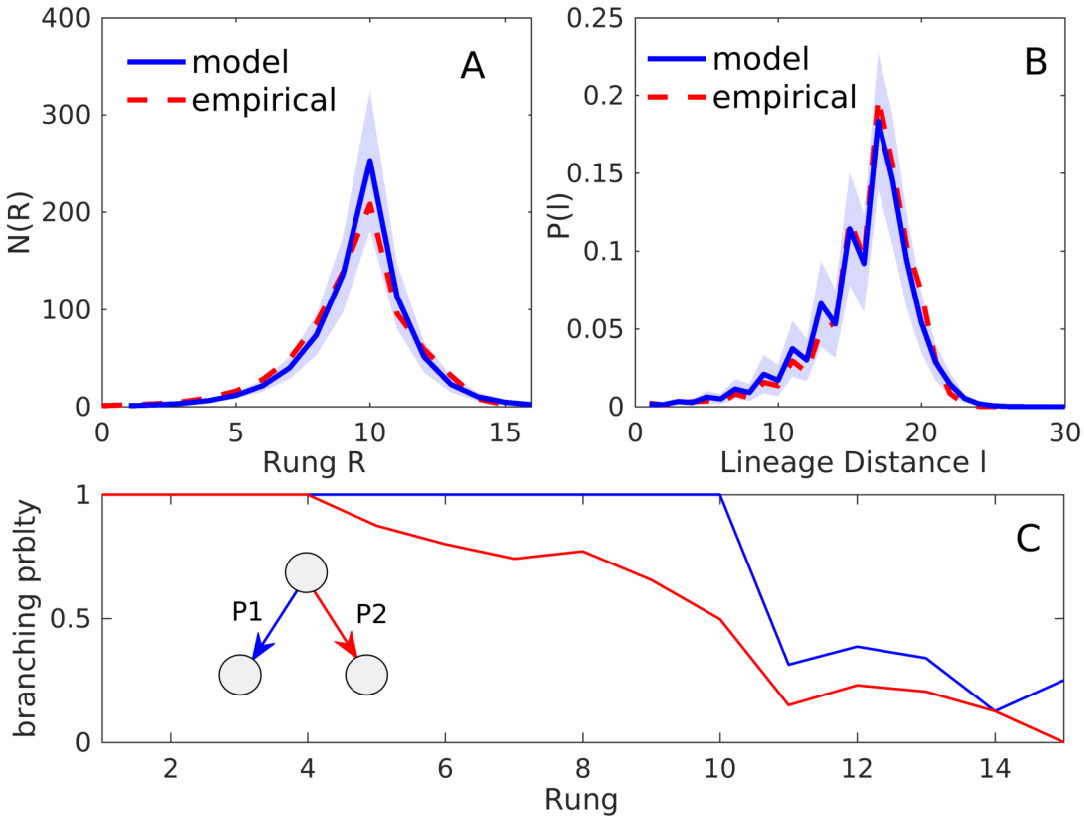


Figure A6: A stochastic branching model for the lineage tree of cells involved in the development of the *C. elegans* somatic nervous system. (A) Comparison of the distribution of rung R occupied by each cell (progenitor cells of the neurons, as well as, differentiated neurons) in the lineage tree obtained empirically (broken curve) with that generated by the model (solid curve shows the mean computed over an ensemble of 10^3 realizations, the dispersion being indicated by the shaded area). (B) Comparison of the distribution of lineage distance l between pairs of differentiated neurons of *C. elegans* (broken curve) with that obtained from the model (solid curve showing the mean computed over an ensemble of 10^3 realizations, the dispersion being indicated by the shaded area). The high degree of overlap between the empirical and simulated distributions indicates that the stochastic branching model is a reasonably accurate description of the lineage tree of neurons. (C) The branching probabilities $P1$ (blue curve) and $P2$ (red curve) of a progenitor cell at each rung, estimated from the empirical lineage tree (by definition, $P1 \geq P2$). Note that both of the branching probabilities show a prominent dip after rung 10. Guided by this, in the stochastic branching model, $P1, P2$ have been chosen to have a constant high value upto rung 10 (viz., $P1 = 1$, $P2 = 0.85$), after which both are decreased to a constant low value (viz., $P1 = 0.25$, $P2 = 0.2$). The inset shows a schematic of the stochastic branching model where a node, occurring at any rung, can branch (or not) based on the probabilities $P1$ and $P2$ which will result in any one of the following three possibilities: (i) proliferation occurs along both branches, (ii) only one branch appears (the other branch leading to either apoptosis or a non-neural cell fate), and, (iii) there is no branching so that we obtain a terminal node of the tree (i.e., the cell differentiates into a neuron).

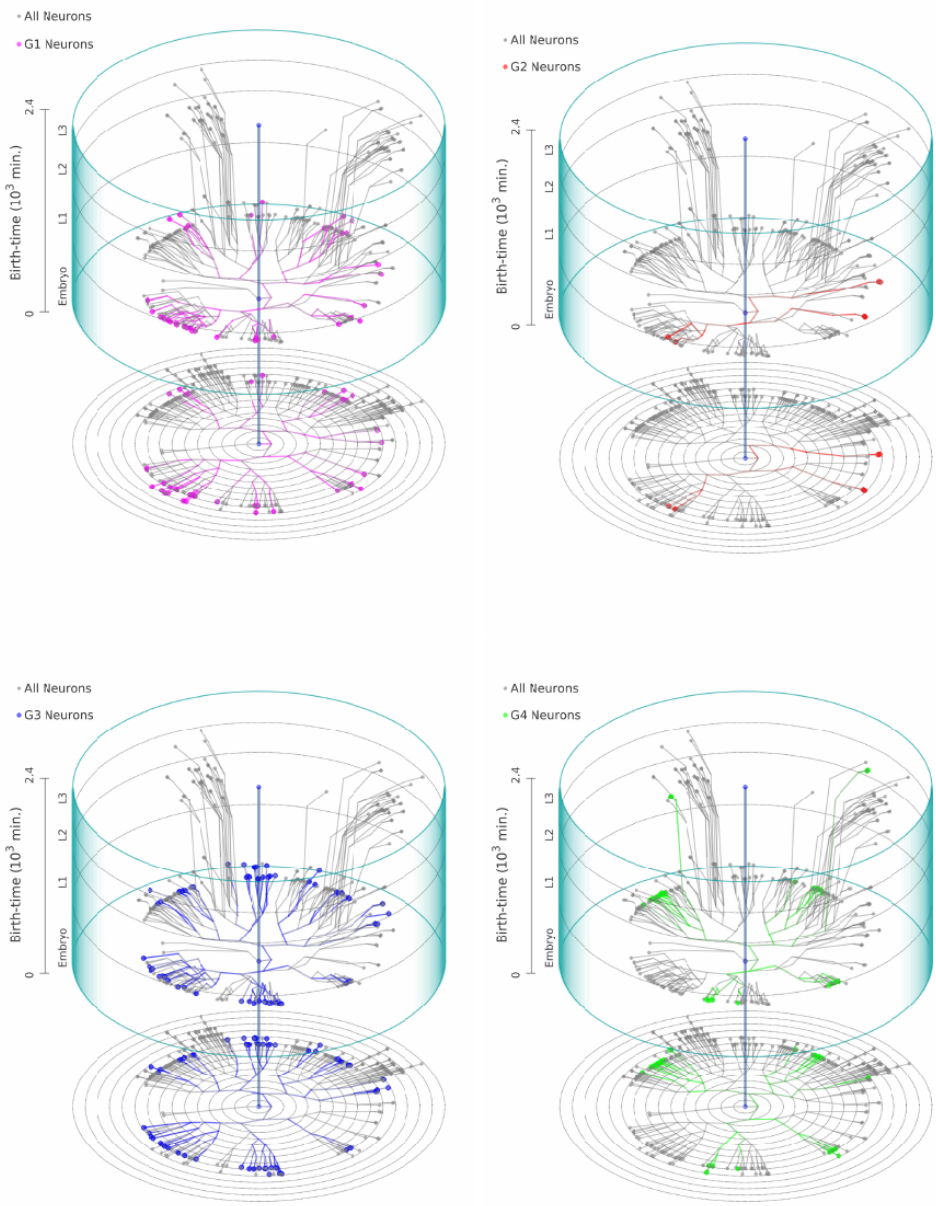


Figure A7: Developmental chrono-dendrograms for the Anterior (G1, top left), Dorsal (G2, top right), Lateral (G3, bottom left) and Ventral (G4, bottom right) ganglia, showing that each comprises multiple localized clusters of neurons. Colored nodes represent neurons belonging to the specified ganglion while gray nodes show other neurons. Branching lines trace all cell divisions starting from the single cell zygote (located at the origin) and terminating at each differentiated neuron. The time and rung of each cell division is indicated by its position along the vertical and radial axis respectively. The entire time period is divided into four stages, viz., Embryo (indicated as E), L1, L2 and L3. A planar projection at the base of each cylinder shows the rung (concentric circles) of each progenitor cell and differentiated neuron.

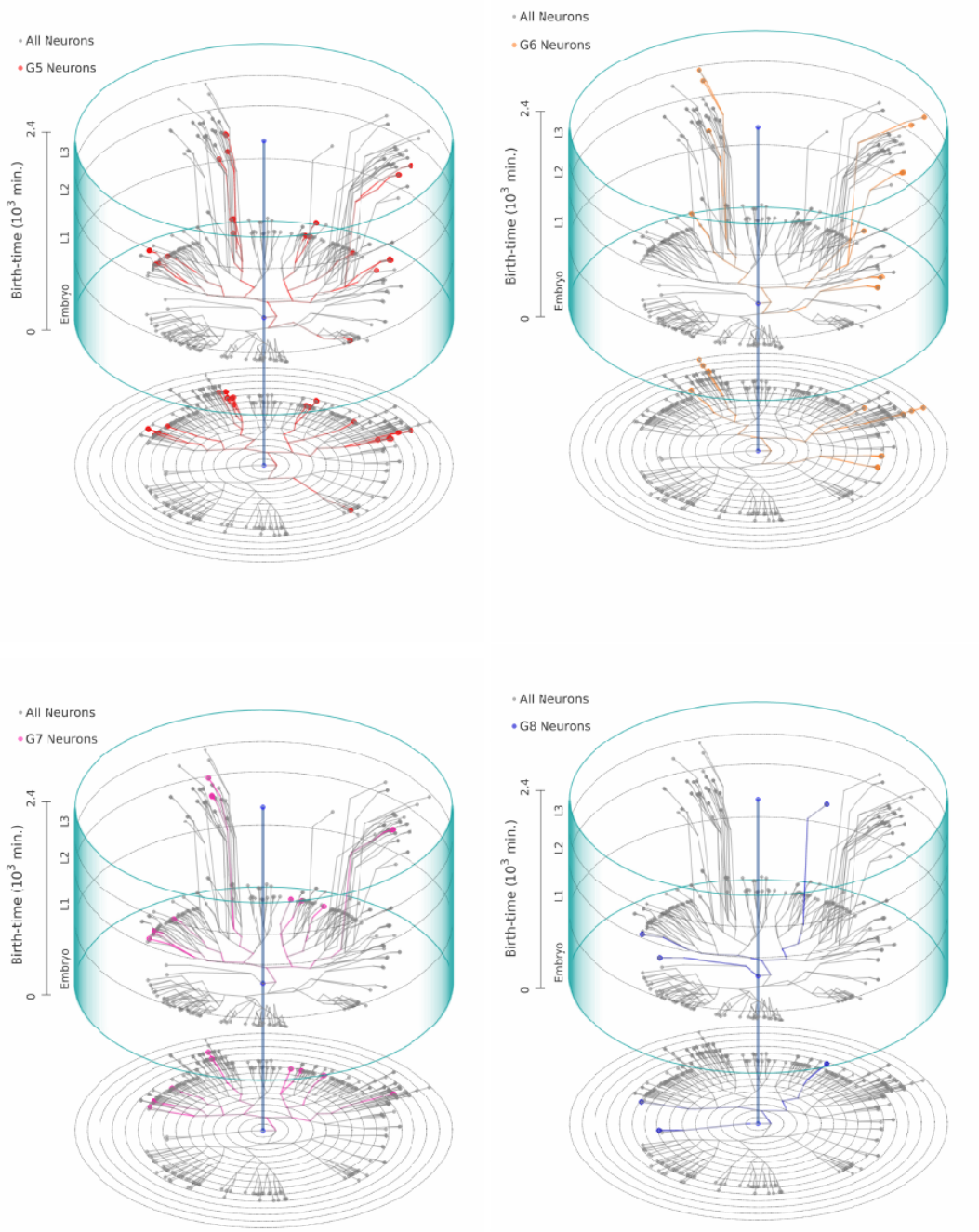


Figure A8: Developmental chrono-dendrograms for the Retrovesicular (G5, top left), Posterolateral (G6, top right), Preanal (G7, bottom left) and Dorsorectal (G8, bottom right) ganglia, showing that each comprises multiple localized clusters of neurons. Colored nodes represent neurons belonging to the specified ganglion while gray nodes show other neurons. Branching lines trace all cell divisions starting from the single cell zygote (located at the origin) and terminating at each differentiated neuron.

Figure A8 ***(previous page)*****:** The time and rung of each cell division is indicated by its position along the vertical and radial axis respectively. The entire time period is divided into four stages, viz., Embryo (indicated as E), L1, L2 and L3. A planar projection at the base of each cylinder shows the rung (concentric circles) of each progenitor cell and differentiated neuron.

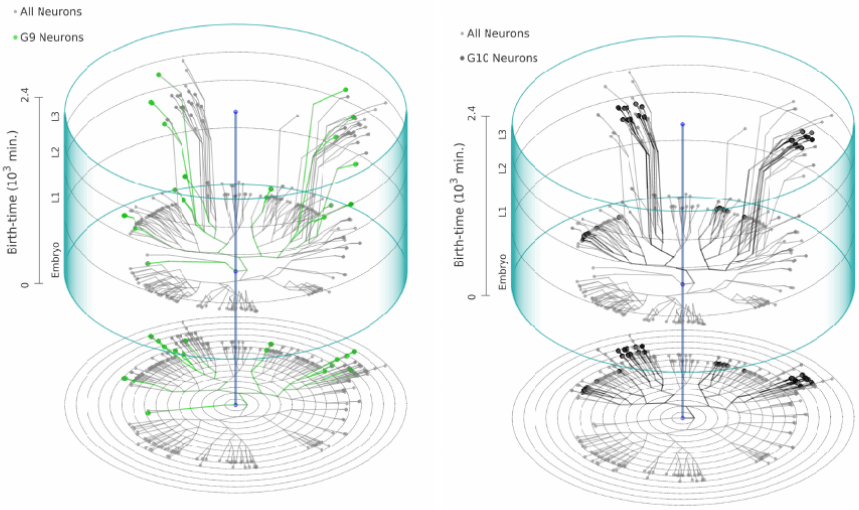


Figure A9: Developmental chrono-dendrograms for the Lumbar ganglion (G9, left) and the Ventral Cord (G10, right), showing that each comprises multiple localized clusters of neurons. Colored nodes represent neurons belonging to the specified ganglion while gray nodes show other neurons. Branching lines trace all cell divisions starting from the single cell zygote (located at the origin) and terminating at each differentiated neuron. The time and rung of each cell division is indicated by its position along the vertical and radial axis respectively. The entire time period is divided into four stages, viz., Embryo (indicated as E), L1, L2 and L3. A planar projection at the base of each cylinder shows the rung (concentric circles) of each progenitor cell and differentiated neuron.

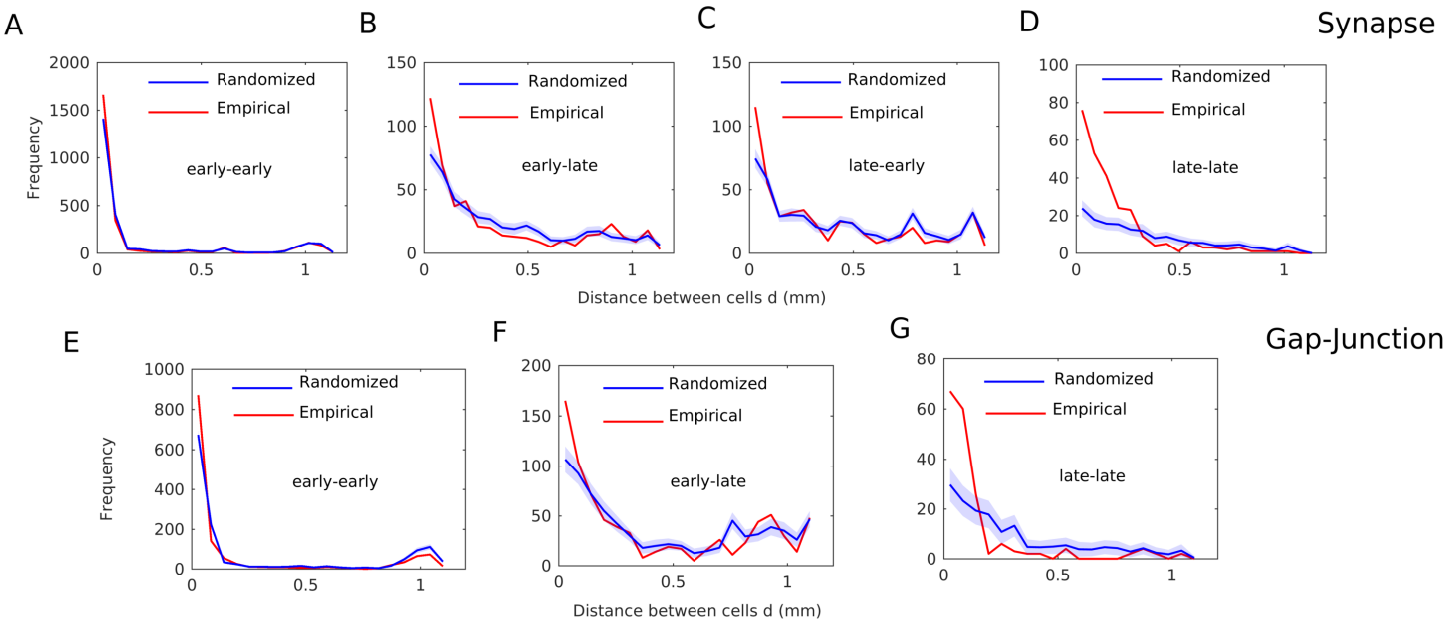


Figure A10: Connections between neurons born at different developmental epochs are over-represented when the cell bodies are far apart, suggesting the presence of active processes facilitating such links. Frequency distributions of the distance d between cell bodies of all neuronal pairs that are connected via synapses (A-D) or gap-junctions (E-G). The distributions for the empirical network (shown in red) are compared with distributions obtained from surrogate ensembles of randomized networks (blue curve shows the average over 100 realizations, the dispersion being indicated by the shaded area). The latter are constructed from the empirical network by randomly rewiring the connections while keeping the total number of connections (degree) for each neuron, the spatial location of its cell body and its process length unchanged. In addition, to allow only physically possible connections between neurons, we have imposed process-length constraint which disallow linking two cells if the distance between their cell bodies is greater than the sum of their individual process lengths. The different panels correspond to the situations where (A,E) both cells in a connected pair are born in the early developmental burst, (B,C,F) one is born early and the other is born late [in (B) it is the pre-synaptic neuron which is born early, while in (C) the post-synaptic neurons appears in the early developmental burst], and (D,G) both cells are born late. When two neurons are born in the same developmental epoch (either early or late), the empirical frequency distribution is seen to have significantly higher values than the randomized distribution at low d (seen in panels A and D, and even more prominently in panels E and G), indicating that neurons prefer to connect to other members of their birth cohort whose cell bodies are in close proximity. This is particularly evident for neurons born in the late developmental epoch. Note that this result complements the earlier observation that birth cohort homophily is seen specifically for neurons whose cell bodies are located relatively close to each other (Fig. A1).

Figure A10 ***(previous page)*****:** More intriguingly, connections between neurons whose cell bodies lie far apart are seen to occur more frequently than expected by chance when the pre-synaptic neuron is born early and the post-synaptic neuron is born late (see panel B). A similar phenomenon is also seen in the case of early- and late-born neurons connected by gap junctions (see panel F). These results suggest the presence of an active process forming connections between neurons born in different epochs.

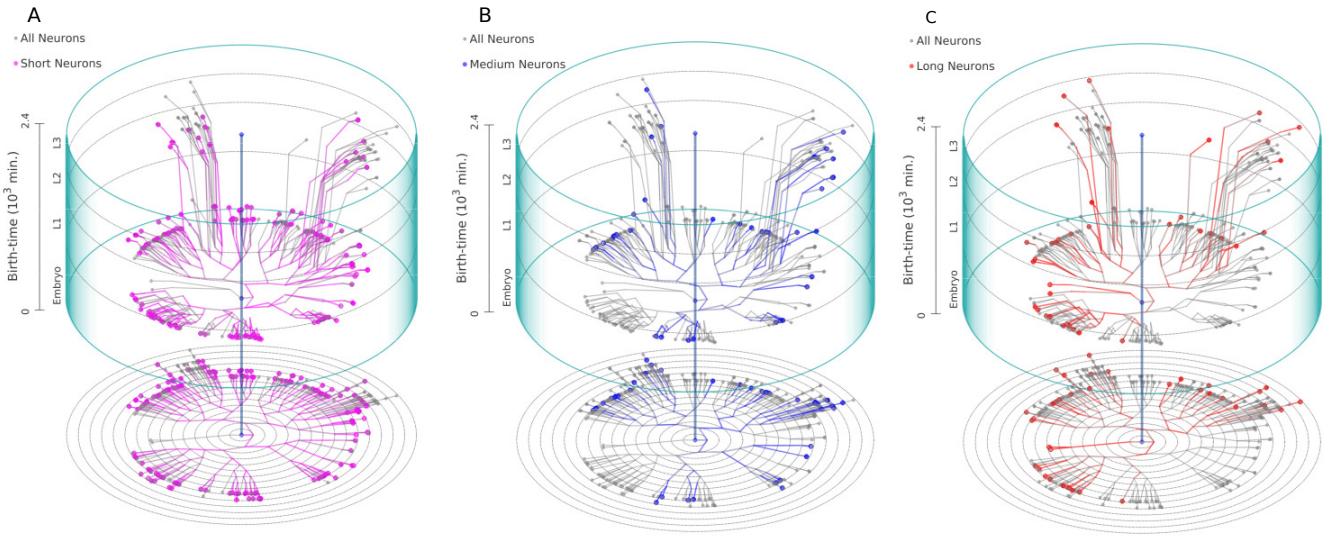


Figure A11: Absence of segregated clusters in the developmental chrono-dendrograms for neurons having similar process lengths (viz., short, medium and long) suggest that process length is not exclusively determined by lineage. Colored nodes represent neurons having a specified process length, viz., short in (A), medium in (B) and long in (C), while grey nodes show other neurons. Branching lines trace all cell divisions starting from the single cell zygote (located at the origin) and terminating at each differentiated neuron. The time and rung of each cell division is indicated by its position along the vertical and radial axis respectively. The entire time period is divided into four stages, viz., Embryo (indicated as E), L1, L2 and L3. A planar projection at the base of each cylinder shows the rung (concentric circles) of each progenitor cell and differentiated neuron.

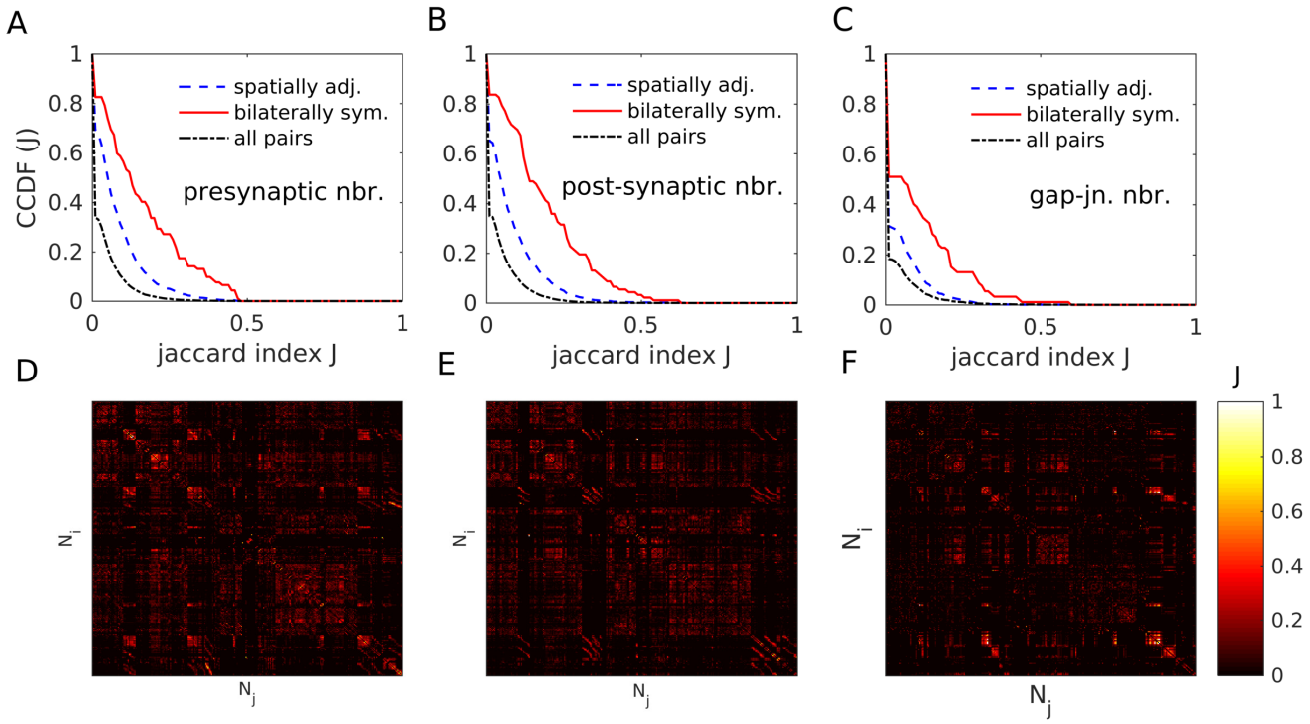


Figure A12: Physical proximity alone cannot explain the high degree of overlap between the cells that each member of a bilaterally symmetric pair of neurons connect to. Complementary cumulative probability distributions (CCDF) of the overlap between the sets $\mathcal{N}_G(i)$, $\mathcal{N}_G(j)$ of the neighbors (defined for a network G) of neurons N_i and N_j, where the indices i and j can run over (i) all pairs of neurons (dash-dotted curves), (ii) only bilaterally symmetric pairs (solid curves) or (iii) pairs whose cell bodies are spatially adjacent to each other ($d < 0.05$ mm, broken curves). The overlap is measured in terms of the Jaccard index J, defined for the pair i, j as $J(i,j) = [\mathcal{N}_G(i) \cap \mathcal{N}_G(j)]/[\mathcal{N}_G(i) \cup \mathcal{N}_G(j)]$, where $\cap$ and $\cup$ refers to intersection and union of two sets, respectively. The different panels correspond to different networks G used to define neighbors for a neuron, viz., (A) pre-synaptic neighbors, i.e., cells from which the neuron receives a synaptic connection, (B) post-synaptic neighbors, i.e., cells to which the neuron sends a synaptic connection, and (C) gap-junctional neighbors, i.e., cells to which a neuron is coupled via a gap junction. We note that the overlaps between the neighborhoods (for all three types of network neighbors considered here) of bilaterally symmetric neurons are consistently higher than that of all pairs of neurons, as well as, of pairs whose cell bodies are in close physical proximity. Thus, bilaterally symmetric neurons share neighbors to a much greater extent than that expected by their cell bodies being located close to each other. (D-F) The Jaccard index matrices J showing overlaps between the neighbors for every pair of neurons N_i, N_j when the network neighborhood defined is that of (D) pre-synaptic partners, (E) post-synaptic partners and (F) gap-junctional partners. The large overlap between neighbors of bilaterally symmetric neurons is indicated by the occurrence of bands of brightly colored entries along the diagonal (note that bilaterally symmetric neurons are always located on adjacent rows/columns).

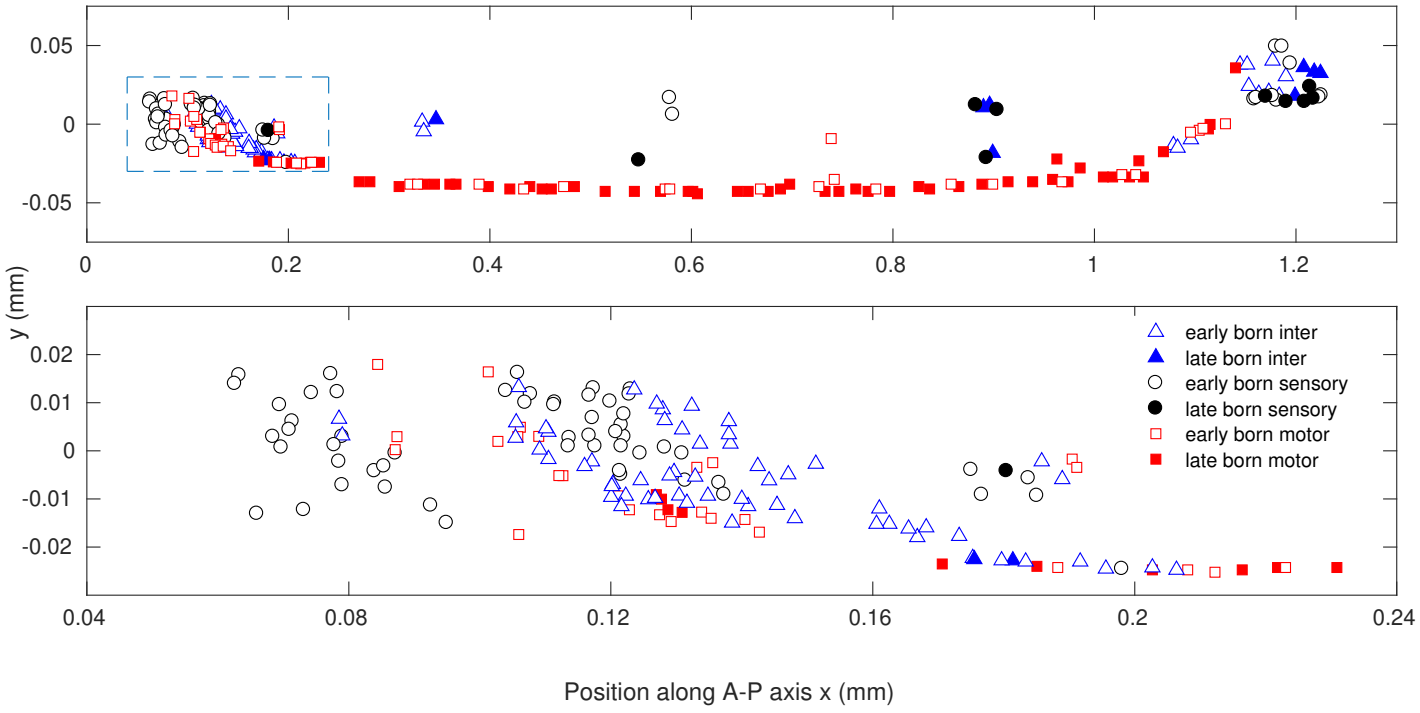


Figure A13: Spatial distribution of the cell bodies of sensory, inter and motor neurons of the somatic nervous system of *Caenorhabditis elegans*. Projections of the physical locations of the neuronal cell bodies, distinguished according to functional type (sensory: circles, inter: triangles and motor: squares) and whether they appear in the early (unfilled symbols) or late (filled symbols) developmental epochs, on the two-dimensional plane formed by the anterior-posterior (AP) axis (x, along the horizontal) and the dorsal-ventral axis (y, along the vertical). Top panel shows the entire worm, with its body oriented such that the head is located near the left end and tail near the right end of the plane. The bottom panel shows a magnified view of the region near the head (bounded by broken lines in the top panel). We note that almost all cells in this region appear at the embryonic stage, during the early burst of development. By contrast, the ventral cord predominantly comprises neurons that appear in the post-embryonic stage (see top panel).

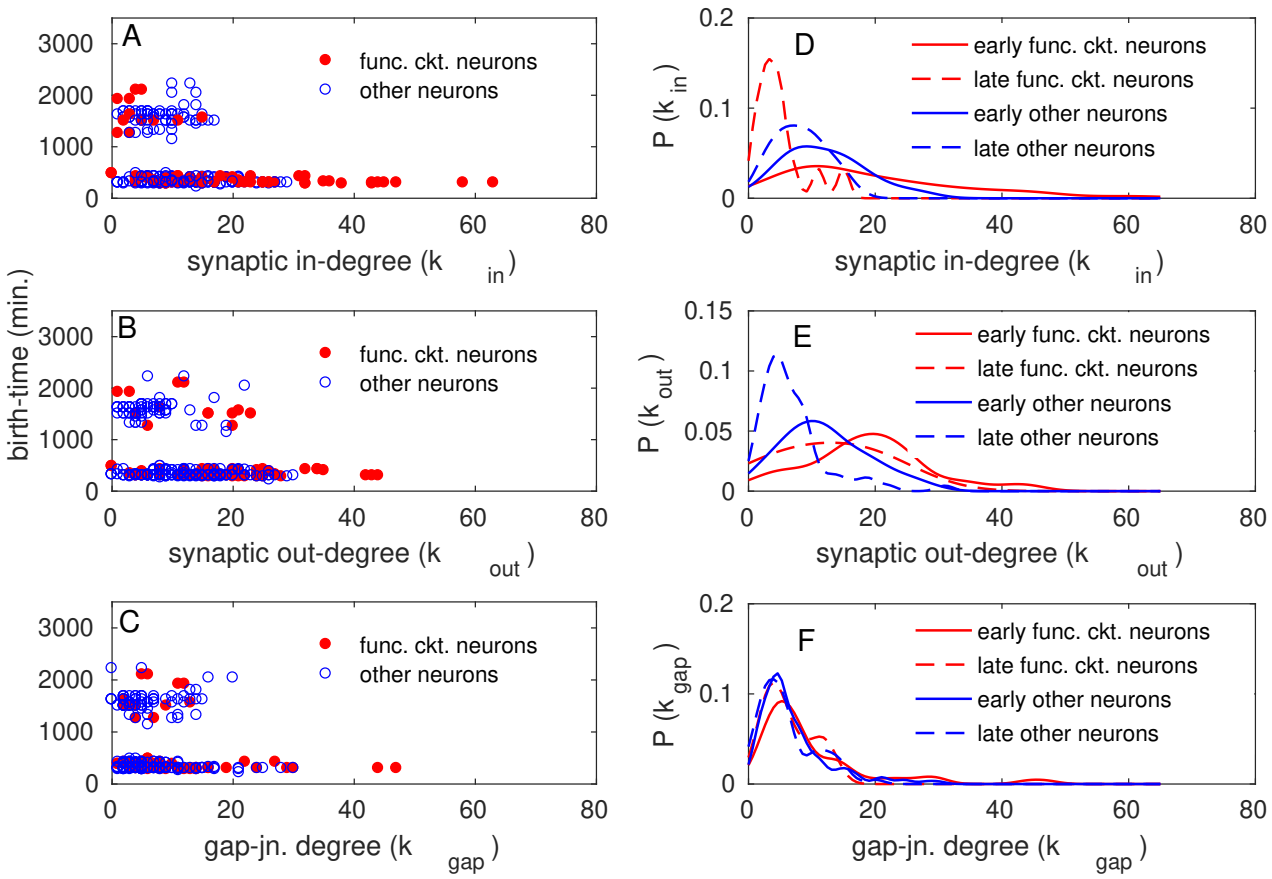


Figure A14: The number of synaptic connections of a neuron is influenced by its functional criticality, as well as, the developmental epoch in which it appeared. (A-C) Scatter plots indicating the relation between the time of appearance of a neuron and its number of (A) incoming synaptic connections from other cells (synaptic in-degree), (B) outgoing synaptic connections to other cells (synaptic out-degree) and (C) gap junctions with other cells (gap-junctional degree). Filled circles represent neurons belonging to any of nine previously identified functional circuits (see Fig. 2.8 in Chapter 2) while unfilled circles show other neurons. (D-F) Probability distributions of different types of connections for neurons categorized in terms of those which are functionally critical, i.e., belong to a functional circuit (red), or not (blue), and whether they appear in the early (solid curve) or late (broken curve) developmental epochs. The different panels correspond to (D) synaptic in-degree, (E) synaptic out-degree and (F) gap-junctional degree. Synaptic in-degree for functionally critical, early-born neurons is seen to have a heavy-tailed distribution which is significantly different from that of the other types of neurons (in terms of a 2-sample Kolmogorov-Smirnov test at 5% level of significance). For the case of synaptic out-degree, however, the distributions for functionally critical neurons that are born at different epochs are statistically indistinguishable. However, for other neurons, the distribution of those that are born in the later, post-embryonic developmental burst are distinct from those that are born early (demonstrated by a 2-sample Kolmogorov-Smirnov test at 5% level of significance). This statistically significant difference between the outgoing connections of early and late-born neurons could arise from the former neurons being present for a much longer period during which they can send out synapses.

Figure A14 ***(previous page)*****:** Distributions of gap junctional connections for all categories of neuron appear to be statistically indistinguishable, suggesting that gap junction formation is relatively unaffected by the functional criticality or time of appearance of the neurons.

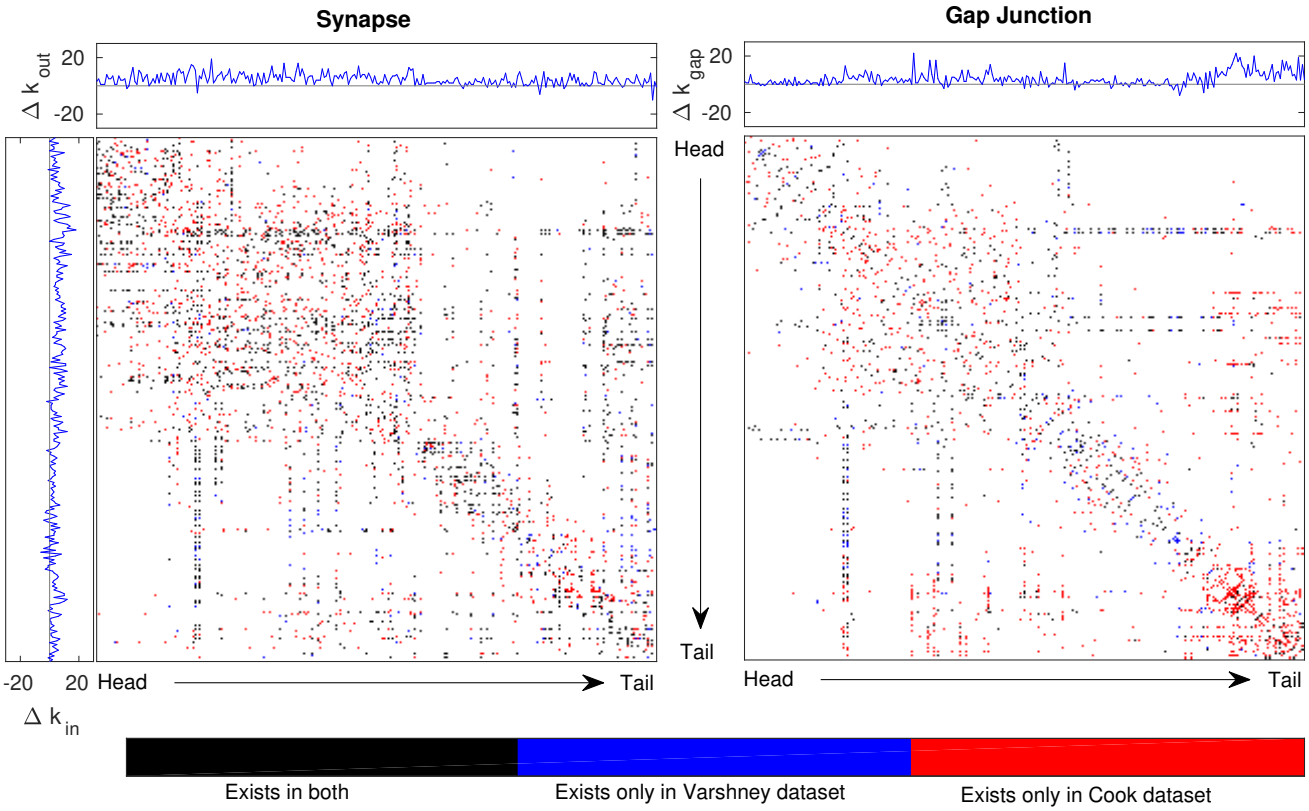


Figure A15: Comparison of the *Caenorhabditis elegans* connectome data obtained from Cook *et al.* [97] and Varshney *et al.* [61]. Difference between the adjacency matrices corresponding to the two connectivity data-sets are shown for synapses (left) and gap-junctions (right). Black-colored entries represent connections common to both datasets, while red- and blue-colored entries denote connections that are found exclusively in the Cook and Varshney datasets, respectively. The columns and rows are arranged according to the spatial ordering of the neuronal cell bodies along the head-tail axis. The horizontal and vertical panels for the synapse matrix shows the difference in the out-degree and in-degree, respectively, of each neuron across the two data-sets. For the gap-junction matrix, the horizontal panel shows the change in degree for each neuron across the data-sets. We note that the two data-sets differ substantially, especially in terms of a large number of additional gap-junctions near the tail region that appear in the database of Cook *et al.*

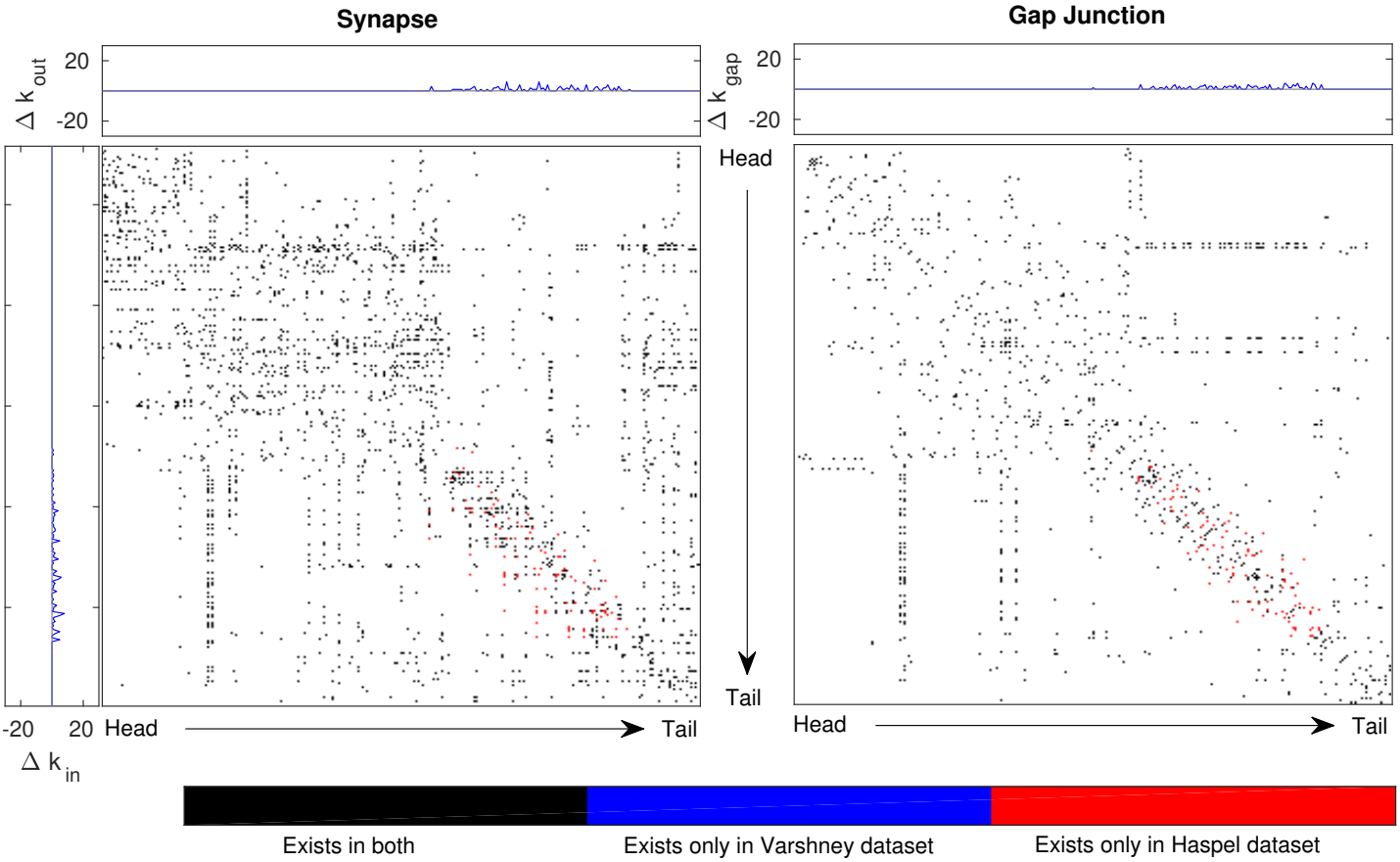


Figure A16: Comparison of the *Caenorhabditis elegans* connectome data obtained from Varshney *et al.* [61] and Haspel *et al.* [96]. Difference between the adjacency matrices corresponding to the two connectivity data-sets are shown for synapses (left) and gap-junctions (right). Black-colored entries represent connections common to both datasets, while red- and blue-colored entries denote connections that are found exclusively in the Haspel and Varshney datasets, respectively. The columns and rows are arranged according to the spatial ordering of the neuronal cell bodies along the head-tail axis. The horizontal and vertical panels for the synapse matrix shows the difference in the out-degree and in-degree, respectively, of each neuron across the two data-sets. For the gap-junction matrix, the horizontal panel shows the change in degree for each neuron across the data-sets. We note that the two data-sets differ relatively little.

Appendix B

The Structure of the Macaque Connectome

Fig. B1 shows some of the macroscopic properties of the network, viz., the cumulative distributions for the number of in-coming, out-going and total connections, and the correlation between the number of in-coming and out-going connections for each node. Fig. B1 (a) shows that the total degree distribution of the nodes [shown in the bottom panel] follows an exponential distribution, in agreement with the observed properties of the network investigated by Modha and Singh [159]. This suggests that the network that we have worked with, which has been processed extensively from the original network of 383 nodes (which contained many redundancies, as explained in Chapter 3), shares the same macroscopic features as the original network. The top and middle panels show the in-degree and out-degree distributions. While both of these appear to also follow an exponential form, the former shows a deviation in the tail indicating that there exist regions that have more in-coming connections than is expected given the form of the distribution. In particular, the four nodes having the highest in-degrees [top panel] that show the largest deviation from the best-fit exponential distribution are all located in the pre-frontal cortex and are also seen to belong to the same module, viz.,#1. This is in accordance with the known cognitive function of prefrontal cortex regions which is high-level multi-modal sensory integration.

In order to see whether regions which attract many in-coming connections also tend to have many out-going connections, we have looked at the correlation between in- and out-degrees in terms of a scatter plot [Fig. B1 (b)]. Here the nodes are colored according

to the module to which they belong, while the relative volumes are represented by the size of the corresponding markers. We note that while most of the nodes are fit well by a linear relation between in-degree and out-degree, there does appear to be several nodes which have a disproportionately higher number of out-going connections than is expected from their in-degree, given the linear relation between the two. Specifically, there are 38 nodes whose out-degree deviate significantly from the value that is expected from the best-fit linear relation with their in-degree, i.e., they are larger than the upper bound given by the root mean square deviation (the upper lighter curve, see figure caption for details). Although the membership of these outliers span across all modules and functional categories (in terms of the role they play in the mesoscopic organization of the connectome), we note that 32% of all connector hubs (R6) and 26% of all satellite connectors (R3) belong to these outliers. As nodes belonging to both of these categories are characterized by having their connections are distributed over several modules, it suggests a possible functional importance of the outlier nodes in coordinating information processing in the Macaque brain.

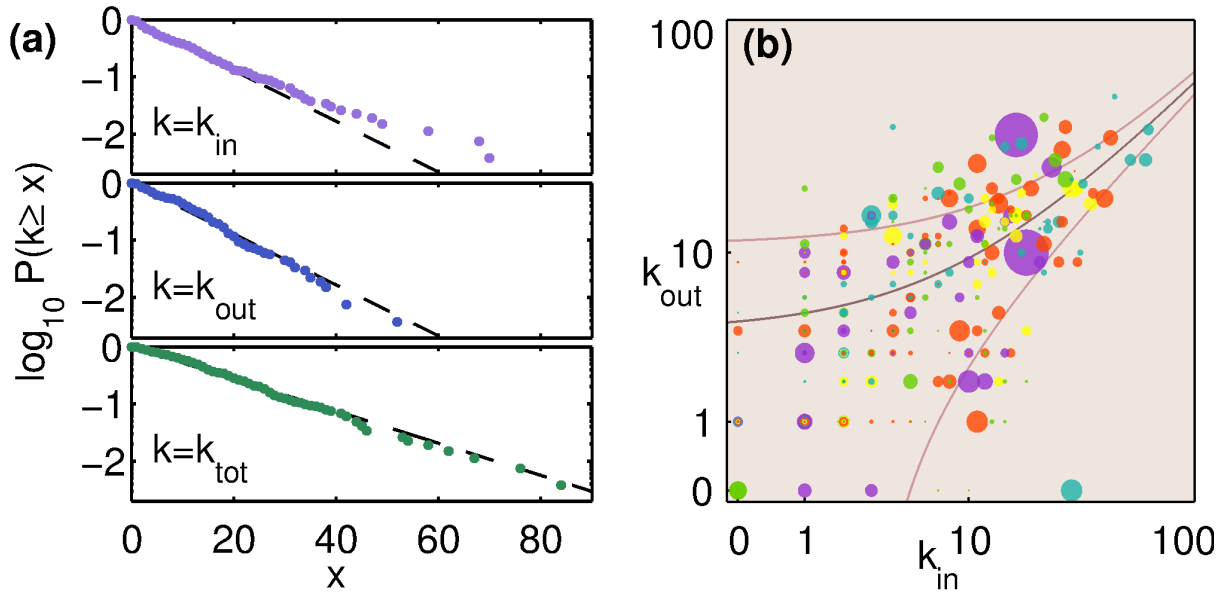


Figure B1: (a) The distributions of the (top) in-degree, (center) out-degree and (bottom) total degree of the Macaque connectome, indicating the best-fit exponential distribution (broken line) in each case. (b) Scatter plot indicating the correlation between in-degree and out-degree of the different nodes in the connectome. The relatively darker central curve represents the best-fit linear relation between k_{in} and k_{out} (the linear correlation coefficient is $r = 0.62$, with a p-value of 0) corresponding to a slope of 0.49. The two lighter curves on either side indicate the root mean square deviation of the empirical data from the best-fit linear relation. The color and sizes of the nodes are same as in Fig. 3.1 of Chapter 3.

Modular Organization of the Connectome

Establishing the robustness of the modular decomposition

As described in the Methods section of Chapter 3, we have ensured that the partitioning of the connectome is not sensitively dependent on the specific method used for the decomposition. Fig. B2 shows that the communities obtained using the Infomap method [168], which is based upon optimally compressing information about dynamic processes on the network, have a high degree of overlap with those obtained using a spectral method [9] that maximizes the modularity Q (for details, see Methods in Chapter 3). While the Infomap method generates a larger number of modules (specifically, 17), not only are many of these extremely small (in some cases comprising only a single node), but several of them are

in fact further subdivisions of the relatively fewer modules (specifically, 5) obtained using the spectral method. The relatively high degree of correspondence between the partitions generated by using techniques that employ completely different principles suggests that the modular decomposition reported here is an intrinsic property of the network, and is not strongly affected by the partitioning method used.

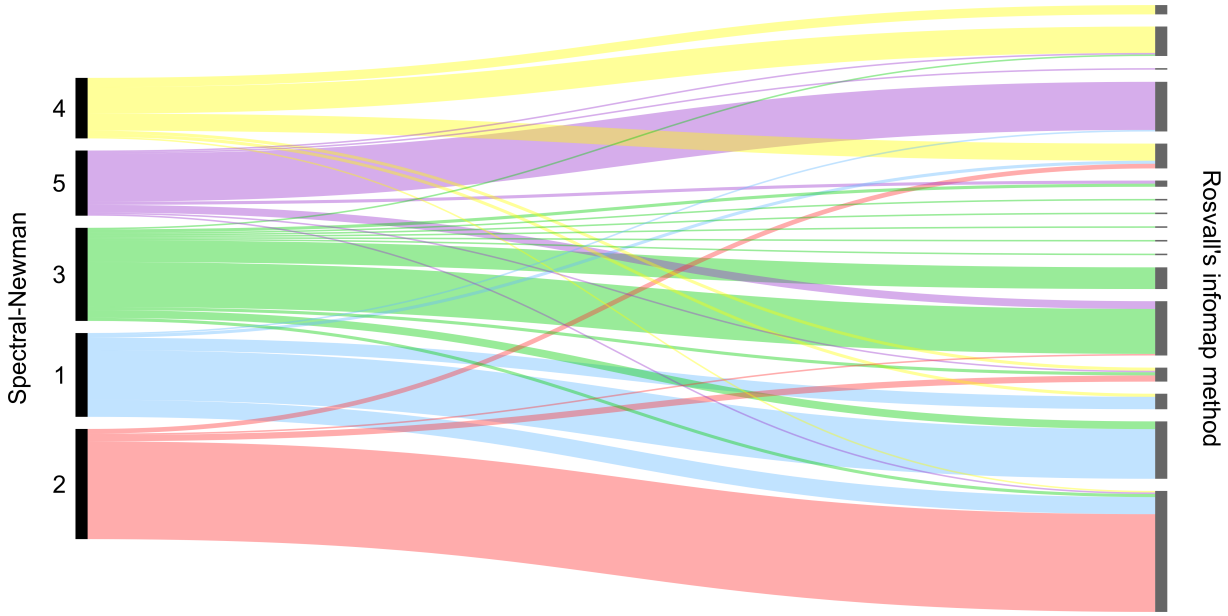


Figure B2: Visual representation of the comparison between the modular decomposition of the Macaque connectome obtained using spectral partitioning [left] with that obtained using the Infomap method [right]. The modules are represented as vertical bars, connected by bands which are colored according to the module obtained using the spectral method from which they originate [using the same color scheme as in Fig. 3.1 (d) of Chapter 3]. This alluvial diagram has been created using the online visualization tool RAW [177].

To verify that the method used for maximizing Q does not alter our results significantly, we have performed 10^3 realizations of a stochastic simulated annealing algorithm for detecting communities [166]. As mentioned in the Methods (see Chapter 3), by comparing between these large number of optimal partitionings of the network, we can determine the extent to which the modular groupings among the different nodes is robust. Fig. B3 shows the Modularity Q values corresponding to these realizations, using a representation such that similar partitionings (corresponding to the circles) occur close to each other in the two-dimensional plane orthogonal to the axis representing Q. The two-dimensional coor-

dinates of each circle in this plane is obtained by Curvilinear Component Analysis (CCA, see Ref. [271]) as described in Ref. [166].

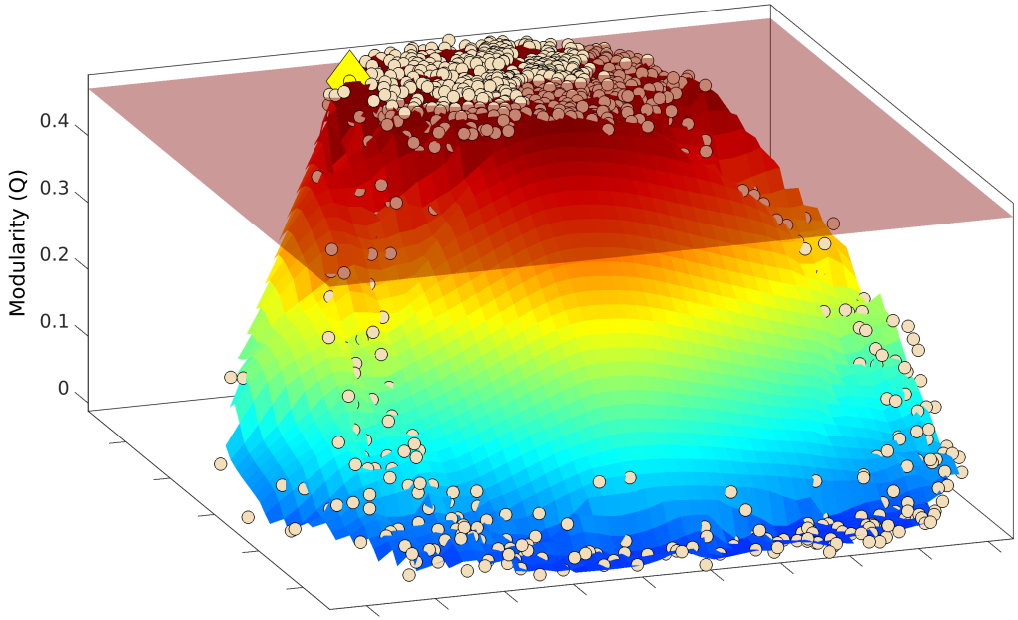


Figure B3: Modularity of the Macaque connectome, shown as a function reconstructed from 10^3 partitionings (circles) obtained through a simulated annealing method for determining communities [166]. The axes on the horizontal plane orthogonal to the vertical axis that corresponds to modularity Q represent embedding dimensions that are themselves complex functions of the partition space, such that the scale of these axes are irrelevant. The distance between the partitionings (whose positions on the horizontal plane are obtained by CCA) are indicative of the degree of dissimilarity between the corresponding modular partitions of the network. The partition obtained by the deterministic spectral method yielding a Q-value of $Q_{spectral} = 0.485$ (diamond), and which has been used for our analysis, is seen to occur in the high-modularity plateau comprising a large number of similar partitions, all having a high value of Q. The 291 partitionings that occur at the top of this surface, whose Q values differ by less than 3% from $Q_{spectral} = 0.485$ (specifically, the circles lying above the translucent plane corresponding to $Q = 0.47$ shown in the figure), have been used to determine the robustness of the modular identities of the different nodes in the connectome, as shown in Fig. B4 (left).

As can be seen from Fig. B3, there are a large number of partitionings having high values of Q that occur close to each other in the plateau and where the partition obtained from the spectral method (diamond, having a Q-value of $Q_{spectral} = 0.485$) that has been for our analysis is also seen. This suggests that the modular decomposition of the nodes in these high Q partitionings are similar to that determined by the spectral method. Fig. B4 shows

the brain regions whose modular identity is invariant across all the partitionings whose Q differs by less than 3% (i.e., $Q > 0.47$, left panel) and 7% (i.e., $Q > 0.45$, right panel) from $Q_{spectral}$. The conserved modular memberships of a large fraction ($\sim$ 70%) of the brain regions across all the different partitionings possessing high modularity (highlighted nodes in Fig. B4 [left]; see Table B1 for their identities) emphasizes that the modular mesoscopic organization we have described here does not depend sensitively on the method used to partition the network, underlining that it is an intrinsic property of the Macaque connectome.

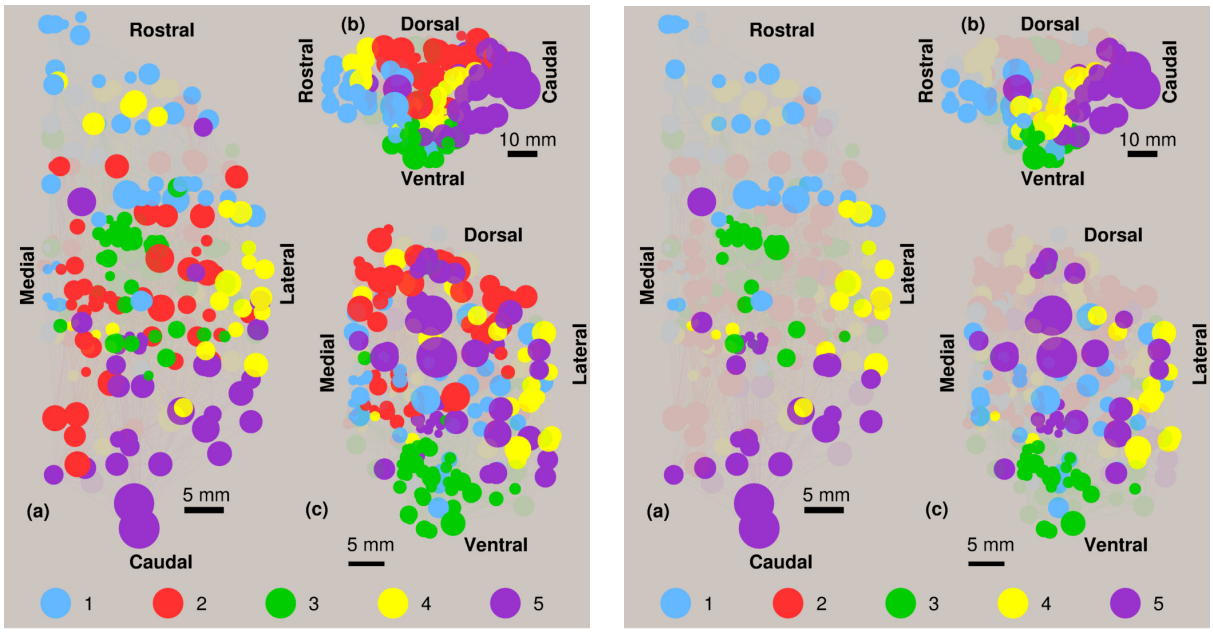


Figure B4: The network of brain regions shown in (a) horizontal, (b) sagittal and (c) coronal projections, indicating the regions (highlighted) whose modular memberships are invariant across the partitionings obtained by the spectral method (used in our analysis) as well as those obtained by simulated annealing, whose Q differs by less than (left) 3% and (right) 7% from $Q_{spectral} = 0.485$. As in Fig. 3.1 (a-c) in Chapter 3, the modular membership of each node is represented by its color (see color key at the bottom of each panel), the spatial positions of the nodes are specified by the three-dimensional stereotaxic coordinates of the corresponding regions, and node sizes provide a representation of the relative volumes of the corresponding brain regions (the spatial scale being indicated by the horizontal bar shown next to each projection). (Left) Within the 291 partitionings that have $Q > 0.47$, around 70% of the 266 brain regions have the same modular membership as that seen in the spectral modular decomposition, underlining the robustness of their modular identities. (Right) For the 625 partitionings that have $Q > 0.45$, we see a much higher degree of variation in the modular identities of the regions across the partitionings as a result of including those with much lower values of Q. Specifically, modules #1, #3 and #5 are seen to have several nodes that are robust (i.e., consistently belong to the corresponding module) across the partitionings, while, for module #4, only some of the nodes belonging to the temporal lobe have this property. The nodes belonging to module #2, on the other hand, change from one partitioning to another.

Table B1: Brain regions highlighted in Fig. B4 (left) whose modular membership is conserved across all the 291 distinct partitionings with $Q > 0.47$, arranged according to the modules, and subsequently into the largest anatomical subdivision (viz., lobe / nuclei), to which they belong. For each region, the corresponding within-module degree z-score and the participation coefficient are displayed in the last two columns (see Methods for details).

Module #1			
Lobe/nuclei	region	zscore	pcoeff
FL	14r	1.154	0.344
	13L	0.481	0.148
	13M	0.443	0.102
	13a	2.837	0.471
	32	3.061	0.688
	10m	-0.230	0.000
	10v	-0.791	0.000
	10d	-0.866	0.000
	10o	0.443	0.000
	12o	2.613	0.619
	12m	0.593	0.263
	12r	0.107	0.069
	12l	2.239	0.659
	11l	0.368	0.108
	11m	1.042	0.447
	PrCO	0.032	0.559
	6Vb	-0.305	0.457
	6Va	-0.267	0.438
TL	TPag	-0.529	0.000
	TPg	-0.305	0.000
	TPdgv	-0.604	0.000

	TPdgd	-0.604	0.000
	Su#2	-0.754	0.000
	Sb	-0.754	0.000
cing	24b	0.780	0.549
Insula	Ial	-0.267	0.000
	Iam	0.630	0.373
	Iapm	0.705	0.515
	Iai	1.752	0.432
Thal	AM#1	-0.492	0.569
	Cim	-0.828	0.000
	Cif	-0.716	0.444
	Cdc	-0.754	0.560
	MDcd	-0.828	0.000
	MDpm	-0.754	0.000
	MDfi	-0.679	0.000
BG	SI#2	0.219	0.410
	Pu_r	0.219	0.429
OFC	OFC	-0.679	0.408
Module #2			
Lobe/nuclei	region	zscore	pcoeff
FL	F5	2.706	0.515
	4c	-0.575	0.180
	F4	0.807	0.355
	F7	3.512	0.551
	F2	2.418	0.360
	M2-FL	-0.862	0.480

	F6	1.267	0.512
	M1-FL	2.073	0.256
	MI-of	-0.402	0.272
PL	1#1	0.691	0.174
	2#1	0.634	0.121
	3b	0.346	0.322
	3a	0.173	0.227
	SII-f	-0.920	0.000
	PR#4	-0.690	0.000
	PFop	-0.805	0.000
	PGop	-0.690	0.000
	PFG#1	0.404	0.140
	PF#1	0.807	0.353
	AIP	-0.517	0.165
	MIP	0.749	0.307
	PEm	0.979	0.463
	5_Foot	-1.035	0.000
	PEc#1	1.152	0.228
	PGm	2.188	0.627
	PECg	1.267	0.399
OL	V6A	0.231	0.375
cing	24d	-0.114	0.290
	23c	2.188	0.563
	TSA	-0.057	0.355
Insula	Ri#1	-0.632	0.494

Thal	Pcn	0.289	0.650
	CM#2	0.231	0.226
	Csl	-0.287	0.640
	Ret	-0.287	0.338
	Pul.o	0.576	0.291
	X	-0.287	0.427
	VPS	-0.862	0.000
	VPM	-0.460	0.278
	VPLo	-0.517	0.000
	VPLc	-0.057	0.194
	VLm	-0.575	0.180
	VLps	-0.575	0.320
	VLo	0.058	0.308
	VLc	-0.114	0.198
	VApc	-0.460	0.375
BG	Pu_c	-0.460	0.375
Module #3			
Lobe/nuclei	region	zscore	pcoeff
TL	TFM	-0.309	0.000
	TFL	0.190	0.254
	35	2.044	0.469
	36c	1.117	0.111
	36r	2.329	0.177
	36p	-0.452	0.000
	EI	0.547	0.453
	ER#1	0.048	0.499

	28m	-0.594	0.180
	ECL	-0.309	0.408
	EC#2	0.689	0.487
	Pros.	-0.166	0.430
	PaS	-0.737	0.406
	TH	3.327	0.666
	PrS	-0.452	0.514
	CA1	0.333	0.159
	DG	-0.808	0.245
cing	29d	-0.951	0.560
BG	Bla	0.261	0.442
	Abpc	0.832	0.320
	Bi	0.903	0.412
	ABd	0.261	0.000
	Bvl	-0.024	0.000
	ABv	0.261	0.000
	MB	-0.024	0.290
	ABvm	0.547	0.204
	ABmg	0.832	0.420
	A	-0.095	0.111
	I#2	-0.879	0.278
	ME#1	-0.166	0.430
	CE#1	-0.095	0.360
	AHA	-0.594	0.180
	PAC2	-0.238	0.000
	COp	-0.808	0.000

	NLOT	-0.808	0.000
	COa	-0.523	0.397
	Ldi	1.117	0.216
	Ld#2	-0.166	0.219
	Lv	0.974	0.229
	Lvl	0.618	0.137
Module #4			
Lobe/nuclei	region	zscore	pcoeff
FL	M9	0.227	0.677
	D9	-0.257	0.602
	46v	2.437	0.741
	46d	0.848	0.477
	8B	2.299	0.629
TL	A1	0.641	0.340
	STPg	-0.326	0.418
	ProK	-0.326	0.231
	paAc	0.089	0.381
	L#1	0.019	0.320
	CL#4	-0.671	0.480
	AL#4	-0.464	0.426
	ST3	1.194	0.370
	ST2	0.641	0.445
	ST1	0.227	0.469
	Tpt	0.848	0.461
	TPOc	0.710	0.571
	TPOr	-0.257	0.492

	TAa	0.434	0.441
Thal	MG	-0.533	0.000
	SG	-0.188	0.519
	Li	0.019	0.615
Module #5			
Lobe/nuclei	region	zscore	pcoeff
FL	45A	0.385	0.562
	8Ac	-1.003	0.000
PL	LIPe	-0.268	0.660
	LIPi	-0.023	0.568
	VIP	0.793	0.507
	PIP#1	-0.268	0.000
TL	CITv	0.466	0.560
	TEm	-0.350	0.691
	PITd	-0.023	0.142
	PITv	0.058	0.500
	IPa	0.385	0.710
	MT	2.997	0.314
	FST	1.446	0.454
	MSTp	0.140	0.231
	MSTd	1.446	0.497
OL	V3A	0.793	0.159
	V3v	0.711	0.000
	V4t	0.140	0.338
	DLr	-0.921	0.000
	DLc	-0.921	0.000

	V4v	-0.758	0.000
	VPP	-0.921	0.000
	V6	1.283	0.447
	DP	0.303	0.443
	VOT	-0.595	0.000
	V1	1.283	0.250
	V2	3.160	0.387
Thal	LGN	-0.595	0.278
	PIl-s	-0.921	0.000
	PIp	-0.840	0.000
	PIm	-0.513	0.245
	PIl	-0.431	0.000
	PIc	-0.431	0.219
	PLa#1	-0.921	0.000
	PLvl	-0.758	0.000
	PLvm	-0.758	0.000
BG	Cd_g	-0.105	0.653
MB	MB#2	-0.187	0.298

Modular decomposition of the cortical and sub-cortical subdivisions of the Macaque brain

As mentioned in Chapter 3, there is no simple correspondence between the modules and the anatomical subdivisions of the brain. The nodes of the connectome we have investigated are brain regions that belong to larger subdivisions, such as the prefrontal cortex, which in turn are part of broader anatomical categories such as the frontal lobe. The association between the network modules and the largest subdivisions have been shown in Fig. 3.1 (d) in Chapter 3. A more detailed representation of this relation is given in terms of the modular spectra of the anatomical subdivisions in Table B2 which indicates how the regions belonging to each subdivision are distributed among the five modules. We note that some of the subdivisions constitute a single brain region in the connectome we consider (e.g., Visual area V1 in the Occipital lobe), so that they belong exclusively to one of the modules. Larger subdivisions that comprise multiple regions, on the other hand, can have their constituent regions distributed non-uniformly among several modules. In such cases, we highlight the dominant module(s) of the subdivision, i.e., those amongst the five modules having the largest number of brain regions, in the table. The spatial layout of the brain regions belonging to these larger subdivisions, colored according to the modules to which they belong, are also shown in Figs. B5 and B6. Note that, the regions belonging to the parietal lobe occur predominantly in module #2, while those in the occipital lobe occur predominantly in module #5 (see Fig. B5). Fig. B6 suggests that the regions belonging to the basal ganglia mostly occur in module #3.

Table B2: Modular decomposition of the brain regions in different anatomical subdivisions of the Macaque brain.

Lobe/Nuclei (no. of brain regions)	Subdivision [abbreviation]	modular distrib.				
		1	**2**	**3**	**4**	**5**
Frontal lobe (58)	beltline of sensorymotor syst. [belt_sm]	0	1	0	0	0
	Prefrontal cortex [PFC]	18	5	0	11	2
	Supplementary motor area [Area 6]	3	11	1	0	0
	Primary motor area [MI]	0	4	2	0	0
Temporal lobe (56)	Ventral temporal cortex [TCV]	4	0	6	0	0
	Parahippocampal cortex [PHC]	3	0	11	0	0
	Hippocampus [Hip]	0	0	3	0	0
	Superior temporal gyrus [STG]	0	0	0	11	0
	Inferotemporal area [TE]	0	0	3	0	4
	Superior temporal sulcus [STS]	0	1	0	5	5
Parietal lobe (27)	Primary somatosensory cortex [S1]	0	4	0	0	0
	Secondary somatosensory cortex [S2]	0	1	1	0	0
	beltline of sensory syst. [belt_s]	1	0	0	0	0
	Rostral parietal area [PR#4]	0	1	0	0	0
	Somatosensory association area [7#1]	0	5	0	2	0
	Cortex of intraparietal sulcus [PCip]	0	2	0	0	4

	Dorsal parietal cortex [PCd#2]	0	6	0	0	0
Occipital lobe (16)	Visual anterior cortex [VAC]	0	1	0	0	12
	Visual area V1 [V1]	0	0	0	0	1
	Prostriate cortex [ProST]	1	0	0	0	0
	Visual area V2 [V2]	0	0	0	0	1
Thalamus (53)	Anterior nuclei [AN]	2	1	0	0	1
	Midline nuclei [ML]	5	1	1	0	0
	Geniculate nucleus [GN]	0	0	0	1	1
	Intralaminar nuclei [IL2]	1	3	0	1	0
	Massa intermedia [MI1]	0	1	0	0	0
	Posterior nuclei [PN]	0	0	0	2	0
	Reticularis thalami [Ret]	0	1	0	0	0
	Pulvinaris thalami [Pul#1]	0	1	0	3	8
	Medial dorsal nucleus [MD]	3	2	0	1	0
	Ventrolateral nuclei [VN]	1	12	0	0	0
Basal Ganglia (31)	Amygdala [Amyg]	0	0	22	0	0
	Substantia nigra [SN]	0	0	1	0	0
	Substantia innominata [SI#2]	1	0	0	0	0
	Nucleus subthalamus [Sub.Th]	0	0	1	0	0
	Globus pallidus [GPe]	0	0	0	0	1
	Striatum [STR]	1	1	1	0	1
	Claustrum [Clau]	0	1	0	0	0
Cingulate Gyrus (13	Area 24 [24]	3	1	0	0	0
	Area 23 [23]	2	2	0	0	0
	Area 26 [26]	0	0	2	2	0

	Area 25 [25]	0	0	1	0	0
Insula (9)	Granular insular cortex [Ig#1]	0	0	1	0	0
	Retroinsular cortex [Ri#1]	0	1	0	0	0
	Insular proisocortex [IPro]	0	1	0	0	0
	Parainsular field [Pi#1]	0	0	1	0	0
	Anterior insula [IA]	4	0	1	0	0
Hypothalamus (1)	Hypothalamus [Hyp]	0	0	1	0	0
Midbrain (1)	Midbrain [MB]	0	0	0	0	1
Olfactory complex (1)	Olfactory complex [OFC]	1	0	0	0	0

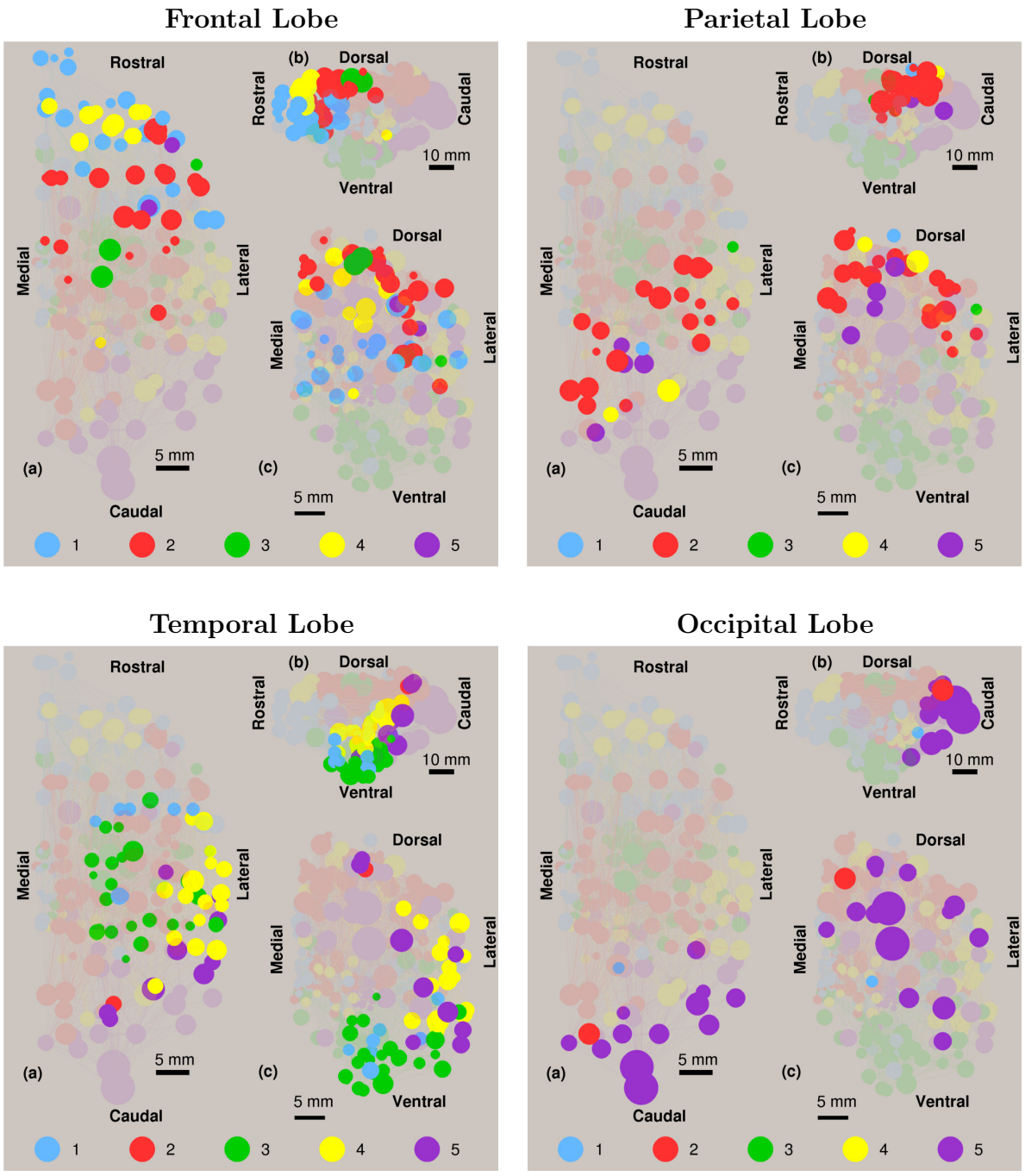


Figure B5: The network of brain regions shown in (a) horizontal, (b) sagittal and (c) coronal projections, indicating the modular memberships of the regions (highlighted) that belong to the frontal (top left), parietal (top right), temporal (bottom left) and occipital (bottom right) lobes. As in Fig. 3.1 (a-c) in Chapter 3, the modular membership of each node is represented by its color (see color key at the bottom of each panel), the spatial positions of the nodes are specified by the three-dimensional stereotaxic coordinates of the corresponding regions, and node sizes provide a representation of the relative volumes of the corresponding brain regions (the spatial scale being indicated by the horizontal bar shown next to each projection).

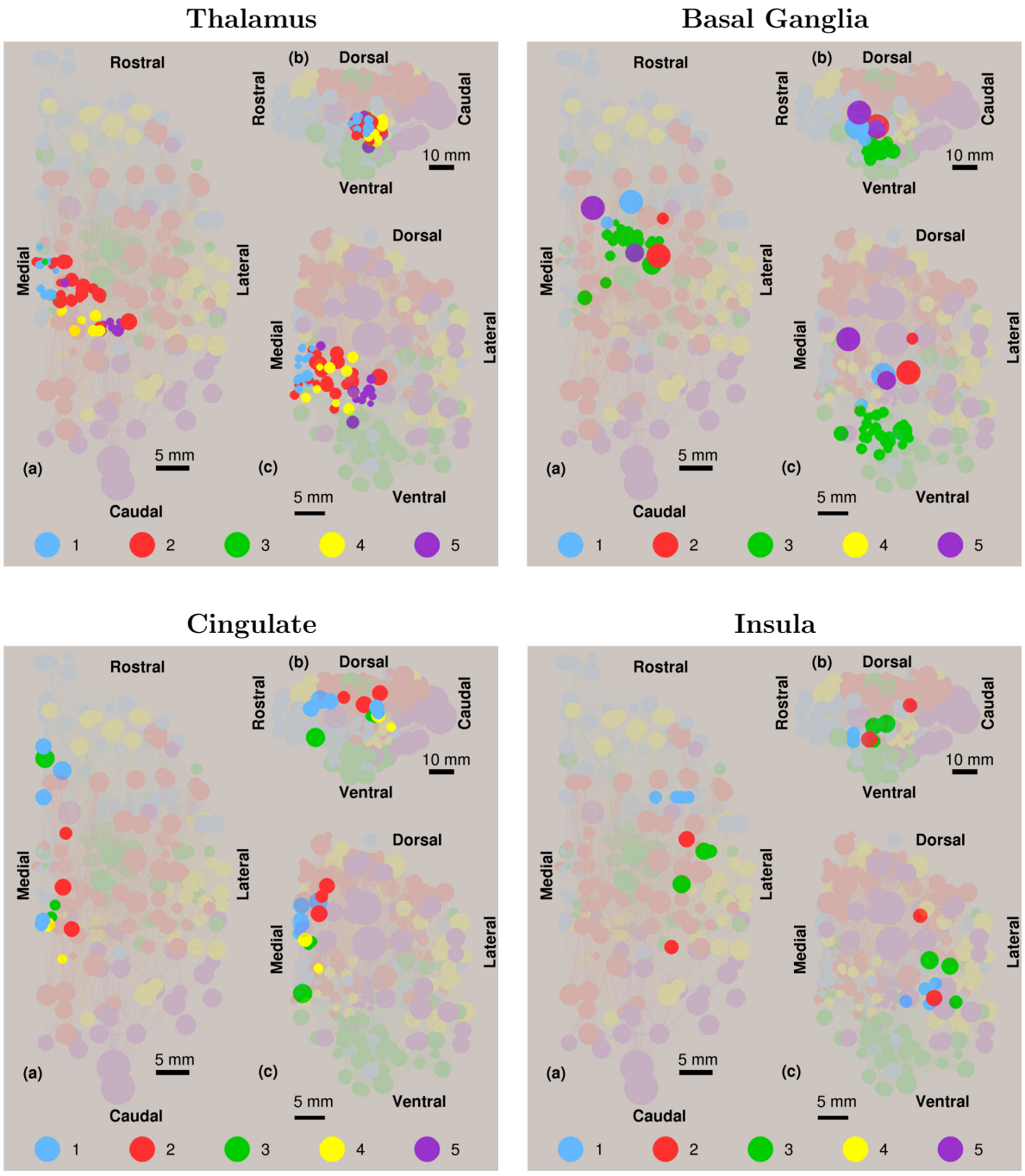


Figure B6: The network of brain regions shown in (a) horizontal, (b) sagittal and (c) coronal projections, indicating the modular memberships of the regions (highlighted) that belong to the thalamus (top left), basal ganglia (top right), cingulate (bottom left) and insula (bottom right). As in Fig. 3.1 (a-c) in Chapter 3, the modular membership of each node is represented by its color (see color key at the bottom of each panel), the spatial positions of the nodes are specified by the three-dimensional stereotaxic coordinates of the corresponding regions, and node sizes provide a representation of the relative volumes of the corresponding brain regions (the spatial scale being indicated by the horizontal bar shown next to each projection).

Functional characterization of modules

As mentioned in Chapter 3, we have investigated a possible structure-function correlation in the mesoscopic organization of the connectome, which would be reflected in the modules being predominantly associated with certain functionalities. In Table B3 we list the known functions (obtained from the literature) either of the brain regions belonging to each of the modules, or of the broader subdivisions to which such regions belong (the first column indicating the lobe or nuclei, and the second specifying the areas comprising it). As the different regions belonging to a subdivision may lie in distinct modules, the fraction of all the regions in a subdivision that are in a specific module are indicated in the third column. The role that these regions play in terms intra- and inter-modular communication can be inferred from the average values (computed over all the regions in the subdivision that are in the same module) of the participation coefficient, $\langle p \rangle$, and the within-module degree z-score, $\langle z \rangle$, which are shown in the fourth and fifth columns, respectively. Some of the brain regions in a subdivision that have been investigated relatively more extensively are mentioned in the sixth column, while the seventh column provides a non-exhaustive list of the functions that are known to be associated with these regions and/or the subdivision to which they belong (along with references to the relevant literature). As alluded to in Chapter 3, regions belonging to the same module have certain functions that appear to complement each other in carrying out a specific cognitive task, e.g., high-level multimodal sensory integration and decision-making (module #1), motor control and somato-sensory processing (module #2), memory and emotion (module #3), auditory processing (module #4) and visual processing (module #5).

Table B3: Functional characterization of modules

Module #1						
Lobe/Nuclei	Subdivision	Frac.	$\langle p \rangle$	$\langle z \rangle$	Notable regions	Known functions
Frontal lobe	PFC	18/36	0.24	0.63	10, 11, 13, 14	sensory integration, decision making [184, 185, 186, 187, 188]
	Area 6	3/15	0.48	-0.18	6Va, 6Vb, PrCo	complex locomotion (e.g. climbing) [272, 273]
Temporal lobe	TCv	4/10	0.00	-0.51		
	PHC	3/14	0.17	-0.68	EO	olfaction [274]
Parietal lobe	belt_s	1/1	0.64	-0.70		
Occipital lobe	ProSt	1/1	0.50	-0.86		
Thalamus	AN	2/4	0.52	-0.62		
	ML	5/7	0.39	-0.78		
	IL#2	1/5	0.00	-0.90		
	MD	3/6	0.00	-0.75		
	VN	1/12	0.57	-0.15		
Basal Ganglia	SI#2	1/1	0.41	0.22		
	STR	1/4	0.43	0.22	Pu_r	motor skills, reinforcement learning [275]

Cingulate gyrus	Area 24	3/4	0.57	0.60	24a, 24b, 24c	emotional behavioural control [276]
	Area 23	2/4	0.61	-0.55	23a,23b	multi-sensory integration [276]
Insula	IA	4/5	0.33	0.70	Iam, Iai	social cognition [277]
olfactory complex	OFC	1/1	0.40	-0.68		olfaction [274]
Module #2						
Lobe/nuclei	Subdi-vision	Frac.	$\langle p \rangle$	$\langle z \rangle$	Notable regions	Known function
Frontal lobe	belt_sm	1/1	0.54	-0.52		
	PFC	5/36	0.48	-0.46	45B, 8Ad	saccadic guidance (frontal eye field) [278]
	Area 6	11/15	0.36	0.66		complex locomotion (climbing etc.) [272]
	MI	4/6	0.27	0.20	M1-FL, M1-HL	voluntary movement (primary motor area) [183]
Temporal lobe	STS	1/11	0.44	-0.98		

Parietal lobe	S1	4/4	0.21	0.46		primary somatosensory cortex [183]
	S2	1/2	0.00	-0.92	SII-f	secondary somatosensory area (face representation) [279]
	PR#4	1/1	0.00	-0.69		
	Area 7	5/7	0.10	-0.28	PF#1, PFG#1	visual-motor coordination [280]
	PCip	2/6	0.24	0.12	AIP, MIP	visual control of reaching & pointing [210, 211]
	PCd#2	6/6	0.39	0.91	PEm, PEc#1	somesthesia & motor control [281]
Occipital lobe	VAC	1/13	0.38	0.23		
Thalamus	AN	1/4	0.72	-0.86		
	ML	1/7	0.59	-0.63		
	IL#2	3/5	0.50	0.07		
	MI#1	1/1	0.00	-1.09		
	Ret	1/1	0.34	-0.29		
	Pul#1	1/12	0.29	0.58		
	MD	2/6	0.60	-0.50		

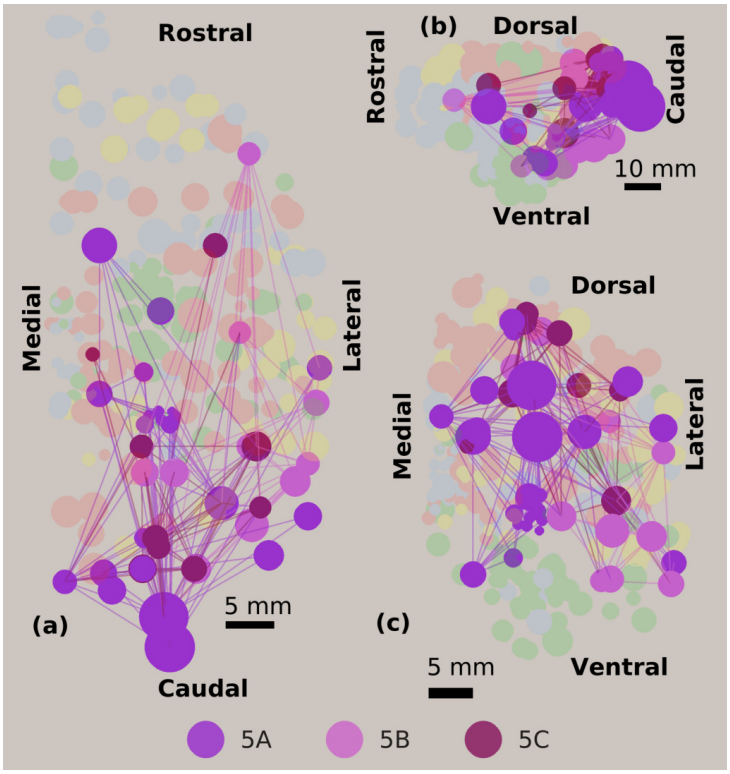
	VN	12/13	0.19	-0.45		somatosensory information relay [183]
Basal Ganglia	STR	1/4	0.37	-0.50	Pu_c	motor skills, reinforcement learning [275]
Cingulate gyrus	Area 24	1/4	0.29	-0.11		
	Area 23	2/4	0.46	1.06		
Insula	Ri#1	1/1	0.49	-0.63		
	Ipro	1/1	0.00	-1.03		
Module #3						
Lobe/nuclei	Subdi-vision	Frac.	$\langle p \rangle$	$\langle z \rangle$	Notable regions	Known function
Temporal lobe	TCv	6/10	0.16	0.82	35, 36c, 36r	visual perception & memory of objects [282]
	PHC	11/14	0.46	0.09	TH	spatial memory [283]
	Hip	3/3	0.28	-0.38		spatial cognition [284, 285] and recognition memory [286]
	TE	3/8	0.65	0.17		
Parietal lobe	S2	1/2	0.50	-1.16		

Basal Ganglia	Amyg	22/22	0.20	0.1		emotional response [189, 190]
	SN	1/1	0.61	-0.88		
	STR	1/4	0.57	-0.80	Cd_t	reinforcement learning [275]
Cingulate gyrus	Area 26	2/2	0.57	-0.88		
	Area 25	1/1	0.65	2.48		
Insula	Ig#1	1/1	0.60	0.83		
	Pi#1	1/1	0.64	0.19		
	IA	1/5	0.70	0.90		
Hypothalamus	Hyp	1/1	0.69	-0.80		
Module #4						
Lobe/nuclei	Subdivision	Frac.	$\langle p \rangle$	$\langle z \rangle$	Notable region	Known function
Frontal lobe	PFC	11/36	0.55	0.25	46d, 46v	working memory [287, 288]
Temporal lobe	STG	11/11	0.39	0.17	A1	auditory cortex [289, 290]
	STS	5/11	0.52	-0.04	TPOc, TAa, Pga	complex sound processing [291, 290]
Parietal lobe	Area 7	2/7	0.69	1.40	PG#1	somato-motor coordination [280]

Thalamus	GN	1/2	0.00	-0.53	MG	auditory information relay [183]
	IL#2	1/5	0.66	-0.67		
	PN	2/2	0.56	-0.08		
	Pul#1	3/12	0.00	-1.22		
	MD	1/6	0.66	-0.60		
Cingulate gyrus	Area 26	2/4	0.30	-0.81		
Module #5						
Lobe/nuclei	Subdi-vision	Frac.	$\langle p \rangle$	$\langle z \rangle$	Notable region	Known function
Frontal lobe	PFC	2/36	0.28	-0.30	45A, 8Ac	saccadic guidance (frontal eye field) [278]
Temporal lobe	TE	4/7	0.47	0.03	CITv, PITd, PITv,TEm	ventral visual pathway [198, 199]
	STS	5/11	0.44	1.28	MT, MST, FST	dorsal visual pathway [198, 199]
Parietal lobe	PCip	4/6	0.43	0.06	LIP, VIP, PIP	visual attention [208, 209]
Occipital lobe	VAC	12/13	0.11	-0.23	V3A, V3B, V6	visual cortex [183]

	V1	1/1	0.25	1.28		primary visual cortex [183]
	V2	1/1	0.39	3.15		secondary visual cortex [183]
Thalamus	AN	1/4	0.69	-0.51		
	GN	1/2	0.28	-0.59	LGN	visual information relay [183]
	Pul#1	8/12	0.06	-0.70		visual processing [292]
Basal Ganglia	Gpe	1/1	0.67	-0.84		
	STR	1/4	0.65	-0.10	Cd_g	reinforcement learning [275]
Mid brain	MB	1/1	0.30	-0.19		

The structure-function correlation associated with the mesoscopic organization, can be seen not only at the level of modules (as indicated by the Table B3 above) but can be extended even further. As mentioned in Chapter 3, we have sbjected module #5 to further partitioning which yields three sub-modules. Fig. B7 shows the nodes in module #5 that belong to these sub-modules. We find that they ae associated with distinct funtionalities, with 5A containing the visual cortex and almost all the sub-cortical components, while the regions identified with different visual processing pathways, viz., the ventral and dorsal streams belong to 5B and 5C, respectively.



5A	5B	5C
LGN	8Ac	45A
PIl-s	VIP	LIPe
PIp	FST	LIPi
PIm	MSTp	CITv
PIl	MSTd	PITd
PIc	V3A	PITv
PLa#1	V4t	IPa
PLvl	DP	V3v
PLvm	LD#1	V4v
PIP#1		VOT
TEm		
MT		
V3d		
DLr		
DLc		
VPP		
DI#1		
V6		
V1		
V2		
MB#2		
Cd_g		
GPe		

Figure B7: The network of brain regions shown in (a) horizontal, (b) sagittal and (c) coronal projections, indicating that the nodes in module #5 (highlighted) can be further grouped into three sub-modules. The sub-modular membership of each node of module #5 is represented by its color (see color key at the bottom) with the list of brain regions belonging to each of the three sub-modules shown in the table in the right. Sub-module #5A is seen to comprise primary visual regions and subcoritcal regions, while sub-modules #5B and #5C contain regions that belong to the ventral and dorsal visual pathways, respectively. The node sizes provide a representation of the relative volumes of the corresponding brain regions (the spatial scale being indicated by the horizontal bar in each panel). The spatial positions of the nodes are specified by the three-dimensional stereotaxic coordinates of the corresponding regions. Links indicate the directed nerve tracts connecting pairs of brain regions, and are colored in accordance with their source nodes.

Categorization of nodes in terms of inter- intra-modular connectivity

As described in Chapter 3, the role played by each of the brain regions in the mesoscopic organization of the connectome can be classified into seven categories according to their intra- and inter-modular connectivity, viz., R1: ultra-peripheral, R2: peripheral, R3: satellite connector, R4: kinless, R5: provincial hub, R6: connector hub, and R7: global hub (note that there are no regions in the Macaque brain belonging to the categories R4 and R7). With the exception of module #4 which has no region playing the role of a provincial hub, each module has a qualitatively similar distribution of its regions across these categories (Fig. B8, top).

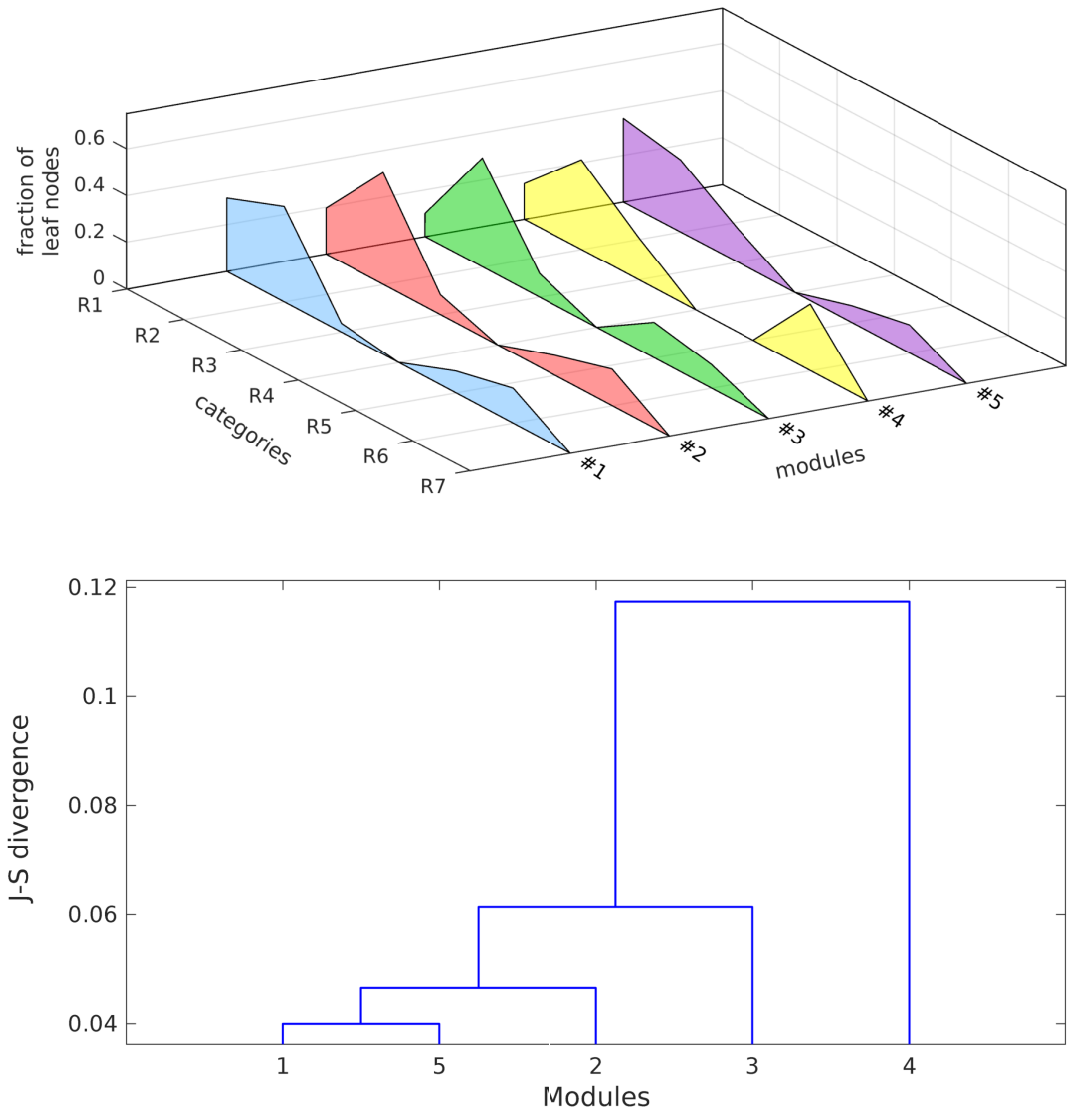


Figure B8: The distribution of the regions across the different categories R1-R7 (see Fig. 3.2 in Chapter 3) is similar for different modules (top), with the sole exception of module #4 which does not possess any provincial hub (R5) nodes. This is illustrated in the dendrogram (bottom) that represents the extent of similarity between these distributions, quantified by the Jensen-Shannon divergence, for the different modules.

The extent of similarity between the modules is represented by the dendrogram shown in Fig. B8 (bottom) in which the distance between the distributions across $x \in \{\mathrm{R1}, \ldots, \mathrm{R7}\}$ for two modules A and B , viz., $P_A(x)$ and $P_B(x)$, is measured in terms of the Jensen-Shannon divergence:

$$JSD(P_A, P_B) = \frac{1}{2} \sum_x \left[P_A(x) \ln P_A(x) + P_B(x) \ln P_B(x) - \{P_A(x) + P_B(x)\} \ln \left(\frac{P_A(x) + P_B(x)}{2} \right) \right].$$

After calculating JSD between all pairs of modules, we perform hierarchical clustering of the nodes to generate a dendrogram showing how the modules are related to each other [293] as follows. We successively merge pairs of modules that are closest to each other in terms of JSD. After each merger, the JSD between the remaining clusters are calculated and the process repeated until only a single cluster is left. The distance between two clusters is obtained using a single-linkage clustering algorithm, i.e., we chose the minimum of the distance between any member of the first cluster to any member of the second cluster.

This quantification of the difference between a pair of distributions is also employed in Fig. B9 to indicate the extent of similarity between the various anatomical subdivisions of the brain, the corresponding distributions of whose regions across the categories R1-R7 is shown in Fig. 3.2 (b) in Chapter 3.

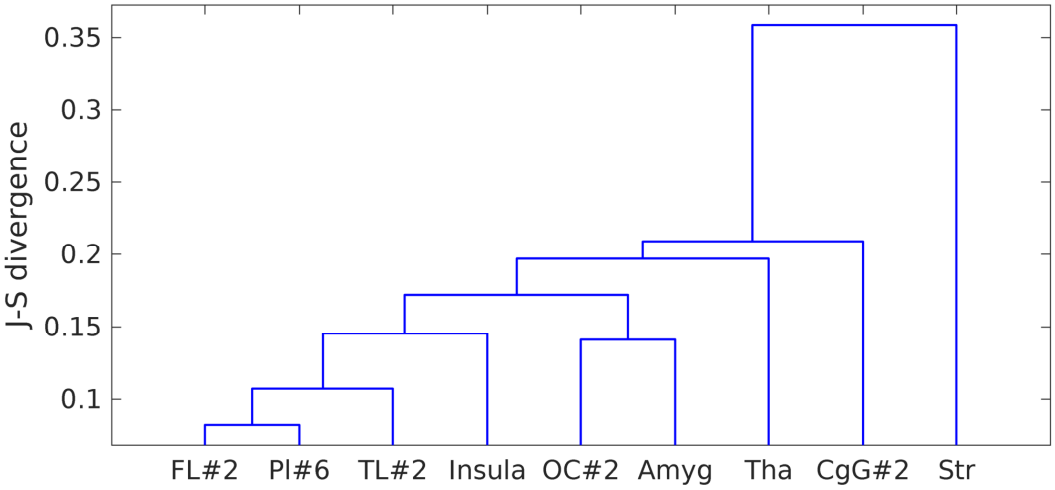


Figure B9: Dendrogram illustrating the extent of similarity between several anatomical subdivisions of the brain, viz., *Tha*: Thalamus, *FL#2*: Frontal Lobe, *P1#6*: Parietal Lobe, *CgG#2*: Cingulate Gyrus, Insula, *TL#2*: Temporal Lobe, *OC#2*: Occipital Lobe, *Amyg*: Amygdala and *STR*: Striatum, in terms of the distribution across the categories R1-R7 of their constituent regions (see Fig. 3.2 (b) in Chapter 3). As in Fig. B8 (bottom), the difference between the distributions corresponding to two subdivisions is measured using the Jensen-Shannon divergence.

Next, we focus on how regions belonging to specific categories connect to each other. In Chapter 3, we mention that within each module, the provincial hubs (R5) connect with each other significantly more often than expected by chance. This intra-modular connectivity between the R5 nodes can be clearly seen from Fig. B10, where these nodes are highlighted and their colors indicate the modules to which they belong (see color key at the bottom). Note that one of the projections shown here is identical to Fig. 3.2 (d) in Chapter 3.

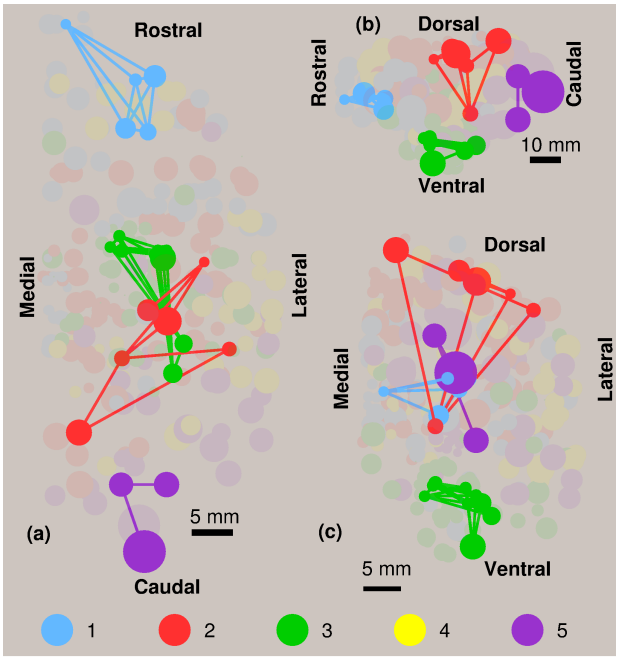


Figure B10: The network of brain regions in (a) horizontal, (b) sagittal and (c) coronal projections, showing that connections between provincial hubs (highlighted nodes) are localized within each module [Figure 3.2 (d) in Chapter 3 is identical to panel (b) above].

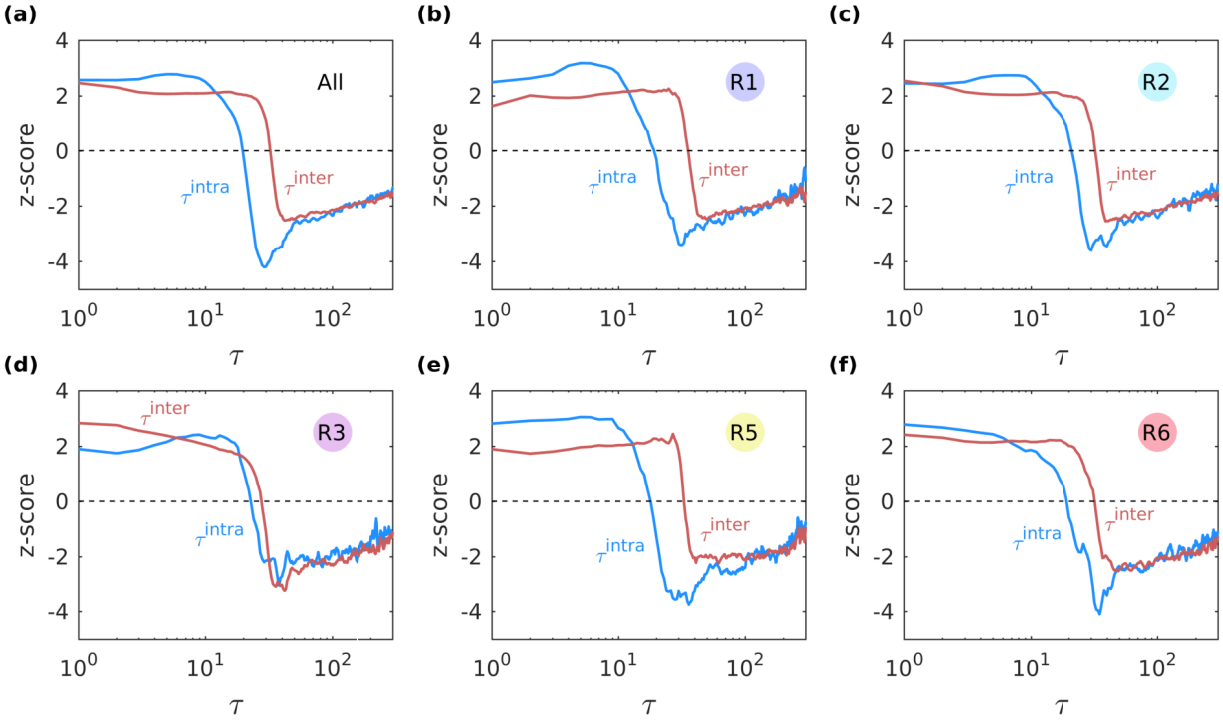


Figure B11: To see how the different categories R1-R7 of brain regions allow spreading to occur faster in the empirical brain network than in equivalent randomized networks, we compare the case where the source node can belong to any category (a) with those where the source is either ultra-peripheral R1 (b), peripheral R2 (c), satellite connector R3 (d), provincial hub R5 (e), or global hub R6 (f). The z-score indicates that there is a statistically significant shift in the empirical distribution towards lower values of τ in all cases. However, while for R3 the increase in the rate of spreading is similar irrespective of whether the target is in the same module or in a different one, we observe that there is a relatively larger shift at lower values for τ^{intra} as compared to τ^{inter} for most of the other categories (in particular, R1 and R5). Indeed, the latter behavior dominates when we consider sources across all categories [see panel (a)]. Note that panels (d) and (e) are identical to see Fig. 3.2 (g-h) in Chapter 3.

In Chapter 3 we have described our investigation of the role played by regions belonging to different categories R1-R7 in facilitating information transmission. For this we simulate diffusive propagation within and between modules and obtain the distribution of first passage times for random walks between a source node and a target node. Fig. B11 (a) shows that the rate of diffusion in the connectome is enhanced both within a module and between modules (as indicated by the statistically significant shift - measured in terms of z-score - in the empirical distributions for both τ^{intra} and τ^{inter} towards lower values) as compared to that seen in equivalent randomized networks.

Fig. B11 (b-f) show how nodes belonging to categories R1, R2, R3, R5 and R6 (respectively), which have distinct intra- and inter-modular connectivity roles, contribute to enhancing communication in the connectome. In each case the source node belongs to the respective category and we quantify the difference in the distributions of both τ^{intra} and τ^{inter} from that obtained from randomized surrogates. We observe that for source nodes of most categories, with the exception of satellite connectors R3, the increase in the rate of diffusion within a module, compared to that in the surrogate networks, is even higher than the increase in the rate of diffusion across modules.

Spatial dependence of connectivity and modular organization

In Chapter 3, we have stated that modular organization of the connectome is not primarily driven by constraints imposed by the physical distance between the brain regions. This is established by using three classes of surrogate random network ensembles to investigate how spatial embedding affects the modular decomposition of a network, with all the regions occupying the same positions in physical space as in the Macaque connectome. The three ensembles we have chosen for our investigation are specified by the dependence of the connection probability P between regions on the physical distance d between them, viz., (i) $P \sim d^0$, i.e., independent of the distance, (ii) $P \sim 1/d$, i.e., power-law dependence

as in the empirical network, and (iii) $P \sim \exp(-d)$, i.e., exponential dependence, for which the constraint of distance most strongly affects the probability of connection. For each category, we have generated 100 different networks that have identical numbers of nodes and links as the empirical connectome. Subsequently, we subject these networks to community detection techniques using information about the connection topology alone, as well as space-independent modular decomposition which explicitly accounts for the dependence of P on d (see Chapter 3 for details).

Fig. B12 shows how the modular nature of the networks belonging to each of the three ensembles described above vary upon two approaches for identifying the modules, viz., (i) using the topological information about the connections alone, and (ii) employing a space-independent partitioning that takes into account the dependence of the probability of connections between regions on the physical distance between them. The similarity between the modules obtained using these two methods is measured using normalized mutual information I_{norm} (see Methods in Chapter 3). Note that, if the two types of partitionings yield identical modules then $I_{norm} = 1$, while $I_{norm} = 0$ implies maximal dissimilarity. Without any spatial dependence, the identified modules arise through fluctuations alone, and hence the similarity between the partitions obtained by the two methods will be entirely stochastic in nature, resulting in the broad distribution for I_{norm} seen in panel (a). In contrast, the ensemble underlying the distribution shown in panel (b) has an inverse relation between connection probability and physical distance, as in the empirical network. The value of I_{norm} obtained for the empirical network (indicated by the arrow) is seen to be significantly larger than those for the random ensemble. This suggests that had the modules arisen exclusively from a distance-dependent constraint on connections, the topological and space-independent approaches would have yielded highly dissimilar partitionings. Qualitatively similar results are obtained when the dependence of connection probability on physical distance is even stronger, viz., P decaying exponentially with d as in the case of the ensemble whose I_{norm} distribution is shown in panel (c). The fact that partitioning the empirical network using either the topological or the space-independent approach results in relatively similar modular decompositions suggest that constraints other than those related

to physical distance plays a significant role in shaping the mesoscopic organization of the Macaque connectome. The results described above are supported by the corresponding distributions of the modularity Q measured for the different partitionings obtained using each of the two approaches (broken and solid curves in panels d-f). Thus, in the absence of any spatial dependence, the distributions of Q obtained using the topological and the space-independent approaches completely overlap [as seen in (d)]. When $P \sim d^{-1}$, the relatively weak spatial dependence gives rise to marginally lower values of Q for the partitionings obtained using the space-independent method, as compared to those obtained using the topological information alone. This is seen to be true for both the empirical network (broken and solid arrows) and the random ensembles [panel (e)]. With the stronger spatial dependence inherent in an exponentially decaying functional relation, we expect to see much larger differences in the Q values for the two types of partitionings, and this is indeed observed in the distributions shown in panel (f). Therefore, the more dominant the role of the constraint on physical distance in determining the connections, the more dissimilar the partitionings obtained by the two methods and the larger the difference in the corresponding Q values.

Fig. B13 illustrates the space independent modular organization of random networks with the three different types of spatial constraints as described above, employing the representation used in Fig. 3.3, Chapter 3. The distributions of the physical distances d between the nodes and the nature of variation of the connection probability between nodes with d are shown in the panels (a) for networks the role of spatial constraint on connectivity is (left) absent, viz., $P \sim d^0$, (center) weak, viz., $P \sim 1/d$, and (right) strong, viz., $P \sim \exp(-d)$. Comparison of the modules obtained using the information about the connection topology alone and those determined using the space-independent method (shown in the panels (b) for each of the networks) indicate that in the absence of any dependence of P on d (left) the partitions overlap to a large extent. Introducing a role for the spatial constraint in determining the connections result in the two types of partitionings differing substantially. This is seen for the power-law dependence of P on d (center), but most prominently when P decays exponentially with d (right). For the latter case, a single

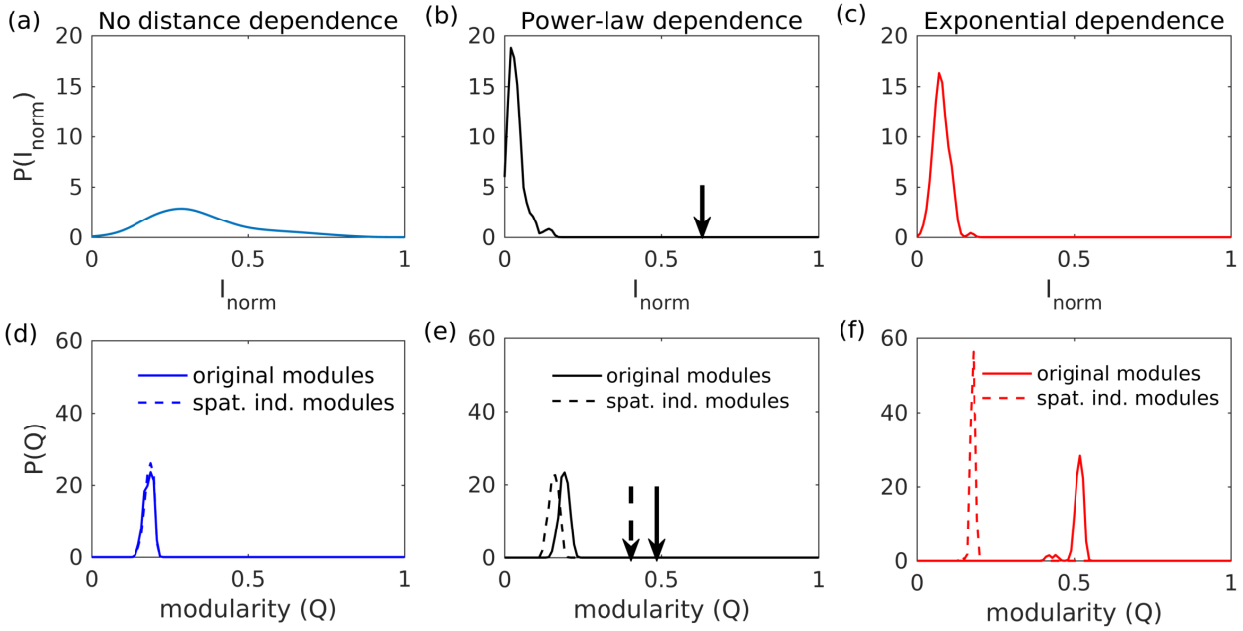


Figure B12: The distributions of (top row) the degree of similarity between the topological and space-independent modular partitionings of a network as measured by normalized mutual information I_{norm} between them, and (bottom row) the corresponding values for the modularity Q obtained using the two methods, for three types of random surrogate network ensembles. These are distinguished by the dependence of connection probability P between a pair of brain regions on the physical distance d between them, viz., $P \sim d^0$ [(a) and (d)], $P \sim 1/d$ [(b) and (e)] and $P \sim \exp(-d)$ [(c) and (f)]. As the Macaque connectome we have investigated also exhibits a power-law dependence, viz., $P \sim 1/d$, similar to that examined in (b) and (e), we have indicated in those panels the corresponding values for the empirical network (arrows).

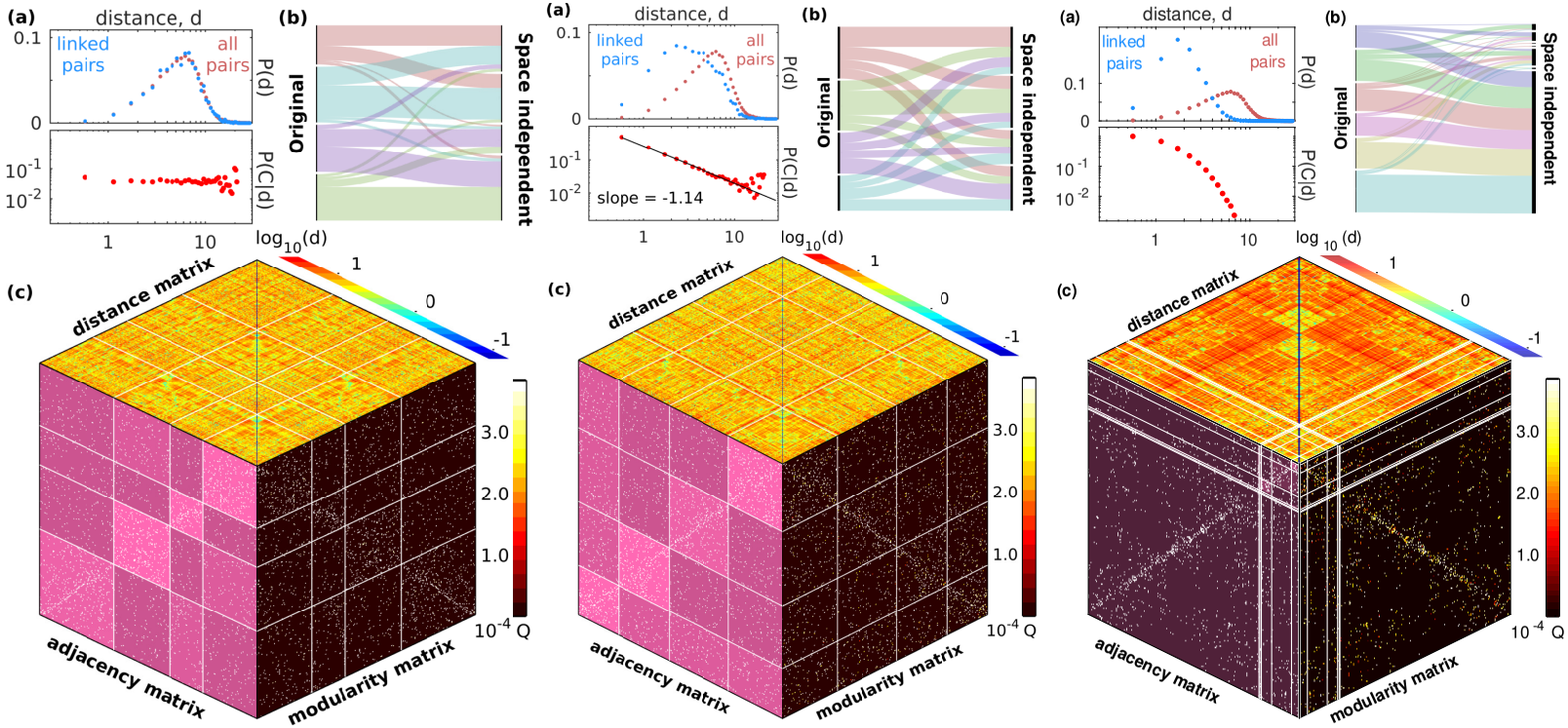


Figure B13: The dependence of the modular organization of a network on the three-dimensional spatial arrangement of the nodes shown for three different functional relations between the physical distance d between a pair of nodes and the connection probability P between them, viz., (left) $P \sim d^0$, (center) $P \sim 1/d$, and (right) $P \sim \exp(-d)$. For each functional relation, we show (a, top) the probability distribution of the physical distances d between all pairs of nodes (red) contrasted with that of the connected pairs (blue), as well as (a, bottom) the variation of the connection probability $P(C|d)$ with physical distance d which characterizes the functional relation (the best fit power-law is indicated with a line, along with its slope, in the central panel). The panels also contain (b) visual representations of the correspondence between the network modules determined using exclusively information about the connection topology ("Original") and those obtained from space-independent partitioning of the network into communities. on the right). This alluvial diagram has been created using the online visualization tool RAW [177]. Joint representation of the space-independent modular organization of each class of network is shown in (c) using three matrices indicating adjacency $\{A_{ij}\}$ (left surface), modularity $\{B_{ij}\}$ (normalized by total number of links L, right surface) and physical distance $\{d_{ij}\}$ (top surface). As in Fig. 3.3 (b) of Chapter 3, the nodes are grouped into partitions corresponding to the space-independent modules of the network with the boundaries indicated by solid lines.

large module is seen to encompass the bulk of the network. As in this case the topological modules arise primarily from the spatial constraint on connections between nodes, on taking this dependence on d into account in the space-independent method the mesoscopic structure becomes relatively homogeneous. The panels (c) show the joint representation of the adjacency, modularity and physical distance matrices for each of the three networks [as per the convention used in Fig. 3.3 (b) of Chapter 3]. The partitions obtained by using the space-independent method are indicated by bounding lines in each matrix. Note that when we take into account the constraint that physical distance places on connectivity between regions, the partitioning results in modules that exhibit only a marginal higher density of connections within them (compared to the overall connection density). This is expected as the modules observed in the topological arrangement of connections in these random networks arise exclusively from the spatial constraint, and therefore the space-independent method should render the networks relatively homogeneous. Thus, the observation of non-trivial modules in the empirical network upon partitioning it with the space-independent method suggests that the observed mesoscopic organization of the Macaque connectome cannot be explained exclusively by the spatial layout of the regions.

Appendix C

Comparing the performance of community detection algorithms on benchmark networks

Although the modules that we obtain using Newman Spectral analysis are robust (as demonstrated by the results from stochastic simulated annealing described in the main text), it remains an open question as to which community detection algorithm would be ideally suited for identifying the modular composition of a given network. A recent study on the efficiency of a range of community detection algorithms on benchmark modular networks [294, 295] has suggested that the *Infomap* algorithm [168], which uses an information-theoretic approach, is most accurate in identifying the community structure. However, we find that the Newman Spectral algorithm is better suited for the type of the empirical network that we consider here. To demonstrate this, we compare the outputs obtained using the two algorithms on an ensemble of benchmark modular networks characterized by a range of degree distributions and variability in module sizes that we have deviced as discussed below.

Generating benchmark networks

In order to systematically examine the comparative efficiency of the community detection methods, we measure their performances for a set of benchmark networks of specified modular structure. In addition to examining the effect of increasing the number of modules

in a network of fixed size, we investigate the role of module size heterogeneity on the outcome. To this end, we develop an algorithm that generates a modular network, where the spread in module sizes can be tuned through a single parameter. The following procedure creates a directional network of size N, with m modules of average size n, whose sizes are drawn from a normal distribution with standard deviation σ (or coefficient of variation $CV = \sigma/n$), where the average number of connections per node is $\langle k \rangle$ and the ratio of inter- to intra-modular connectivity is r:

1. We consider m boxes, each of initial size n, select two boxes at random, and increment the size of one of them by 1 and correspondingly decrease that of the other by 1.

2. We next pick two boxes at random again and repeat the procedure in step [1] for a total of T steps. Note that this essentially corresponds to a random walk in the size of each of the m boxes, and that at every step a pair of boxes exhibits random walks in opposite directions. Thus we effectively have $m/2$ non-concurrent random walks occurring over T steps. From the central limit theorem, we know that the position of a random walker after T steps of size 1 is governed by a normal distribution with variance σ^2, i.e. $T = \sigma^2$. The variance of box sizes as a result of our procedure is thus $\sigma^2 = T/(m/2)$, or in other words we repeat step [1] $T = \sigma^2\, m/2$ times to obtain a set of boxes whose sizes follow a normal distribution with standard deviation σ.

Having finalized the box sizes, we next connect the nodes so as to obtain a network with the desired modularity. We generate benchmark networks with three different types of degree distributions, namely Poisson, exponential and power law distributions, respectively. To generate networks with Poisson degree distributions, we connect nodes in a manner similar to the generation of Erdős-Rényi random graphs (which is also characterized by a Poisson degree distribution), with the constraint that the probability of connection of nodes within modules is different to that across modules. Specifically, if sizes of box i are given by n_i, then the intra and inter-modular connection densities are given by:

$$\rho_{\text{in}} = \frac{N \langle k \rangle}{\Sigma_i n_i \,(n_i - 1) + \Sigma_i r\, n_i \,(N - n_i)}\,, \quad \rho_{\text{out}} = r\, \rho_{\text{in}}\,.$$

Note that in the limit $r \to 1$ this procedure yields an Erdős-Rényi random network with connection density $\rho = \langle k \rangle / (N-1)$.

For the case of networks with exponential and power-law degree distributions, we assign an in-degree and out-degree to every node in the network through a process of repeatedly drawing a random variable from the desired probability distribution. Note that in principle we can generate the required degree sequence from any arbitrary degree distribution, by using the following procedure.

Generating a random variable from a desired probability function

Consider a probability distribution function $p(k)$ such that

$$\int_{k_{\min}}^{k_{\max}} p(k) \cdot dk = 1 \,, \tag{C1}$$

As we require a finite network of size N, the natural bounds $k_{\min}$ and $k_{\max}$ of this degree distribution are 1 and N, respectively. We vary $k_{\min}$ and $k_{\max}$ within these bounds to obtain the desired average degree $\langle k \rangle$. For the given distribution $p(k)$, we define the cumulative distribution function (CDF):

$$F(k) = \int_{k_{\min}}^{k} p(\kappa) \cdot d\kappa = Y \,, \tag{C2}$$

where $0 \leqslant Y \leqslant 1$. In a discrete system where the random variables represent the degree sequence of the network, the CDF can also be considered as the rank order for the degree sequence arranged in ascending order. Hence for any given degree k, we obtain its rank order in Y. Now, for any degree distribution, the rank is always distributed uniformly, since it is just the order of occurrence in an arrangement. Thus in order to generate the random variable k, we first generate Y as a uniform random variable between 0 and 1, and then invert the function $F(k)$.

$$F(k) = Y \in U(0,1) \,, \tag{C3}$$

$$k = F^{-1}(Y)\,, \tag{C4}$$

Following this procedure, we can obtain a degree sequence $S = \{k_1, k_2, \ldots k_N\}$ from any given probability distribution.

Generating an in-degree and out-degree sequence with desired correlation

Consider the uncorrelated sequences S_1, S_2 and S_3 that were randomly generated from the a probability distribution. We can obtain two correlated sequences $S_{\rm in}$ and $S_{\rm out}$ for the in-degree and out-degree of the network, respectively, from the above three sequences with a mixing parameter $m(\in [0,1])$ in the following way:

$$S_{in} = m \cdot S_1 + (1-m) \cdot S_2\,, \tag{C5}$$

$$S_{out} = m \cdot S_1 + (1-m) \cdot S_3\,, \tag{C6}$$

Now, calculating the correlation between S_{in} and S_{out} gives:

$$corr(S_{in}, S_{out}) = c = \frac{m^2}{2m^2 - 2m + 1}\,, \tag{C7}$$

Hence in order to generate two degree sequences with a desired correlation c, we generate three uncorrelated sequences and mix them in the above given way using mixing parameter m given by:

$$m = \frac{1}{1 + \sqrt{\frac{1}{c} - 1}}\,. \tag{C8}$$

Community detection using *Infomap*

Upon generating the benchmark networks, we compare the modules identified using Newman Spectral Analysis with those using the *Infomap* algorithm [168]. This technique operates on the principle of optimally compressing information about a dynamic process (viz. a random walk) on a given graph, such that the whole process can be recovered as accurately as possible upon decoding the information. The optimization is

done using a combination of greedy search and simulated annealing, to yield the underlying community structure. The code for *Infomap* that we use is freely available from `http://www.tp.umu.se/~rosvall/code.html`.

Comparison of the robustness of the *Infomap* and Newman Spectral algorithms

A comparison of the results obtained upon using the *Infomap* and Newman Spectral algorithms for detecting community structure is shown in Fig. C1. While both approaches identify the existence of modules, the *Infomap* algorithm yields a much larger number of modules than the Spectral method. As discussed in Appendix B, the modules obtained using *Infomap* are essentially a finer decomposition of those obtained using the Spectral method. However, a comparison of the accuracy of the two methods on generated modular benchmark networks with Poisson degree distributions shows that the Spectral method is accurate over a broader range of the modularity parameter r. Fig. C1(c-f), displays the probability distributions of the normalized mutual information $I_{\rm norm}$ of the modular structure obtained from the method with the embedded modular structure, over a range of values of r. As can be seen in all cases, the performance of *Infomap* drops close to $r = 0.1$, whereas the Newman Spectral method is more robust even for higher r values, regardless of the choice of m or CV for the benchmark modular network.

Comparison of results obtained for modular benchmark networks with different degree distributions

In order to confirm that the results discussed above are independent of the degree distribution of the modular benchmark network, we compare results obtained using networks with Poisson, exponential and power-law degree distributions, with different choices of average degree $\langle k \rangle$. As can be seen in Fig. C2, the two algorithms yield similar results for the normalized mutual information $I_{\rm norm}$, regardless of degree distribution. Furthermore, for

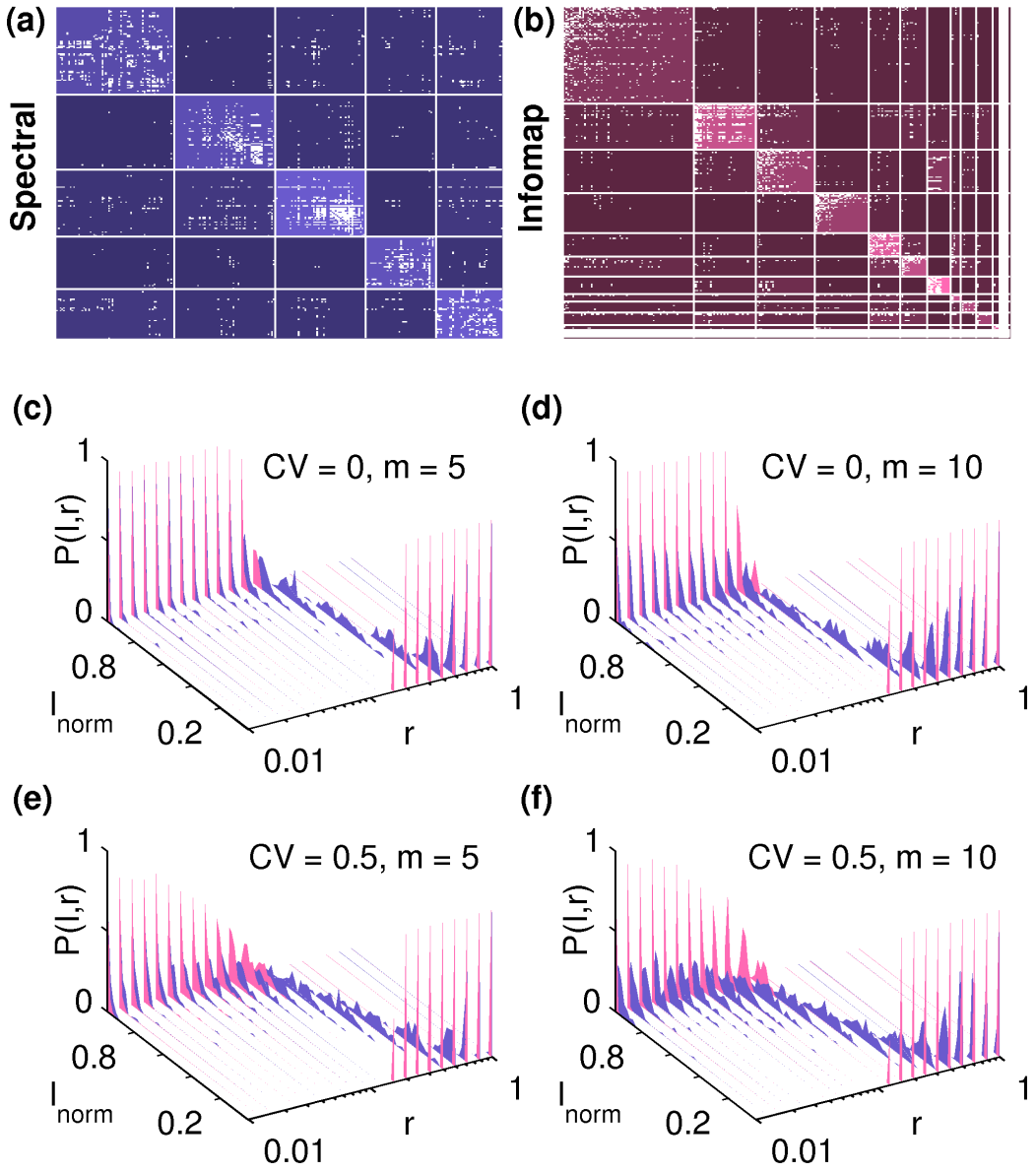


Figure C1: Robustness of the detected community structure of the brain network. (a-b) Adjacency matrices of the macaque brain network, with nodes arranged according to the community partitioning obtained, using (a) Newman Spectral Analysis and (b) Infomap, respectively. The brightness of each block is proportional to the density of connections within it, and the white dots in the diagonal and off-diagonal blocks represent intra- and inter-community links, respectively. (c-f) Probability distributions of the normalized mutual information I_{norm}, which measures the performance of each algorithm (purple: Newman Spectral Analysis, pink: Infomap) in accurately detecting the community structure, on ensembles of benchmark modular networks with Poisson degree distributions. The networks are of size $N = 400$, each comprising m communities (c, e: $m = 5$, d, f: $m = 10$) whose size distribution has a coefficient of variation CV (c, d: $CV = 0$, e, f: $CV = 0.5$) and average degree $\langle k \rangle = 10$. These distributions are obtained from 100 realizations, and are displayed over a range of values of inter- to intra-community connectivity ratio r, which we vary from 0.01 (highly modular) to 1 (homogenous random network).

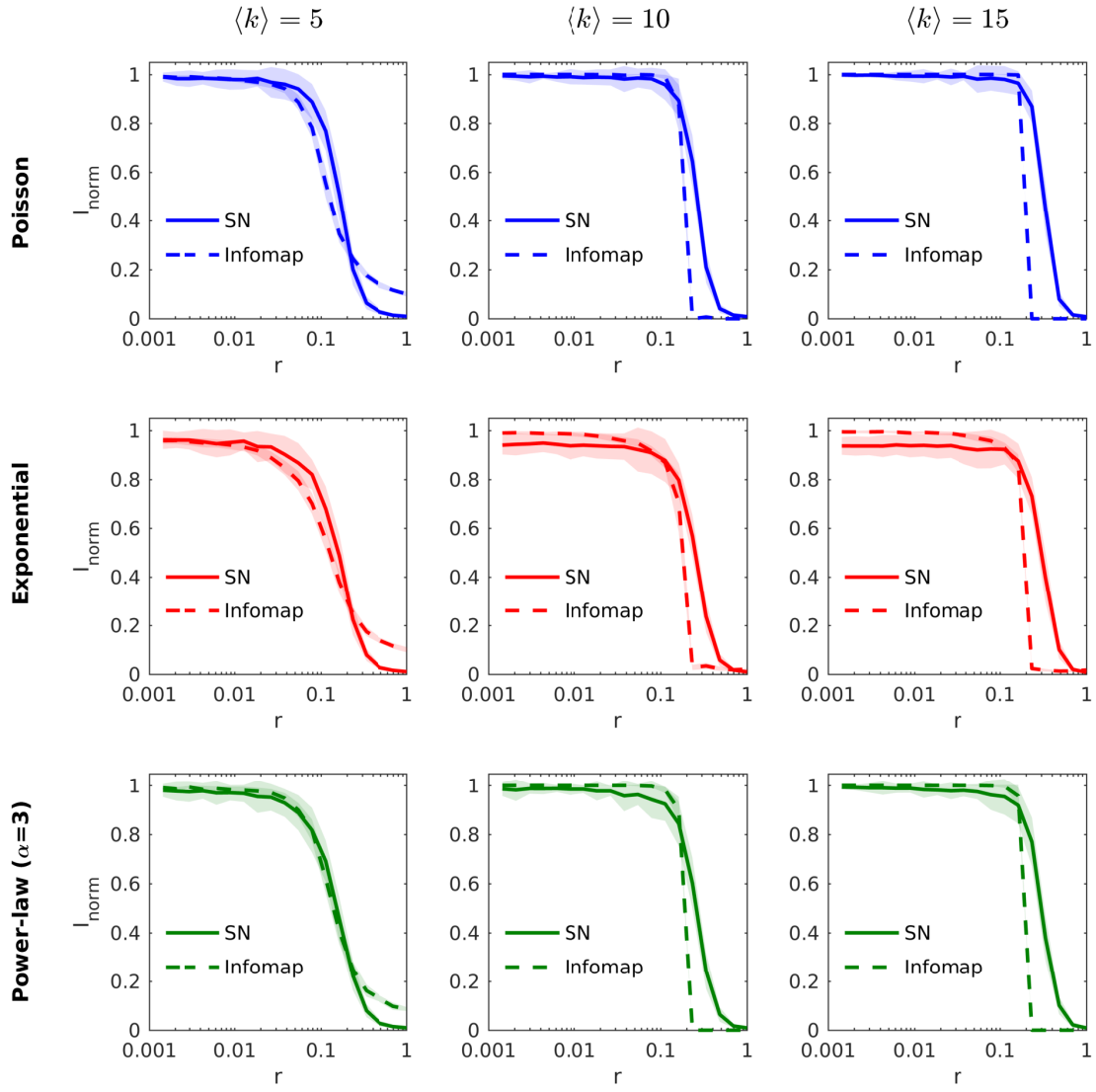


Figure C2: Comparison of the accuracy of the community detection algorithms for different modular benchmark networks using normalized mutual information. Normalized mutual information of the communities obtained using a community detection algorithm with those of a benchmark modular network of size $N = 400$, containing $m = 5$ modules and with average degree $\langle k \rangle$. Results are shown for benchmark networks with three different degree distributions: Poisson, exponential and power-law, for the cases where the module sizes are identical ($CV = 0$). For the case of the exponential distribution, the exponents vary with average degree (0.2 for $\langle k \rangle = 5$, 0.1 for $\langle k \rangle = 10$ and 0.065 for $\langle k \rangle = 15$), while for the power-law distribution, we use an exponent of $\alpha = 3$ in each case. The solid and dashed lines represent the average values of I_{norm} obtained using the Newman Spectral and Infomap algorithms, respectively, over an ensemble of 100 trials, while the shaded regions represents the standard deviation of these quantities. Results are displayed over a range of values of the modularity parameter r.

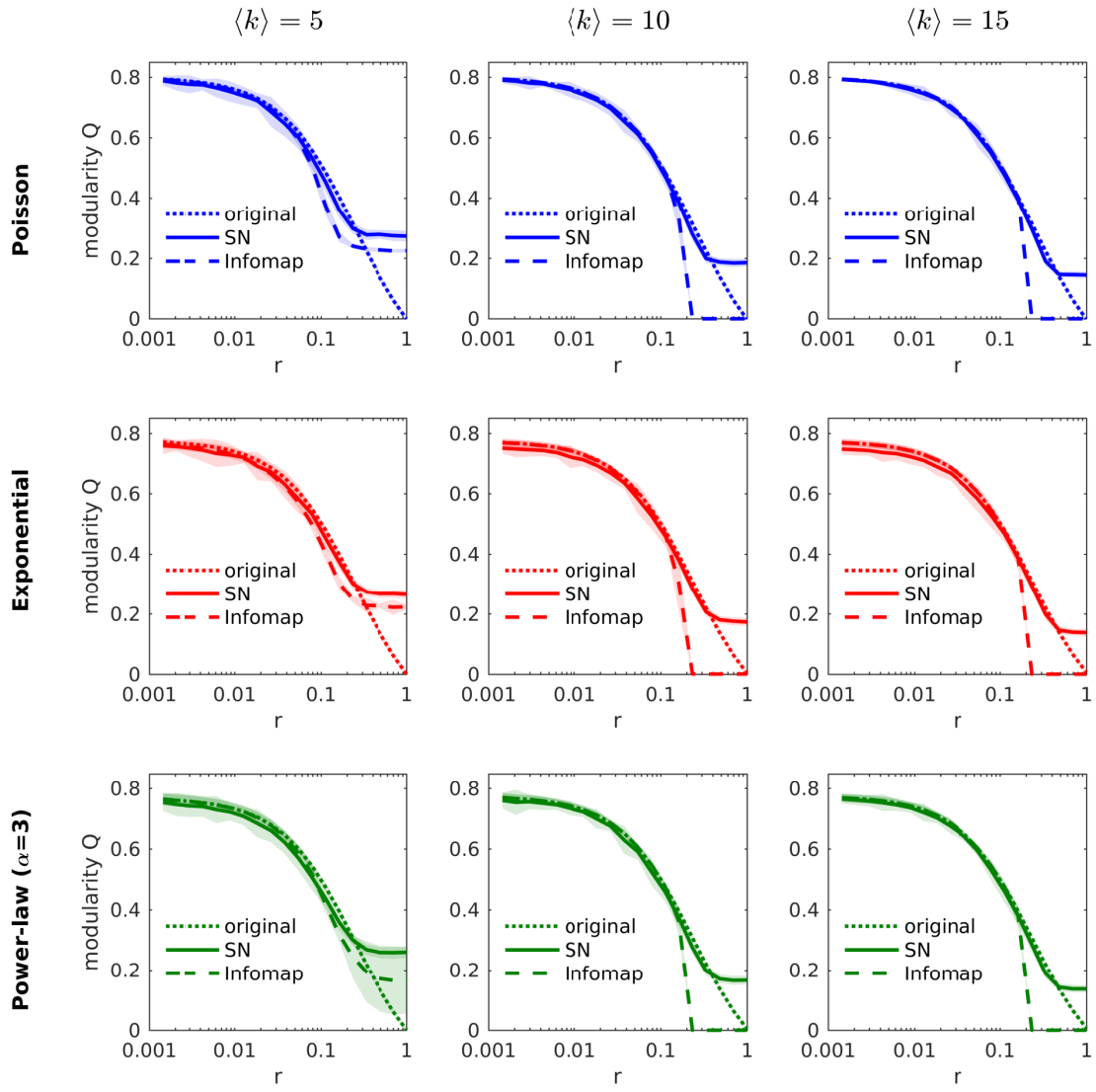


Figure C3: Comparison of the accuracy of the community detection algorithms for different modular benchmark networks using the modularity parameter Q. Modularity parameter Q of the communities obtained using a community detection algorithm with those of a benchmark modular network of size $N = 400$, containing $m = 5$ modules and with average degree $\langle k \rangle$. Results are shown for benchmark networks with three different degree distributions: Poisson, exponential and power-law, for the cases where the module sizes are identical ($CV = 0$). For the case of the exponential distribution, the exponents vary with average degree (0.2 for $\langle k \rangle = 5$, 0.1 for $\langle k \rangle = 10$ and 0.065 for $\langle k \rangle = 15$), while for the power-law distribution, we use an exponent of $\alpha = 3$ in each case. The dotted, solid and dashed lines represent the average values of Q of the original benchmark networks and those obtained using the Newman Spectral and Infomap algorithms, respectively, over an ensemble of 100 trials, while the shaded regions represents the standard deviation of these quantities. Results are displayed over a range of values of the modularity parameter r.

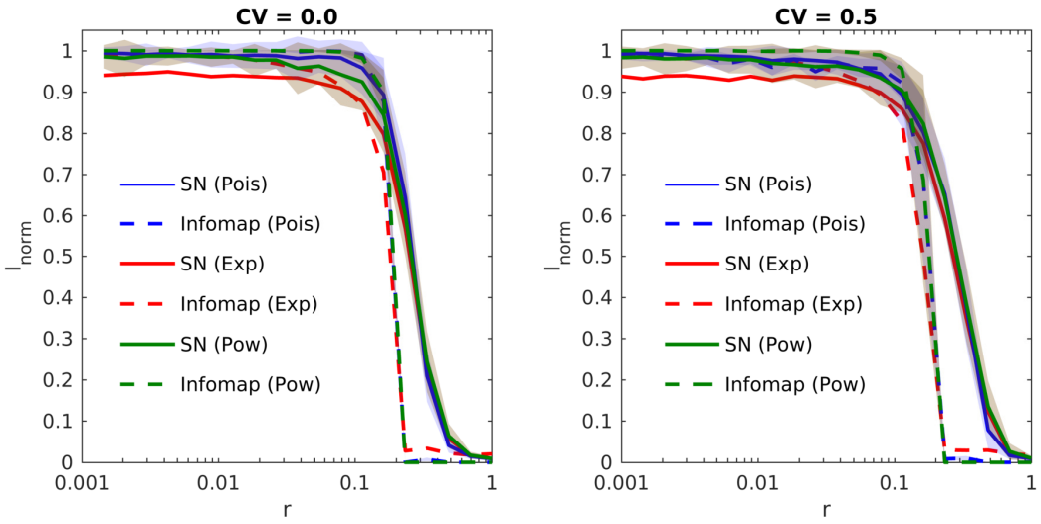


Figure C4: Comparison of results obtained for modular benchmark networks with different values of CV Normalized mutual information of the communities obtained using a community detection algorithm with those of a benchmark modular network of size $N = 400$, containing $m = 5$ modules and with average degree $\langle k \rangle = 10$. Results are shown for networks with three different types of degree distributions: Poisson, exponential and power-law, for the cases where the module sizes are identical (right, $CV = 0$) and variable (left, $CV = 0.5$). For the case of the exponential distribution, we use an exponent of 0.1, while for the power-law distribution, we use an exponent of $\alpha = 3$. The solid and dashed lines represent the average values of I_{norm} obtained using the Newman Spectral and Infomap algorithms, respectively, over an ensemble of 100 trials, while the shaded regions represents the standard deviation of these quantities. Results are displayed over a range of values of the modularity parameter r.

average degree $\langle k \rangle = 10$ and higher, the accuracy of the *Infomap* algorithm drops sharply beyond a critical value of r.

In addition, in Fig. C3 we display a comparison of the modularity parameter Q for the communities obtained in the case displayed in Fig. C2. Note that we expect Q to gradually drop to zero as $r \to 1$, corresponding to the network becoming less modular. We note that the Q values of Infomap drops sharply around the same values as for the case of I_{norm} (for $\langle k \rangle \geq 10$), which indicates that the method is not applicable beyond a critical value of r. However, we also observe that the value of Q does not drop to 0 for the case of Newman Spectral algorithm, even in the limit $r \to 1$. This deviation from the expected trend suggests that the Newman Spectral algorithm is not applicable for networks with high values of r. Finally, in Fig. C4 we compare results obtained for the case where the

module sizes are identical ($CV = 0$) and variable ($CV = 0.5$) for the three choices of degree distributions of benchmark modular networks. We find that the results are similar in both cases, with a slight shift in the values of I_{norm} in the two cases, suggesting that the results are robust with respect to variability in module size.

Hence, we find that while the Infomap algorithm can accurately identify the community structure of highly modular networks, it quickly breaks down beyond a critical value of r for networks with average degree $\langle k \rangle \geq 10$. In contrast, the Newman Spectral method can identify modules even for higher values of r, and only breaks down in the limit $r \to 1$. This provides justification for our choice of the Newman Spectral method for the detection of community structure in the macaque network (which has an average degree of $\langle k \rangle = 9.782$).

Appendix D

The following figures comprise the supporting information for the analysis of human structural connectome, as described in Chapter 4.

Structure-function correlation for each link

In Fig. D1, we display the correspondence between structure and function for every link (i, j) belonging to the "representative" brain network over the population of individuals (see Chapter 4). We observe that the link-wise correlation $Corr(W_{ij}, C_{ij})$ between the structural connection weights W_{ij} and corresponding weights in functional connectivities C_{ij}, calculated across all individuals, tends to be extremely low for most of the links. This suggests that the functional connectivity has a negligible dependence on the structural connectivity when observed at the level of nodes and links. This is in contrast with the macro-level picture shown in Fig. 4.4 of Chapter 4, where we compared the structural and functional connectivities across entire networks, and found a comparatively higher and statistically significant correspondence between the two.

Community structure in the human structural connectome

We have analyzed two types structural connectivity: the first corresponding to original connection weights, as given in the database, and the second consisting of "rescaled" connection weights (for details, see Methods in Chapter 4). We have found modules in the brain network of each individual by implementing two separate community detection methods which are described in Chapter 3 and also in Appendix C, viz., Newman Spectral

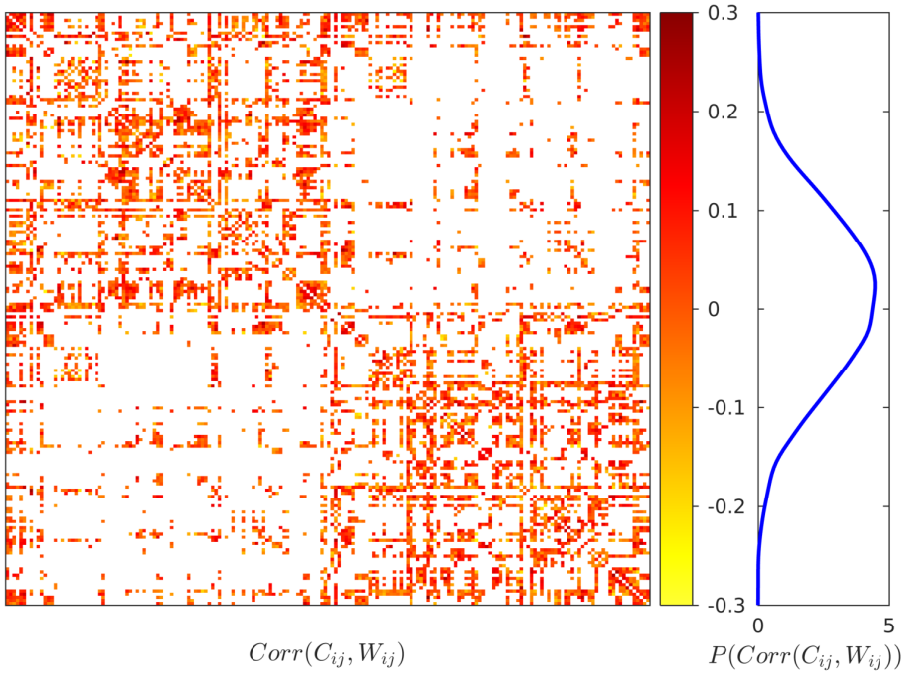


Figure D1: The link-wise correlations between each structural connection and its corresponding functional connection over the population are extremely low. For each link between a pair of regions (i, j) we display the correlation between their structural connection weights W_{ij} and corresponding weights in functional connectivities C_{ij}, calculated across all individuals (left panel). The probability distributions for the correlation values are shown in the right panel. Only those links that are identified as part of the representative network, i.e., whose weights are Poisson distributed over the population, are shown here.

Analysis [9] and the *Infomap Method* [168]. We have included only those links that comprise the "representative" structural network (as described in Chapter 4). Fig. D2 and D3 show the modular decomposition of the original network (Fig. D2) and the rescaled network (Fig. D3) for the same individual, as obtained from Newman Spectral Analysis.

We observe that the modules obtained are spatially contiguous with clearly defined boundaries. There is only a slight variation between the modular partitionings of the two types of the networks shown in Fig. D2 and D3. The similarity of modular partitioning between different individuals is shown in Fig. D4 where we show the normalized mutual information I_{norm} between all pairs of modular partitionings (for details about normalized

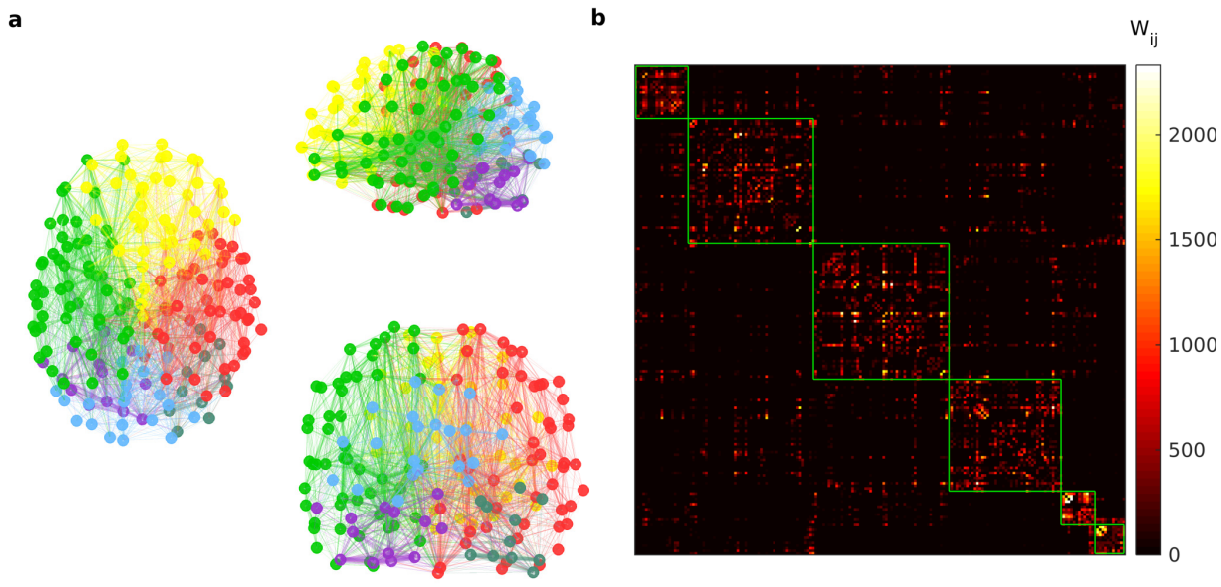


Figure D2: Modules in the structural brain network of an individual subject, obtained using Newman Spectral Analysis. (a) Horizontal, sagittal and coronal projections (left, top right and bottom right, respectively) of the spatial representations for the structural brain network of an individual subject, highlighting the 6 modules obtained from Newman Spectral Analysis. Here, the nodes are colored in accordance with the module to which they belong, and the color of each link corresponds to that of its respective source node, while the thickness of each links is proportional to its connection weight. Only those links that are part of the representative network (as described in Chapter 4) are considered here. (b) Weighted adjacency matrix representing the network shown in panel (a). Here, the nodes are rearranged and grouped according to their modular membership. The matrix elements are colored in accordance with the connection weights W_{ij} of the corresponding links (see legend at the right). The 6 modules that were obtained from the analysis correspond to the diagonal blocks, marked by green lines.

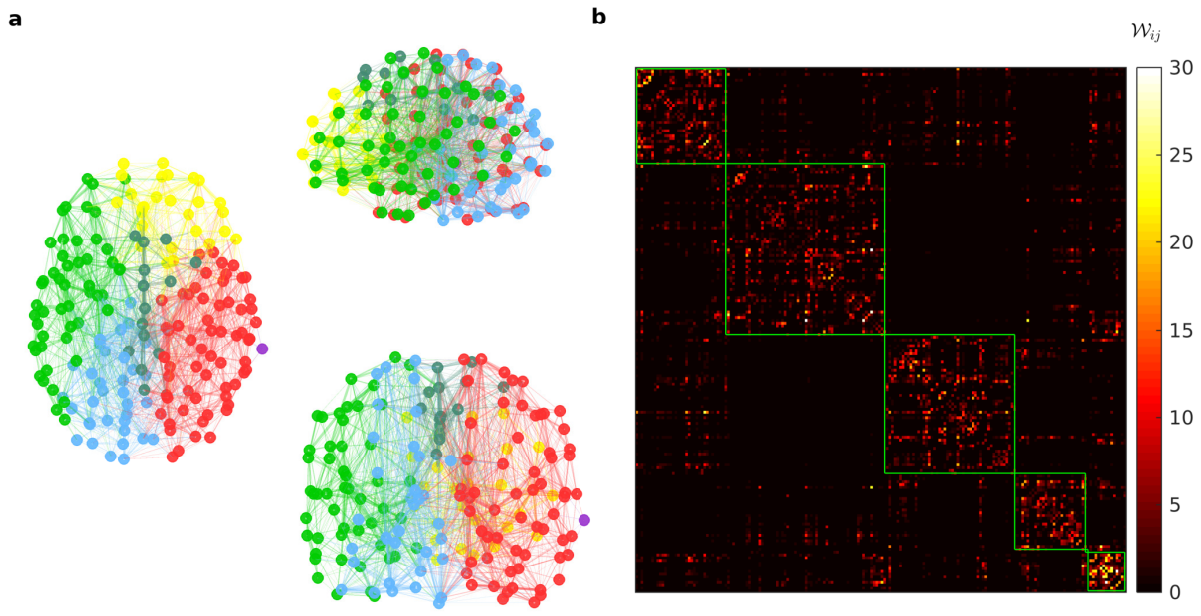


Figure D3: Modules in the rescaled structural brain network of an individual subject, obtained using the Newman Spectral Analysis. (a) Horizontal, sagittal and coronal projections (left, top, right and bottom right, respectively) of the spatial representations for the rescaled structural brain network of an individual subject, highlighting the 6 modules obtained from Newman Spectral Analysis. The individual represented here is the same as that in Fig. D2. Here, the nodes are colored in accordance with the module to which they belong, and the color of each link corresponds to that of its respective source node, while the thickness of each links is proportional to its connection weight. Only those links that are part of the representative network (as described in Chapter 4) are considered here. (b) Weighted adjacency matrix representing the rescaled network shown in panel (a). Here, the nodes are rearranged and grouped according to their modular membership. The matrix elements are colored in accordance with the rescaled connection weights $\mathcal{W}_{ij}$ of the corresponding links (see legend at the right). The 6 modules that were obtained from the analysis correspond to the diagonal blocks, marked by green lines.

mutual information, see Chapter 3). For most pairs of individuals, the I_{norm} values are ≈ 0.5, which indicates that modular partitioning is moderately varying across individuals.

Qualitatively similar results are obtained on using the *Infomap* method to detect modules in original structural network (Fig. D5) and the rescaled structural network (Fig. D6). The networks represented in Fig. D5 and D6 are from the same individual as that in Fig. D2 and D3. Fig. D7 shows that the modules obtained across the individuals using *Infomap* method are relatively more similar to each other, as indicated by higher I_{norm} values, in

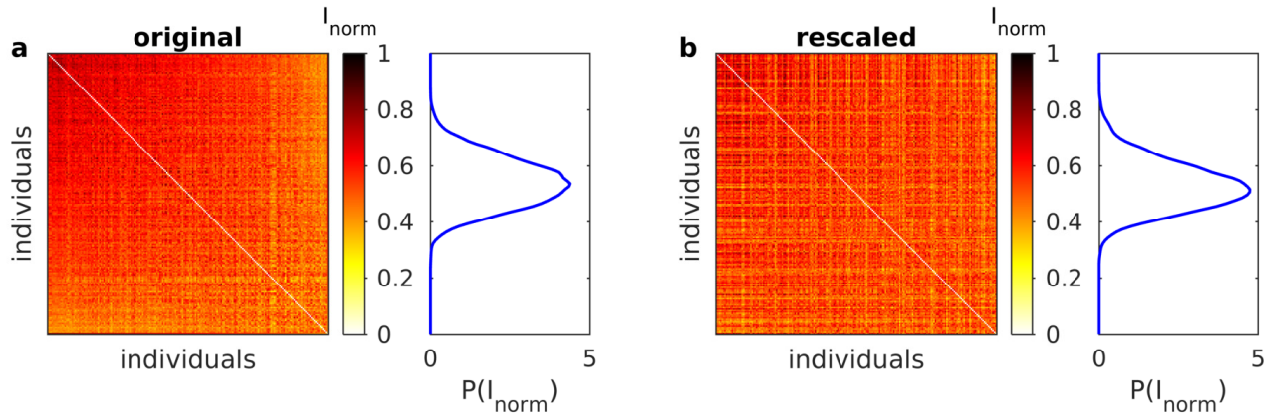


Figure D4: Similarity of Newman Spectral modules across different individual brain networks. Pair-wise values of I_{norm}, which quantifies the similarity of modular paritionings between a pair of individuals, as well as the kernel-smoothened distributions of the corresponding I_{norm} values, are shown for (a) the brain networks having original connection weights, and (b) brain networks having rescaled weights. Note that the mode of the I_{norm} distributions in both cases is ≈ 0.5, indicating that the modular decomposition of individual brain networks varies moderately over the population.

comparison to those obtained using Newman Spectral Analysis.

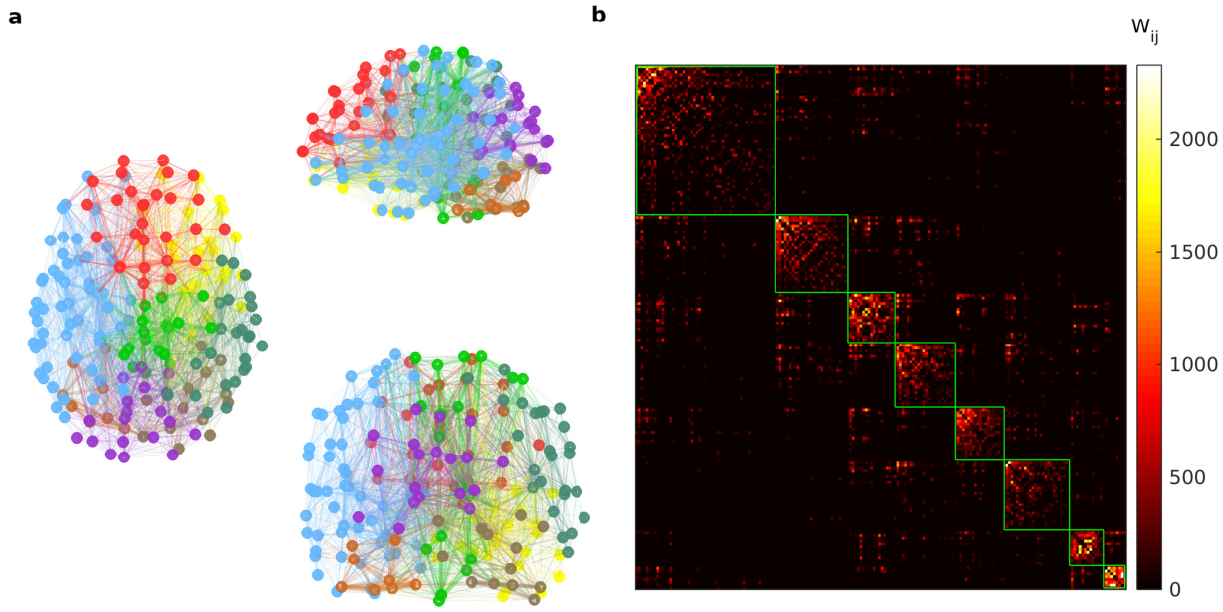


Figure D5: Modules in the structural brain network of an individual subject, obtained using the *Infomap* method. (a) Horizontal, sagittal and coronal projections (left, top, right and bottom right, respectively) of the spatial representations for the structural brain network of an individual subject, highlighting the 8 modules obtained from the *Infomap* method. The individual represented here is the same as that in Fig. D2. Here, the nodes are colored in accordance with the module to which they belong, and the color of each link corresponds to that of its respective source node, while the thickness of each link is proportional to its connection weight. Only those links that are part of the representative network (as described in Chapter 4) are considered here. (b) Weighted adjacency matrix representing the network shown in panel (a). Here, the nodes are rearranged and grouped according to their modular membership. The matrix elements are colored in accordance with the connection weights W_{ij} of the corresponding links (see legend at the right). The 8 modules that were obtained from the analysis correspond to the diagonal blocks, marked by green lines.

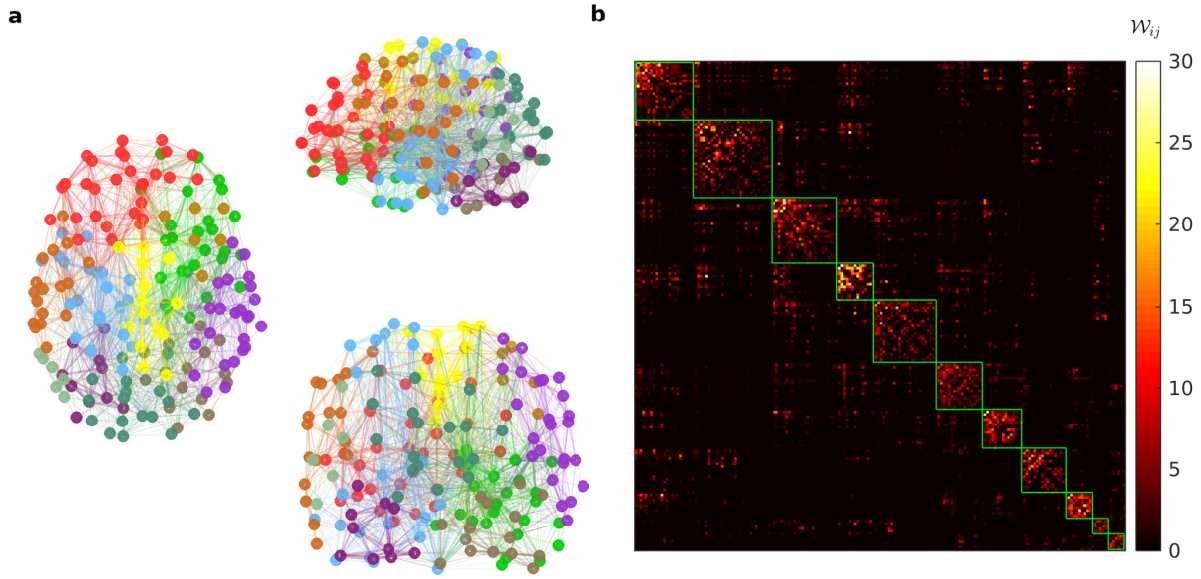


Figure D6: Modules in the rescaled structural brain network of an individual subject, obtained using the *Infomap* method. (a) Horizontal, sagittal and coronal projections (left, top, right and bottom right, respectively) of the spatial representations for the rescaled structural brain network of an individual subject, highlighting the 11 modules obtained from the *Infomap* method. The individual represented here is the same as that in Fig. D2. Here, the nodes are colored in accordance with the module to which they belong, and the color of each link corresponds to that of its respective source node, while the thickness of each link is proportional to its connection weight. Only those links that are part of the representative network (as described in Chapter 4) are considered here. (b) Weighted adjacency matrix representing the rescaled network shown in panel (a). Here, the nodes are rearranged and grouped according to their modular membership. The matrix elements are colored in accordance with the rescaled connection weights $\mathcal{W}_{ij}$ of the corresponding links (see legend at the right). The 11 modules that were obtained from the analysis correspond to the diagonal blocks, marked by green lines.

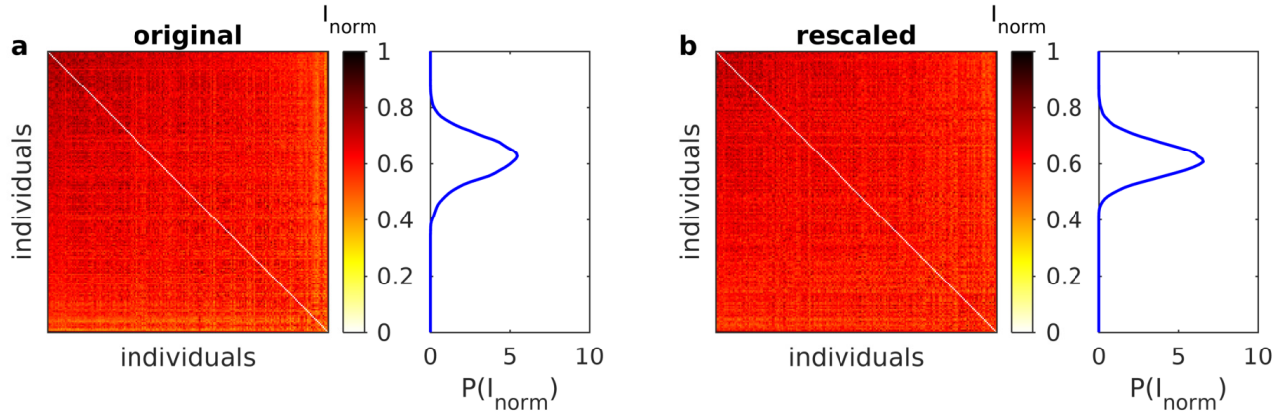


Figure D7: Similarity of *Infomap* modules across different individual brain networks. Pair-wise values of I_{norm}, which quantifies the similarity of modular paritionings between a pair of individuals, as well as the kernel-smoothened distributions of the corresponding I_{norm} values, are shown for (a) the brain networks having original connection weights, and (b) brain networks having rescaled weights. Note that the mode of the I_{norm} distributions in both cases is ≈ 0.6, indicating that the modular decomposition of individual brain networks varies moderately over the population.

Bibliography

[1] Philip W Anderson. More is different. *Science*, 177(4047):393–396, 1972. doi:10.1126/science.177.4047.393.

[2] James Sethna et al. *Statistical mechanics: entropy, order parameters, and complexity*, volume 14. Oxford University Press, 2006.

[3] Ed Bullmore and Olaf Sporns. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186, 2009. doi:10.1038/nrn2575.

[4] Olaf Sporns. *Networks of the Brain*. MIT press, 2010.

[5] Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nature neuroscience*, 20 (3):353–364, 2017. doi:10.1038/nn.4502.

[6] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298 (5594):824–827, 2002. doi:10.1126/science.298.5594.824.

[7] Anna D Broido and Aaron Clauset. Scale-free networks are rare. *Nat Commun*, 10 (1):1–10, 2019. doi:10.1038/s41467-019-08746-5.

[8] Danielle S Bassett and Edward T Bullmore. Small-world brain networks revisited. *Neuroscientist*, 23(5):499–516, 2017. doi:10.1177/1073858416667720.

[9] Mark E. J. Newman. Modularity and community structure in networks. *Proc Natl Acad Sci USA*, 103(23):8577–8582, 2006. doi:10.1073/pnas.0601602103.

[10] Aaron Clauset, Cristopher Moore, and Mark EJ Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008. doi:10.1038/nature06830.

[11] Jack W. Scannell, Gully A. P. C. Burns, Claus C. Hilgetag, Marc A. O'Neil, and Malcolm P. Young. The connectional organization of the cortico-thalamic system of the cat. *Cereb Cortex*, 9(3):277–299, 1999. doi:10.1093/cercor/9.3.277.

[12] Claus C. Hilgetag, Gully A. P. C. Burns, Marc A. O'Neill, Jack W. Scannell, and Malcolm P. Young. Anatomical connectivity defines the organization of clusters of cortical areas in the macaque and the cat. *Philos Trans R Soc London, Ser B*, 355 (1393):91–110, 2000. doi:10.1098/rstb.2000.0551.

[13] Danielle S Bassett, Daniel L Greenfield, Andreas Meyer-Lindenberg, Daniel R. Weinberger, Simon W. Moore, and Edward T. Bullmore. Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Comput Biol*, 6(4):e1000748, 2010. doi:10.1371/journal.pcbi.1000748.

[14] Raj Kumar Pan, Nivedita Chatterjee, and Sitabhra Sinha. Mesoscopic organization reveals the constraints governing *caenorhabditis elegans* nervous system. *PLoS One*, 5(2):e9240, 02 2010. doi:10.1371/journal.pone.0009240.

[15] Quanxin Wang, Olaf Sporns, and Andreas Burkhalter. Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *J Neurosci*, 32(13):4386–4399, 2012. doi:10.1523/JNEUROSCI.6063-11.2012.

[16] Logan Harriger, Martijn P. Van Den Heuvel, and Olaf Sporns. Rich club organization of macaque cerebral cortex and its role in network communication. *PloS One*, 7(9): e46497, 2012. doi:10.1371/journal.pone.0046497.

[17] Murray Shanahan, Verner P. Bingman, Toru Shimizu, Martin Wild, and Onur Güntürkün. Large-scale network organization in the avian forebrain: a connectivity matrix and theoretical analysis. *Front Comput Neurosci*, 7:89, 2013. doi:10.3389/fncom.2013.00089.

[18] Chi-Tin Shih, Olaf Sporns, Shou-Li Yuan, Ta-Shun Su, Yen-Jen Lin, Chao-Chun Chuang, Ting-Yuan Wang, Chung-Chuang Lo, Ralph J Greenspan, and Ann-Shyn Chiang. Connectomics-based analysis of information flow in the *drosophila* brain. *Curr Biol*, 25(10):1249–1258, 2015. doi:10.1016/j.cub.2015.03.021.

[19] Olaf Sporns and Richard F. Betzel. Modular brain networks. *Annu Rev Psychol*, 67: 613–640, 2016. doi:10.1146/annurev-psych-122414-033634.

[20] Petter Holme. Core-periphery organization of complex networks. *Phys Rev E*, 72(4): 046111, 2005. doi:10.1103/PhysRevE.72.046111.

[21] Nivedita Chatterjee and Sitabhra Sinha. Understanding the mind of a worm: Hierarchical network structure underlying nervous system function in *c. elegans*. *Progress in Brain Research*, 168:145–153, 2007. doi:10.1016/S0079-6123(07)68012-1.

[22] Kelly Shen, Alexandros Goulas, David S Grayson, John Eusebio, Joseph S Gati, Ravi S Menon, Anthony R McIntosh, and Stefan Everling. Exploring the limits of network topology estimation using diffusion-based tractography and tracer studies in the macaque cortex. *NeuroImage*, 191:81–92, 2019. doi:10.1016/j.neuroimage.2019.02.018.

[23] Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns. Mapping the structural core of human cerebral cortex. *PLoS Biol*, 6(7):e159, 2008.

[24] Federico Battiston, Jeremy Guillon, Mario Chavez, Vito Latora, and Fabrizio de Vico Fallani. Multiplex core-periphery organization of the human connectome. *J R Soc Interface*, 15(146):20180514, 2018. doi:10.1098/rsif.2018.0514.

[25] Vittoria Colizza, Alessandro Flammini, M Angeles Serrano, and Alessandro Vespignani. Detecting rich-club ordering in complex networks. *Nat Phys*, 2(2):110–115, 2006. doi:10.1038/nphys209.

[26] Emma K. Towlson, Petra E. Vértes, Sebastian E. Ahnert, William R. Schafer, and Edward T. Bullmore. The rich club of the *c. elegans* neuronal connectome. *Journal of Neuroscience*, 33(15):6380–6387, 2013. ISSN 0270-6474. doi:10.1523/JNEUROSCI.3784-12.2013.

[27] Martijn P Van Den Heuvel and Olaf Sporns. Rich-club organization of the human connectome. *J Neurosci*, 31(44):15775–15786, 2011. doi:10.1523/JNEUROSCI.3539-11.2011.

[28] Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: a structural description of the human brain. *PLoS Comput Biol*, 1(4):e42, 2005. doi:10.1371/journal.pcbi.0010042.

[29] Javier DeFelipe. From the connectome to the synaptome: an epic love story. *Science*, 330(6008):1198–1201, 2010. doi:10.1126/science.1193378.

[30] Sebastian Seung. *Connectome: How the brain's wiring makes us who we are*. HMH, 2012.

[31] Larry W Swanson and Mihail Bota. Foundational model of structural connectivity in the nervous system with a schema for wiring diagrams, connectome, and basic plan architecture. *Proc Natl Acad Sci USA*, 107(48):20610–20617, 2010. doi:10.1073/pnas.1015128107.

[32] Steven E Petersen and Olaf Sporns. Brain networks and cognitive architectures. *Neuron*, 88(1):207–219, 2015. doi:10.1016/j.neuron.2015.09.027.

[33] Bratislav Mišić and Olaf Sporns. From regions to connections and networks: new bridges between brain and behavior. *Curr Opin Neurobiol*, 40:1–7, 2016. doi:10.1016/j.conb.2016.05.003.

[34] Ann M Hermundstad, Kevin S Brown, Danielle S Bassett, and Jean M Carlson. Learning, memory, and the role of neural network architecture. *PLoS Comput Biol*, 7(6):e1002063, 2011. doi:10.1371/journal.pcbi.1002063.

[35] Rishidev Chaudhuri and Ila Fiete. Computational principles of memory. *Nat Neurosci*, 19(3):394, 2016. doi:10.1038/nn.4237.

[36] Kanaka Rajan, Christopher D Harvey, and David W Tank. Recurrent network models of sequence generation and memory. *Neuron*, 90(1):128–142, 2016. doi:10.1016/j.neuron.2016.02.009.

[37] Shin-ya Takemura, Arjun Bharioke, Zhiyuan Lu, Aljoscha Nern, Shiv Vitaladevuni, Patricia K Rivlin, William T Katz, Donald J Olbris, Stephen M Plaza, Philip Winston, et al. A visual motion detection circuit suggested by drosophila connectomics. *Nature*, 500(7461):175–181, 2013. doi:10.1038/nature12450.

[38] Mei Zhen and Aravinthan DT Samuel. C. elegans locomotion: small circuits, complex functions. *Curr Opin Neurobiol*, 33:117–126, 2015. doi:10.1016/j.conb.2015.03.009.

[39] John G. White, Eileen Southgate, James N. Thomson, and Sydney Brenner. The structure of the nervous system of the nematode *caenorhabditis elegans. Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 314(1165):1–340, 11 1986. doi:10.1098/rstb.1986.0056.

[40] Henry Markram, Eilif Muller, Srikanth Ramaswamy, Michael W Reimann, Marwan Abdellah, Carlos Aguado Sanchez, Anastasia Ailamaki, Lidia Alonso-Nanclares, Nicolas Antille, Selim Arsever, et al. Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2):456–492, 2015. doi:10.1016/j.cell.2015.09.029.

[41] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013. doi:10.1016/j.neuroimage.2013.05.041.

[42] Hideyuki Okano, Atsushi Miyawaki, and Kiyoto Kasai. Brain/minds: brain-mapping project in japan. *Phil Trans R Soc B*, 370(1668):20140310, 2015. doi:10.1098/rstb.2014.0310.

[43] Mu-ming Poo, Jiu-lin Du, Nancy Y Ip, Zhi-Qi Xiong, Bo Xu, and Tieniu Tan. China brain project: basic neuroscience, brain diseases, and brain-inspired computing. *Neuron*, 92(3):591–596, 2016. doi:10.1016/j.neuron.2016.10.050.

[44] Javier DeFelipe. The anatomical problem posed by brain complexity and size: a potential solution. *Front Neuroanat*, 9:104, 2015. doi:10.3389/fnana.2015.00104.

[45] Michel A Hofman. Evolution of the human brain: when bigger is better. *Front Neuroanat*, 8:15, 2014. doi:10.3389/fnana.2014.00015.

[46] Suzana Herculano-Houzel. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience*, 3:31, 2009. doi:10.3389/neuro.09.031.2009.

[47] Martijn P Van den Heuvel, Edward T Bullmore, and Olaf Sporns. Comparative connectomics. *Trends Cogn Sci*, 20(5):345–361, 2016. doi:10.1016/j.tics.2016.03.001.

[48] Marcus Kaiser. Mechanisms of connectome development. *Trends in Cognitive Sciences*, 21(9):703–717, 9 2017. ISSN 1364-6613. doi:10.1016/j.tics.2017.05.010.

[49] Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nat Rev Neurosci*, 13(5):336, 2012. doi:10.1038/nrn3214.

[50] Richard F Betzel and Danielle S Bassett. Generative models for network neuroscience: prospects and promise. *J R Soc Interface*, 14(136):20170623, 2017. doi:10.1098/rsif.2017.0623.

[51] Karl Friston, James Kilner, and Lee Harrison. A free energy principle for the brain. *J Physiol Paris*, 100(1-3):70–87, 2006. doi:10.1016/j.jphysparis.2006.10.001.

[52] Karl Friston. The free-energy principle: a unified brain theory? *Nat Rev Neurosci*, 11(2):127–138, 2010. doi:10.1038/nrn2787.

[53] Jeff W. Lichtman, Hanspeter Pfister, and Nir Shavit. The big data challenges of connectomics. *Nature Neuroscience*, 17(11):1448, 2014. doi:10.1038/nn.3837.

[54] Bassem A. Hassan and P. Robin Hiesinger. Beyond molecular codes: simple rules to wire complex brains. *Cell*, 163(2):285–291, 2015. doi:10.1016/j.cell.2015.09.031.

[55] Kevin J. Mitchell. The genetics of brain wiring: from molecule to mind. *PLoS Biology*, 5(4):e113, 2007. doi:10.1371/journal.pbio.0050113.

[56] Ralph Adolphs. The unsolved problems of neuroscience. *Trends in Cognitive Sciences*, 19(4):173–175, 2015. doi:10.1016/j.tics.2015.01.007.

[57] Sofia J. Araújo and Guy Tear. Axon guidance mechanisms and molecules: lessons from invertebrates. *Nature Reviews Neuroscience*, 4(11):910, 2003. doi:10.1038/nrn1243.

[58] Shih-Yu Chen and Hwai-Jong Cheng. Functions of axon guidance molecules in synapse formation. *Current Opinion in Neurobiology*, 19(5):471–478, 2009. doi:10.1016/j.conb.2009.09.005.

[59] Daniel A. Colón-Ramos. Synapse formation in developing neural circuits. volume 87, pages 53–79. Elsevier, 2009. doi:10.1016/S0070-2153(09)01202-2.

[60] Alex L. Kolodkin and Marc Tessier-Lavigne. Mechanisms and molecules of neuronal wiring: a primer. *Cold Spring Harbor Perspectives in Biology*, 3(6):a001727, 2011. doi:10.1101/cshperspect.a001727.

[61] Lav R. Varshney, Beth L. Chen, Eric Paniagua, David H. Hall, and Dmitri B. Chklovskii. Structural properties of the *caenorhabditis elegans* neuronal network. *PLOS Computational Biology*, 7(2):e1001066, 02 2011. doi:10.1371/journal.pcbi.1001066.

[62] Sydney Brenner. In the beginning was the worm... *Genetics*, 182(2):413–415, 2009. doi:10.1534/genetics.109.104976.

[63] Donald L. Riddle, Thomas Blumenthal, Barbara J. Meyer, and James R. Priess, editors. *C. elegans II*, volume 33 of *Cold Spring Harbor Monograph Series*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997. ISBN 9780879694883.

[64] Eric S. Haag, David H. A. Fitch, and Marie Delattre. From "the worm" to "the worms" and back again: The evolutionary developmental biology of nematodes. *Genetics*, 210(2):397–433, 2018. doi:10.1534/genetics.118.300243.

[65] LaDeana W Hillier, Alan Coulson, John I Murray, Zhirong Bao, John E Sulston, and Robert H Waterston. Genomics in c. elegans: so many genes, such a little worm. *Genome Research*, 15(12):1651–1660, 2005. doi:10.1101/gr.3729105.

[66] John E. Sulston and Howard R. Horvitz. Post-embryonic cell lineages of the nematode, *caenorhabditis elegans*. *Developmental Biology*, 56(1):110 – 156, 1977. ISSN 0012-1606. doi:10.1016/0012-1606(77)90158-0.

[67] John E. Sulston, Einhard Schierenberg, John G. White, and James N. Thomson. The embryonic cell lineage of the nematode *caenorhabditis elegans*. *Developmental Biology*, 100(1):64 – 119, 1983. ISSN 0012-1606. doi:10.1016/0012-1606(83)90201-4.

[68] Scott W. Emmons. Connectomics, the final frontier. In *Current Topics in Developmental Biology*, volume 116, pages 315–330. Elsevier, 2016. doi:10.1016/bs.ctdb.2015.11.001.

[69] Sydney Brenner. The genetics of *caenorhabditis elegans*. *Genetics*, 77(1):71–94, 1974.

[70] Joseph G. Culotti. Axon guidance mechanisms in *caenorhabditis elegans*. *Current Opinion in Genetics & Development*, 4(4):587–595, 1994. doi:10.1016/0959-437X(94)90077-G.

[71] Milica A. Margeta, Kang Shen, and Brock Grill. Building a synapse: Lessons on synaptic specificity and presynaptic assembly from the nematode *c. elegans*. *Current Opinion in Neurobiology*, 18(1):69–76, 2008. doi:10.1016/j.conb.2008.04.003.

[72] Salvatore J. Cherra and Yishi Jin. Advances in synapse formation: forging connections in the worm. *WIREs Developmental Biology*, 4(2):85–97, 2015. doi:10.1002/wdev.165.

[73] Geoffrey J. Goodhill. Can molecular gradients wire the brain? *Trends in Neurosciences*, 39(4):202–211, 2016. doi:10.1016/j.tins.2016.01.009.

[74] David Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Co., San Francisco, CA, 1982. ISBN 0716712849.

[75] Kent A Stevens. The vision of david marr. *Perception*, 41(9):1061–1072, 2012. doi:10.1068/p7297.

[76] Danielle S Bassett, Edward Bullmore, Beth A Verchinski, Venkata S Mattay, Daniel R Weinberger, and Andreas Meyer-Lindenberg. Hierarchical organization of human cortical networks in health and schizophrenia. *Journal of Neuroscience*, 28(37):9239–9248, 2008. doi:10.1523/JNEUROSCI.1929-08.2008.

[77] Hae-Jeong Park and Karl Friston. Structural and functional brain networks: From connections to cognition. *Science*, 342(6158):1238411, 2013. doi:10.1126/science.1238411.

[78] Cornelis J Stam. Modern network science of neurological disorders. *Nature Reviews Neuroscience*, 15(10):683, 2014. doi:10.1038/nrn3801.

[79] Alex Fornito, Andrew Zalesky, and Michael Breakspear. The connectomics of brain disorders. *Nature Reviews Neuroscience*, 16(3):159, 2015. doi:10.1038/nrn3901.

[80] Manuel Schröter, Ole Paulsen, and Edward T Bullmore. Micro-connectomics: Probing the organization of neuronal networks at the cellular scale. *Nature Reviews Neuroscience*, 18(3):131–146, 2017. doi:10.1038/nrn.2016.182.

[81] Markus Reigl, Uri Alon, and Dmitri B. Chklovskii. Search for computational modules in the *c. elegans* brain. *BMC Biology*, 2(1):25, 2004. doi:10.1186/1741-7007-2-25.

[82] Raj Kumar Pan, Nivedita Chatterjee, and Sitabhra Sinha. Mesoscopic organization reveals the constraints governing *caenorhabditis elegans* nervous system. *PLOS ONE*, 5(2):e9240, 02 2010. doi:10.1371/journal.pone.0009240.

[83] Sreedevi Varier and Marcus Kaiser. Neural development features: Spatio-temporal development of the *caenorhabditis elegans* neuronal network. *PLOS Computational Biology*, 7(1):e1001044, 01 2011. doi:10.1371/journal.pcbi.1001044.

[84] Bradly Alicea. The emergent connectome in *caenorhabditis elegans* embryogenesis. *Biosystems*, 173:247–255, 2018. doi:10.1016/j.biosystems.2018.09.016.

[85] Yong-Yeol Ahn, Hawoong Jeong, and Beom Jun Kim. Wiring cost in the organization of a biological neuronal network. *Physica A: Statistical Mechanics and its Applications*, 367:531 – 537, 2006. ISSN 0378-4371. doi:10.1016/j.physa.2005.12.013.

[86] Beth L. Chen, David H. Hall, and Dmitri B. Chklovskii. Wiring optimization can relate neuronal structure and function. *Proceedings of the National Academy of Sciences USA*, 103(12):4723–4728, 2006. doi:10.1073/pnas.0506806103.

[87] Alfonso Pérez-Escudero and Gonzalo G. de Polavieja. Optimally wired subnetwork determines neuroanatomy of *caenorhabditis elegans*. *Proceedings of the National Academy of Sciences USA*, 104(43):17180–17185, 2007. ISSN 0027-8424. doi:10.1073/pnas.0703183104.

[88] Marta Rivera-Alba, Hanchuan Peng, Gonzalo G de Polavieja, and Dmitri B Chklovskii. Wiring economy can account for cell body placement across species and brain areas. *Current Biology*, 24(3):R109–R110, 2014. doi:10.1016/j.cub.2013.12.012.

[89] Andrey Gushchin and Ao Tang. Total wiring length minimization of *c. elegans* neural network: A constrained optimization approach. *PLOS ONE*, 10(12):e0145029, 12 2015. doi:10.1371/journal.pone.0145029.

[90] Irving E. Wang and Thomas R. Clandinin. The influence of wiring economy on nervous system evolution. *Current Biology*, 26(20):R1101–R1108, 2016. doi:10.1016/j.cub.2016.08.053.

[91] Vincenzo Nicosia, Petra E. Vértes, William R. Schafer, Vito Latora, and Edward T. Bullmore. Phase transition in the economically modeled growth of a cellular nervous system. *Proceedings of the National Academy of Sciences USA*, 110(19):7880–7885, 2013. ISSN 0027-8424. doi:10.1073/pnas.1300753110. URL `https://www.pnas.org/content/110/19/7880`.

[92] Martin Chalfie, John E. Sulston, John G. White, Eileen Southgate, James N. Thomson, and Sydney Brenner. The neural circuit for touch sensitivity in *caenorhabditis elegans*. *Journal of Neuroscience*, 5(4):956–964, 1985. ISSN 0270-6474. doi:10.1523/JNEUROSCI.05-04-00956.1985.

[93] Stephen R. Wicks and Catherine H. Rankin. Integration of mechanosensory stimuli in *caenorhabditis elegans*. *Journal of Neuroscience*, 15(3):2434–2444, 1995. ISSN 0270-6474. doi:10.1523/JNEUROSCI.15-03-02434.1995.

[94] Elizabeth R. Sawin. *Genetic and cellular analysis of modulated behaviors in* Caenorhabditis elegans. PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1996.

[95] Emily R. Troemel, Bruce E. Kimmel, and Cornelia I. Bargmann. Reprogramming chemotaxis responses: Sensory neurons define olfactory preferences in c. elegans. *Cell*, 91(2):161 – 169, 1997. ISSN 0092-8674. doi:10.1016/S0092-8674(00)80399-2.

[96] Gal Haspel and Michael J. O'Donovan. A perimotor framework reveals functional segmentation in the motoneuronal network controlling locomotion in *caenorhabditis elegans*. *Journal of Neuroscience*, 31(41):14611–14623, 2011. doi:10.1523/JNEUROSCI.2186-11.2011.

[97] Steven J. Cook, Travis A. Jarrell, Christopher A. Brittin, Yi Wang, Adam E. Bloniarz, Maksim A. Yakovlev, Ken C. Q. Nguyen, Leo T.-H. Tang, Emily A. Bayer, Janet S. Duerr, Hannes E. Bülow, Oliver Hobert, David H. Hall, and Scott W. Emmons. Whole-animal connectomes of both *caenorhabditis elegans* sexes. *Nature*, 571 (7763):63–71, 2019. doi:10.1038/s41586-019-1352-7.

[98] WormAtlas, Zeynep F. Altun, Laura A. Herndon, Cathy A. Wolkow, Chris Crocker, Robyn Lints, and David H. Hall. A database featuring behavioural and structural anatomy of *caenorhabditis elegans*, 2002-2019. URL `http://www.wormatlas.org/`.

[99] Nikhil Bhatla. An interactive visualization of the *c. elegans* cell lineage, 2011. URL `http://wormweb.org/celllineage`.

[100] Yoonsuck Choe, Bruce H. McCormick, and W. Koh. Network connectivity analysis on the temporally augmented c. elegans web: A pilot study. In *Soc. Neurosci. Abstr.*, volume 30, 2004.

[101] William B. Wood, editor. *The Nematode* Caenorhabditis Elegans, volume 17 of *Cold Spring Harbor Monograph Series*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988. ISBN 9780879694333.

[102] Theodore B. Achacoso and William S. Yamamoto. *AY's Neuroanatomy of* C. Elegans *for Computation*. CRC Press, Boca Raton, FL, 1st edition, 1991. ISBN 9780849342349.

[103] Laura E. Waggoner, G. Tong Zhou, Ronald W. Schafer, and William R. Schafer. Control of alternative behavioral states by serotonin in *caenorhabditis elegans*. *Neuron*, 21(1):203 – 214, 1998. ISSN 0896-6273. doi:10.1016/S0896-6273(00)80527-9.

[104] I. Amy Bany, Meng-Qiu Dong, and Michael R. Koelle. Genetic and cellular basis for acetylcholine inhibition of *caenorhabditis elegans* egg-laying behavior. *Journal of Neuroscience*, 23(22):8060–8069, 2003. ISSN 0270-6474. doi:10.1523/JNEUROSCI.23-22-08060.2003.

[105] Laura A. Hardaker, Emily Singer, Rex Kerr, Guotong Zhou, and William R. Schafer. Serotonin modulates locomotory behavior and coordinates egg-laying and movement in *caenorhabditis elegans*. *Journal of Neurobiology*, 49(4):303–313, 2001. doi:10.1002/neu.10014.

[106] Ikue Mori and Yasumi Ohshima. Neural regulation of thermotaxis in *caenorhabditis elegans*. *Nature*, 376:344–348, 1995. ISSN 1476-4687. doi:10.1038/376344a0.

[107] Matthew Beverly, Sriram Anbil, and Piali Sengupta. Degeneracy and neuromodulation among thermosensory neurons contribute to robust thermosensory behaviors in *caenorhabditis elegans*. *Journal of Neuroscience*, 31(32):11718–11727, 2011. doi:10.1523/JNEUROSCI.1098-11.2011.

[108] Shu Liu, Ekkehard Schulze, and Ralf Baumeister. Temperature-and touch-sensitive neurons couple cng and trpv channel activities to control heat avoidance in *caenorhabditis elegans*. *PloS one*, 7(3):e32360, 2012. doi:10.1371/journal.pone.0032360.

[109] Jesse M. Gray, Joseph J. Hill, and Cornelia I. Bargmann. A circuit for navigation in *caenorhabditis elegans*. *Proceedings of the National Academy of Sciences USA*, 102 (9):3184–3191, 2005. ISSN 0027-8424. doi:10.1073/pnas.0409009101.

[110] Stephen R. Wicks and Catherine H. Rankin. The integration of antagonistic reflexes revealed by laser ablation of identified neurons determines habituation kinetics of the caenorhabditis elegans tap withdrawal response. *Journal of Comparative Physiology A*, 179(5):675–685, Nov 1996. ISSN 1432-1351. doi:10.1007/BF00216131.

[111] Andy J. Chang, Nikolas Chronis, David S. Karow, Michael A. Marletta, and Cornelia I. Bargmann. A distributed chemosensory circuit for oxygen preference in *c. elegans*. *PLoS Biology*, 4(9):e274, 2006. doi:10.1371/journal.pbio.0040274.

[112] Manuel Zimmer, Jesse M Gray, Navin Pokala, Andy J Chang, David S Karow, Michael A Marletta, Martin L Hudson, David B Morton, Nikos Chronis, and Cornelia I Bargmann. Neurons detect increases and decreases in

oxygen levels using distinct guanylate cyclases. *Neuron*, 61(6):865–879, 2009. doi:10.1016/j.neuron.2009.02.013.

[113] Andrew Jonathan Bretscher, Karl Emanuel Busch, and Mario de Bono. A carbon dioxide avoidance behavior is integrated with responses to ambient oxygen and food in *caenorhabditis elegans*. *Proceedings of the National Academy of Sciences USA*, 105 (23):8044–8049, 2008. doi:10.1073/pnas.0707607105.

[114] Andrew Jonathan Bretscher, Eiji Kodama-Namba, Karl Emanuel Busch, Robin Joseph Murphy, Zoltan Soltesz, Patrick Laurent, and Mario de Bono. Temperature, oxygen, and salt-sensing neurons in *c. elegans* are carbon dioxide sensors that control avoidance behavior. *Neuron*, 69(6):1099–1113, 2011. doi:10.1016/j.neuron.2011.02.023.

[115] Mark E. J. Newman. Detecting community structure in networks. *Eur Phys J B*, 38 (2):321–330, 2004. doi:10.1140/epjb/e2004-00124-y.

[116] Roland Pfister, Katharina A. Schwarz, Markus Janczyk, Rick Dale, and Jon Freeman. Good things peak in pairs: A note on the bimodality coefficient. *Frontiers in Psychology*, 4:700, 2013. doi:10.3389/fpsyg.2013.00700.

[117] Roger Guimerà and Luís A. N. Amaral. Functional cartography of complex metabolic networks. *nature*, 433(7028):895–900, 2005. doi:10.1038/nature03288.

[118] John S. Long. *Regression Models for Categorical and Limited Dependent Variables*. SAGE, 1997. ISBN 0803973748.

[119] David A. Belsley, Edwin Kuh, and Roy E. Welsch. *Regression diagnostics: identifying influential data and sources of collinearity*. Wiley, 1980. ISBN 0471058564.

[120] Frank J. Massey. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[121] Adrian W. Bowman and Adelchi Azzalini. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-PLUS Illustrations*. Oxford University Press, Oxford, 1997. ISBN 9780198523963.

[122] Marcus Kaiser and Claus C. Hilgetag. Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLOS Computational Biology*, 2(7):e95, 07 2006. doi:10.1371/journal.pcbi.0020095.

[123] Mark E.J. Newman. *Networks: An Introduction*. Oxford University Press, Oxford, 2010. ISBN 9780191500701.

[124] Tianqi Xu, Jing Huo, Shuai Shao, Michelle Po, Taizo Kawano, Yangning Lu, Min Wu, Mei Zhen, and Quan Wen. Descending pathway facilitates undulatory wave propagation in *caenorhabditis elegans* through gap junctions. *Proceedings of the National Academy of Sciences USA*, 115(19):E4493–E4502, 2018. doi:10.1073/pnas.1717022115.

[125] David H. Hall. Gap junctions in *c. elegans*: Their roles in behavior and development. *Developmental Neurobiology*, 77(5):587–596, 2017. doi:10.1002/dneu.22408.

[126] Claudiu A. Giurumescu and Andrew D. Chisholm. Cell identification and cell lineage analysis. In Joel H. Rothman and Andrew Singson, editors, *Methods in Cell Biology*, volume 106, pages 323–341. Elsevier, Waltham, MA, 2011.

[127] William Schafer. Nematode nervous systems. *Current Biology*, 26(20):R955–R959, 2016. doi:10.1016/j.cub.2016.07.044.

[128] Marc Tessier-Lavigne and Corey S. Goodman. The molecular biology of axon guidance. *Science*, 274(5290):1123–1133, 1996. doi:10.1126/science.274.5290.1123.

[129] Oliver Hobert. Development of left/right asymmetry in the *caenorhabditis elegans* nervous system: from zygote to postmitotic neuron. *Genesis*, 52(6):528–543, 2014. doi:10.1002/dvg.22747.

[130] Oliver Hobert. Specification of the nervous system (august 8, 2005). In The C. elegans Research Community, editor, *WormBook*. www.wormbook.org, 2005. doi:10.1895/wormbook.1.12.1.

[131] Elizabeth M. DiLoreto, Christopher D. Chute, Samantha Bryce, and Jagan Srinivasan. Novel technological advances in functional connectomics in *c. elegans*. *Journal of Developmental Biology*, 7(2), 2019. doi:10.3390/jdb7020008.

[132] Athanasios Metaxakis, Dionysia Petratou, and Nektarios Tavernarakis. Multimodal sensory processing in *caenorhabditis elegans*. *Open Biology*, 8(6), 2018. doi:10.1098/rsob.180049.

[133] Oliver Hobert. Behavioral plasticity in *c. elegans*: paradigms, circuits, genes. *Journal of Neurobiology*, 54(1):203–223, 2003. doi:10.1002/neu.10168.

[134] Oliver Hobert, Robert J Johnston Jr, and Sarah Chang. Left–right asymmetry in the nervous system: the *caenorhabditis elegans* model. *Nature Reviews Neuroscience*, 3(8):629, 2002. doi:10.1038/nrn897.

[135] Oleg Tolstenkov, Petrus Van der Auwera, Wagner S. Costa, Olga Bazhanova, Tim M. Gemeinhardt, Amelie C. F. Bergs, and Alexander Gottschalk. Functionally asymmetric motor neurons contribute to coordinating locomotion of *caenorhabditis elegans*. *eLife*, 7:e34997, 2018. doi:10.7554/eLife.34997.

[136] Roger Guimerà and Luís A. N. Amaral. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005 (02):P02001, 2005. doi:10.1088/1742-5468/2005/02/P02001.

[137] Cornelia I. Bargmann. Beyond the connectome: how neuromodulators shape neural circuits. *Bioessays*, 34(6):458–465, 2012. doi:10.1002/bies.201100185.

[138] Barry Bentley, Robyn Branicky, Christopher L. Barnes, Yee Lian Chew, Eviatar Yemini, Edward T. Bullmore, Petra E. Vértes, and William R. Schafer. The mul-

tilayer connectome of *caenorhabditis elegans*. *PLoS Computational Biology*, 12(12): e1005283, 2016. doi:10.1371/journal.pcbi.1005283.

[139] Lewis Wolpert and Cheryll Tickle. *Principles of Development (4th Edition)*. Oxford University Press, 2011. ISBN 9780199549078.

[140] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning about a highly connected world*. Cambridge University Press, Cambridge, 2010. ISBN 9780521195331.

[141] Jörn Davidsen, Holger Ebel, and Stefan Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks. *Physical Review Letters*, 88(12): 128701, 2002. doi:10.1103/PhysRevLett.88.128701.

[142] Hilla Brot, Lev Muchnik, and Yoram Louzoun. Directed triadic closure and edge deletion mechanism induce asymmetry in directed edge properties. *The European Physical Journal B*, 88(1):12, 2015. doi:10.1140/epjb/e2014-50220-4.

[143] Aharon Azulay, Eyal Itskovits, and Alon Zaslaver. The *c. elegans* connectome consists of homogenous circuits with defined functional roles. *PLoS Computational Biology*, 12(9):e1005021, 2016. doi:10.1371/journal.pcbi.1005021.

[144] Karl Zilles and Katrin Amunts. Centenary of brodmann's map—conception and fate. *Nat Rev Neurosci*, 11(2):139–145, 2010. doi:10.1038/nrn2776.

[145] Alexandros Goulas, Richard F Betzel, and Claus C Hilgetag. Spatiotemporal ontogeny of brain wiring. *Sci Adv*, 5(6):eaav9694, 2019. doi:10.1126/sciadv.aav9694.

[146] Peter T Fox and Karl J Friston. Distributed processing; distributed functions? *Neuroimage*, 61(2):407–426, 2012. doi:10.1016/j.neuroimage.2011.12.051.

[147] Karl Friston. Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annu Rev Neurosci*, 25(1):221–250, 2002. doi:10.1146/annurev.neuro.25.112701.142846.

[148] Gustavo Deco, Viktor K Jirsa, Peter A Robinson, Michael Breakspear, and Karl Friston. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comput Biol*, 4(8):e1000092, 2008. doi:10.1371/journal.pcbi.1000092.

[149] Yuhan Chen, Zi-Ke Zhang, Yong He, and Changsong Zhou. A large-scale high-density weighted structural connectome of the macaque brain acquired by predicting missing links. *Cereb Cortex*, Advance online publication, 2020. doi:10.1093/cercor/bhaa060.

[150] David Meunier, Renaud Lambiotte, and Edward T Bullmore. Modular and hierarchically modular organization of brain networks. *Front Neurosci*, 4:200, 2010. doi:10.3389/fnins.2010.00200.

[151] Raj Kumar Pan and Sitabhra Sinha. Modular networks emerge from multiconstraint optimization. *Phys Rev E*, 76(4):045103, 2007. doi:10.1103/PhysRevE.76.045103.

[152] Jerry A. Fodor. *The Modularity of Mind*. A Bradford Book. MIT Press, Cambridge MA, 1983. ISBN 9780262260701.

[153] Larry W Swanson and Jeff W Lichtman. From cajal to connectome and beyond. *Annu Rev Neurosci*, 39:197–216, 2016. doi:10.1146/annurev-neuro-071714-033954.

[154] Raj Kumar Pan and Sitabhra Sinha. Modularity produces small-world networks with dynamical time-scale separation. *EPL*, 85(6):68006, 2009. doi:10.1209/0295-5075/85/68006/meta.

[155] David Mumford. On the computational architecture of the neocortex. *Biol Cybern*, 65(2):135–145, 1991. doi:10.1007/BF00202389.

[156] Chris G Antonopoulos, Shambhavi Srivastava, Sandro E de S Pinto, and Murilo S Baptista. Do brain networks evolve by maximizing their information flow capacity? *PLoS Comput Biol*, 11(8):e1004372, 2015. doi:10.1371/journal.pcbi.1004372.

[157] Yutaka Yamaguti and Ichiro Tsuda. Mathematical modeling for evolution of heterogeneous modules in the brain. *Neural Netw*, 62:3–10, 2015. doi:10.1016/j.neunet.2014.07.013.

[158] Jennifer Stiso and Danielle S Bassett. Spatial embedding imposes constraints on neuronal network architectures. *Trends Cogn Sci*, 22(12):1127–1142, 2018. doi:10.1016/j.tics.2018.09.007.

[159] Dharmendra S. Modha and Raghavendra Singh. Network architecture of the long-distance pathways in the macaque brain. *Proc Natl Acad Sci USA*, 107(30):13485–13490, 2010. ISSN 0027-8424. doi:10.1073/pnas.1008054107.

[160] Klaas E. Stephan, Karl Zilles, and Rolf Kötter. Coordinate-independent mapping of structural and functional data by objective relational transformation (ort). *Philos Trans R Soc London, Ser B*, 355(1393):37–54, 2000. doi:10.1098/rstb.2000.0548.

[161] Klass E. Stephan, Lars Kamper, Ahmet Bozkurt, Gully A. P. C. Burns, Malcolm P. Young, and Rolf Kötter. Advanced database methodology for the collation of connectivity data on the macaque brain (cocomac). *Philos Trans R Soc London, Ser B*, 356(1412):1159–1186, 2001. doi:10.1098/rstb.2001.0908.

[162] Rolf Kötter. Online retrieval, processing, and visualization of primate connectivity data from the cocomac database. *Neuroinformatics*, 2(2):127–144, 2004. doi:10.1385/NI:2:2:127.

[163] Scalable brain atlas: Macaque. `https://scalablebrainatlas.incf.org/macaque/PHT00`.

[164] George Paxinos, Xu F. Huang, and Arthur W. Toga. *The Rhesus Monkey Brain in Stereotaxic Coordinates*. Academic Press, San Diego, CA, 2000. ISBN 9780123582553. URL `https://books.google.co.in/books?id=x-6bQgAACAAJ`.

[165] Mark E. J. Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Phys Rev E*, 69(2):026113, 2004. doi:10.1103/PhysRevE.69.026113.

[166] Benjamin H Good, Yves Alexandre De Montjoye, and Aaron Clauset. Performance of modularity maximization in practical contexts. *Phys Rev E*, 81(4):046106, 2010. doi:10.1103/PhysRevE.81.046106.

[167] On modularity maximization. `http://tuvalu.santafe.edu/~aaronc/modularity/`.

[168] Martin Rosvall and Carl T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci USA*, 105(4):1118–1123, 2008. doi:10.1073/pnas.0706851105.

[169] Roger Guimera, Marta Sales-Pardo, and Luis A. N. Amaral. Classes of complex networks defined by role-to-role connectivity profiles. *Nat Phys*, 3(1):63, 2007. doi:10.1038/nphys489.

[170] Paul Expert, Tim S. Evans, Vincent D. Blondel, and Renaud Lambiotte. Uncovering space-independent communities in spatial networks. *Proc Natl Acad Sci USA*, 108 (19):7663–7668, 2011. doi:10.1073/pnas.1018962108.

[171] David J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge, UK, 2003. ISBN 9780521642989.

[172] Jason N MacLean, Brendon O Watson, Gloster B Aaron, and Rafael Yuste. Internal dynamics determine the cortical response to thalamic stimulation. *Neuron*, 48(5): 811–823, 2005. doi:10.1016/j.neuron.2005.09.035.

[173] Leo J. Van Hemmen and Terrence J. Sejnowski. *23 Problems in Systems Neuroscience*. Computational Neuroscience Series. Oxford University Press, USA, 2006. ISBN 9780195148220. URL `https://books.google.co.in/books?id=bk1nDAAAQBAJ`.

[174] Mircea Steriade, David A. McCormick, and Terrence J. Sejnowski. Thalamocortical oscillations in the sleeping and aroused brain. *Science*, 262(5134):679–685, 1993. doi:10.1126/science.8235588.

[175] Ray W. Guillery. Anatomical evidence concerning the role of the thalamus in corticocortical communication: a brief review. *J Anat*, 187(3):583, 1995.

[176] S. Murray Sherman and Ray W. Guillery. Functional organization of thalamocortical relays. *J Neurophysiol*, 76(3):1367–1395, 1996. doi:10.1152/jn.1996.76.3.1367.

[177] Michele Mauri, Tommaso Elli, Giorgio Caviglia, Giorgio Uboldi, and Matteo Azzi. Rawgraphs: A visualisation platform to create open outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*, CHItaly '17, pages 28:1–28:5, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-5237-6. doi:10.1145/3125571.3125585.

[178] Claude M. J. Braun, Mathieu Dumont, Julie Duval, Isabelle Hamel-Hébert, and Lucie Godbout. Brain modules of hallucination: an analysis of multiple patients with brain lesions. *J Psychiatry Neurosci*, 28(6):432, 2003.

[179] Saul Sternberg. Modular processes in mind and brain. *Cogn Neuropsychol*, 28(3-4): 156–208, 2011. doi:10.1080/02643294.2011.557231.

[180] Michael S Gazzaniga. Brain modules and belief formation. In *Self and consciousness*, pages 96–110. Psychology Press, Abingdon, UK, 2014.

[181] Leland H Hartwell, John J Hopfield, Stanislas Leibler, and Andrew W Murray. From molecular to modular cell biology. *Nature*, 402(6761supp):C47, 1999. doi:https://doi.org/10.1038/35011540.

[182] Dale Ed Purves, George J Augustine, David Ed Fitzpatrick, William C Hall, Anthony-Samuel Ed LaMantia, James O McNamara, and Leonard E White. *Neuroscience*. Sinauer Associates, Sunderland (MA), Sunderland, MA, 2001.

[183] Eric Kandel, James Schwartz, and Thomas Jessell. *Principles of Neural Science, Fourth Edition*, volume 4. McGraw-hill, New York, 2000.

[184] Mark E. Walton, Timothy E. J. Behrens, Mark J. Buckley, Peter H. Rudebeck, and Matthew F. S. Rushworth. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron*, 65(6):927–939, 2010. doi:10.1016/j.neuron.2010.02.027.

[185] Lesley K. Fellows. The role of orbitofrontal cortex in decision making: a component process account. *Ann N Y Acad Sci*, 1121(1):421–430, 2007. doi:10.1196/annals.1401.023.

[186] Mark E. Walton, Joseph T. Devlin, and Matthew F. S. Rushworth. Interactions between decision making and performance monitoring within prefrontal cortex. *Nat Neurosci*, 7(11):1259, 2004. doi:10.1038/nn1339.

[187] Alicia Izquierdo, Robin K. Suda, and Elisabeth A. Murray. Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J Neurosci*, 24(34):7540–7548, 2004. doi:10.1523/JNEUROSCI.1921-04.2004.

[188] Edmund T. Rolls. The orbitofrontal cortex and reward. *Cereb Cortex*, 10(3):284–294, 2000. doi:10.1093/cercor/10.3.284.

[189] Lawrence Weiskrantz. Behavioral changes associated with ablation of the amygdaloid complex in monkeys. *J. Comp. Physiol. Psychol.*, 49(4):381, 1956.

[190] Michael Davis. The role of the amygdala in fear and anxiety. *Annu Rev Neurosci*, 15(1):353–375, 1992. doi:10.1146/annurev.ne.15.030192.002033.

[191] Simon Killcross, Trevor W Robbins, and Barry J Everitt. Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. *Nature*, 388 (6640):377, 1997. doi:10.1038/41097.

[192] Stephen Maren. Long-term potentiation in the amygdala: a mechanism for emotional learning and memory. *Trends Neurosci*, 22(12):561–567, 1999. doi:10.1016/S0166-2236(99)01465-4.

[193] Elizabeth A. Kensinger and Suzanne Corkin. Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Mem Cogn*, 31(8):1169–1180, 2003. doi:10.3758/BF03195800.

[194] Mark P. Richardson, Bryan A. Strange, and Raymond J. Dolan. Encoding of emotional memories depends on amygdala and hippocampus and their interactions. *Nat Neurosci*, 7(3):278, 2004. doi:10.1038/nn1190.

[195] Adam K. Anderson, Peter E. Wais, and John D. E. Gabrieli. Emotion enhances remembrance of neutral events past. *Proc Natl Acad Sci USA*, 103(5):1599–1604, 2006. doi:10.1073/pnas.0506308103.

[196] Arielle Tambini, Ulrike Rimmele, Elizabeth A Phelps, and Lila Davachi. Emotional brain states carry over and enhance future memory formation. *Nat Neurosci*, 20(2): 271, 2017. doi:10.1038/nn.4468.

[197] Patricia S. Churchland and Terrence J. Sejnowski. Perspectives on cognitive neuroscience. *Science*, 242(4879):741–745, 1988. doi:10.1126/science.3055294.

[198] Melvyn A. Goodale and A. David Milner. Separate visual pathways for perception and action. *Trends Neurosci*, 15(1):20–25, 1992. doi:10.1016/0166-2236(92)90344-8.

[199] Melvyn A. Goodale and A. David Milner. Two visual pathways—where have they taken us and where will they lead in future? *Cortex*, 98:283–292, 2018. doi:10.1016/j.cortex.2017.12.002.

[200] Bruno B Averbeck and Moonsang Seo. The statistical neuroanatomy of frontal networks in the macaque. *PLoS Comput Biol*, 4(4):e1000050, 2008. doi:10.1371/journal.pcbi.1000050.

[201] Suzana Herculano-Houzel, Bruno Mota, Peiyan Wong, and Jon H. Kaas. Connectivity-driven white matter scaling and folding in primate cerebral cortex. *Proc Natl Acad Sci USA*, 107(44):19008–19013, 2010. doi:10.1073/pnas.1012590107.

[202] Dmitri B. Chklovskii. Exact solution for the optimal neuronal layout problem. *Neural Comput*, 16(10):2067–2078, 2004. doi:10.1162/0899766041732422.

[203] Marcus Kaiser and Claus C Hilgetag. Modelling the development of cortical systems networks. *Neurocomputing*, 58:297–302, 2004. doi:10.1016/j.neucom.2004.01.059.

[204] György Buzsáki, Caroline Geisler, Darrell A Henze, and Xiao-Jing Wang. Interneuron diversity series: circuit complexity and axon wiring economy of cortical interneurons. *Trends Neurosci*, 27(4):186–193, 2004. doi:10.1016/j.tins.2004.02.007.

[205] Jeremy E Niven and Simon B Laughlin. Energy limitation as a selective pressure on the evolution of sensory systems. *J Exp Biol*, 211(11):1792–1804, 2008. doi:10.1242/jeb.017574.

[206] Marc Barthélemy. Spatial networks. *Phys Rep*, 499(1):1–101, 2011. doi:10.1016/j.physrep.2010.11.002.

[207] James W Gnadt and Richard A Andersen. Memory related motor planning activity in posterior parietal cortex of macaque. *Exp Brain Res*, 70(1):216–220, 1988. doi:10.1007/BF00271862.

[208] Jean-René Duhamel, Carol L Colby, and Michael E Goldberg. Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *J Neurophysiol*, 79(1):126–136, 1998. doi:10.1152/jn.1998.79.1.126.

[209] Bijan Pesaran, John S Pezaris, Maneesh Sahani, Partha P Mitra, and Richard A Andersen. Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat Neurosci*, 5(8):805–811, 2002. doi:10.1038/nn890.

[210] Hideo Sakata, Masato Taira, Akira Murata, and Seiichiro Mine. Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. *Cereb Cortex*, 5(5):429–438, 1995. doi:10.1093/cercor/5.5.429.

[211] Emad N Eskandar and John A Assad. Dissociation of visual, motor and predictive signals in parietal cortex during visual guidance. *Nat Neurosci*, 2(1):88–93, 1999. doi:10.1038/4594.

[212] Christian Grefkes and Gereon R Fink. The functional organization of the intraparietal sulcus in humans and monkeys. *J Anat*, 207(1):3–17, 2005. doi:10.1111/j.1469-7580.2005.00426.x.

[213] Matteo Colombo. Moving forward (and beyond) the modularity debate: A network perspective. *Philosophy of Science*, 80(3):356–377, 2013. doi:DOI: 10.1086/670331.

[214] James L McClelland. Integration of information: Reflections on the theme of attention and performance xvi. In Toshio Inui and James L. McClelland, editors, *Attention and performance XVI: Information integration in perception and communication*, pages 633–656. MIT Press, Cambridge, MA, 1996. doi:10.7551/mitpress/1479.003.0037.

[215] Aleksandr R. Luria. *Higher Cortical Functions in Man*. Springer US, Manhattan, NY, 2nd edition, 1966. ISBN 9780306107405.

[216] Aleksandr R. Luria. *The Working Brain: An Introduction To Neuropsychology*. Basic Books, New York, NY, 1st edition, 1973. ISBN 9780465092086.

[217] Michael K Tanenhaus, Michael J Spivey-Knowlton, Kathleen M Eberhard, and Julie C Sedivy. Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. In Toshio Inui and James L. McClelland, editors, *Attention and performance XVI: Information integration in perception and communication*, pages 457–478. MIT Press, Cambridge, MA, 1996. doi:10.7551/mitpress/1479.001.0001.

[218] James L McClelland, Felix Hill, Maja Rudolph, Jason Baldridge, and Hinrich Schütze. Extending machine language models toward human-level language understanding. *arXiv preprint arXiv:1912.05877*, 2019.

[219] David E Rumelhart. Toward an interactive model of reading. In Stanislav Dornič, editor, *Attention and performance VI*, pages 573–603. L. Erlbaum Associates, Hillsdale, NJ, 1977.

[220] Timothy T Rogers, Matthew A Lambon Ralph, Peter Garrard, Sasha Bozeat, James L McClelland, John R Hodges, and Karalyn Patterson. Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychol Rev*, 111(1):205, 2004. doi:10.1037/0033-295X.111.1.205.

[221] Karalyn Patterson, Peter J Nestor, and Timothy T Rogers. Where do you know what you know? the representation of semantic knowledge in the human brain. *Nat Rev Neurosci*, 8(12):976–987, 2007. doi:10.1038/nrn2277.

[222] Matthew A Lambon Ralph, Elizabeth Jefferies, Karalyn Patterson, and Timothy T Rogers. The neural and computational bases of semantic cognition. *Nat Rev Neurosci*, 18(1):42, 2017. doi:10.1038/nrn.2016.150.

[223] Mark T Wallace, Lee K Wilkinson, and Barry E Stein. Representation and integration of multiple sensory inputs in primate superior colliculus. *J. Neurophysiol*, 76(2):1246–1266, 1996. doi:10.1152/jn.1996.76.2.1246.

[224] Andreas Nieder. Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proc Natl Acad Sci USA*, 109(29):11860–11865, 2012. doi:10.1073/pnas.1204580109.

[225] Katja Liebal, Bridget M Waller, Katie E Slocombe, and Anne M Burrows. *Primate Communication: A Multimodal Approach*. Cambridge University Press, Cambridge, UK, 2014. ISBN 9780521195041.

[226] Yuhan Chen, Shengjun Wang, Claus C Hilgetag, and Changsong Zhou. Features of spatial and functional segregation and integration of the primate connectome revealed by trade-off between wiring cost and efficiency. *PLoS Comput Biol*, 13(9):e1005776, 2017. doi:10.1371/journal.pcbi.1005776.

[227] Jérôme Sallet, Rogier B Mars, M P Noonan, Jesper L Andersson, J X O'Reilly, Saad Jbabdi, Paula L Croxson, Mark Jenkinson, Karla L Miller, and Matthew F S Rushworth. Social network size affects neural circuits in macaques. *Science*, 334 (6056):697–700, 2011. doi:10.1126/science.1210027.

[228] Michael L Platt, Robert M Seyfarth, and Dorothy L Cheney. Adaptations for social cognition in the primate brain. *Phil Trans R Soc B*, 371(1687):20150096, 2016. doi:10.1098/rstb.2015.0096.

[229] Dorothy L Cheney and Robert M Seyfarth. Flexible usage and social function in primate vocalizations. *Proc Natl Acad Sci USA*, 115(9):1974–1979, 2018. doi:10.1073/pnas.1717572115.

[230] Benjamin Wilson, Yukiko Kikuchi, Li Sun, David Hunter, Frederic Dick, Kenny Smith, Alexander Thiele, Timothy D Griffiths, William D Marslen-Wilson, and Christopher I Petkov. Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nat Commun*, 6:8901, 2015. doi:10.1038/ncomms9901.

[231] Charles T Snowdon. Learning from monkey "talk". *Science*, 355(6330):1120–1122, 2017. doi:10.1126/science.aam7443.

[232] Josef P Rauschecker. Where did language come from? precursor mechanisms in nonhuman primates. *Curr Opin Behav Sci*, 21:195–204, 2018. doi:10.1016/j.cobeha.2018.06.003.

[233] Oren Kolodny and Shimon Edelman. The evolution of the capacity for language: The ecological context and adaptive value of a process of cognitive hijacking. *Phil Trans R Soc B*, 373(1743):20170052, 2018. doi:10.1098/rstb.2017.0052.

[234] Jonathan C Horton and Daniel L Adams. The cortical column: a structure without a function. *Phil Trans R Soc B*, 360(1456):837–862, 2005. doi:10.1098/rstb.2005.1623.

[235] Richard Passingham. How good is the macaque monkey model of the human brain? *CURR OPIN NEUROBIOL*, 19(1):6–11, 2009. doi:10.1016/j.conb.2009.01.002.

[236] James K Rilling. Comparative primate neuroimaging: insights into human brain evolution. *Trends Cogn Sci*, 18(1):46–55, 2014. doi:10.1016/j.tics.2013.09.013.

[237] Jon H Kaas. The evolution of brains from early mammals to humans. *Wiley Interdiscip Rev Cogn Sci*, 4(1):33–45, 2013. doi:10.1002/wcs.1206.

[238] Alberto Llera, Thomas Wolfers, Peter Mulders, and Christian F Beckmann. Inter-individual differences in human brain structure and morphology link to variation in demographics and behavior. *Elife*, 8:e44443, 2019. doi:10.7554/eLife.44443.

[239] Stephen Smith, Eugene Duff, Adrian Groves, Thomas E Nichols, Saad Jbabdi, Lars T Westlye, Christian K Tamnes, Andreas Engvig, Kristine B Walhovd, Anders M Fjell, et al. Structural variability in the human brain reflects fine-grained functional architecture at the population level. *J Neurosci*, 39(31):6136–6149, 2019. doi:10.1523/JNEUROSCI.2912-18.2019.

[240] James W Madole, Stuart J Ritchie, Simon R Cox, Colin R Buchanan, Maria Valdés Hernández, Susana Muñoz Maniega, Joanna M Wardlaw, Mat A Harris, Mark E Bastin, Ian J Deary, et al. Aging-sensitive networks within the human structural connectome are implicated in late-life cognitive declines. *Biol Psychiatry*, 2020. doi:10.1016/j.biopsych.2020.06.010.

[241] Florian Klimm, Danielle S Bassett, Jean M Carlson, and Peter J Mucha. Resolving structural variability in network models and the brain. *PLoS Comput Biol*, 10(3): e1003491, 2014. doi:10.1371/journal.pcbi.1003491.

[242] Clare Kelly, Bharat B Biswal, R Cameron Craddock, F Xavier Castellanos, and Michael P Milham. Characterizing variation in the functional connectome: promise and pitfalls. *Trends Cogn Sci*, 16(3):181–188, 2012. doi:10.1016/j.tics.2012.02.001.

[243] Chao-Gan Yan, R Cameron Craddock, Xi-Nian Zuo, Yu-Feng Zang, and Michael P Milham. Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. *Neuroimage*, 80:246–262, 2013. doi:10.1016/j.neuroimage.2013.04.081.

[244] Christopher J Honey, Olaf Sporns, Leila Cammoun, Xavier Gigandet, Jean-Philippe Thiran, Reto Meuli, and Patric Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *Proc Natl Acad Sci USA*, 106(6):2035–2040, 2009.

[245] Virendra R Mishra, Karthik R Sreenivasan, Xiaowei Zhuang, Zhengshi Yang, Dietmar Cordes, Sarah J Banks, and Charles Bernick. Understanding white matter structural connectivity differences between cognitively impaired and nonimpaired active professional fighters. *Hum Brain Mapp*, 40(17):5108–5122, 2019. doi:10.1002/hbm.24761.

[246] Fumihiko Taya, Yu Sun, Fabio Babiloni, Nitish Thakor, and Anastasios Bezerianos. Brain enhancement through cognitive training: a new insight from brain connectome. *"Front Syst Neurosci*, 9:44, 2015. doi:10.3389/fnsys.2015.00044.

[247] Kate Brody Nooner, Stanley J Colcombe, Russell H Tobe, Maarten Mennes, Melissa M Benedict, et al. The nki-rockland sample. *Front Neurosci*, 6(OCT): 152, 2012. doi:10.3389/fnins.2012.00152.

[248] Jesse A Brown, Jeffrey D Rudie, Anita Bandrowski, John D Van Horn, and Susan Y Bookheimer. The ucla multimodal connectivity database: a web-based platform for brain connectivity matrix sharing and analysis. *Front Neuroinfo*, 6:28, 2012. doi:10.3389/fninf.2012.00028.

[249] R Cameron Craddock, G Andrew James, Paul E Holtzheimer III, Xiaoping P Hu, and Helen S Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Hum Brain Mapp*, 33(8):1914–1928, 2012.

[250] Kiesha Prem, Alex R Cook, and Mark Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput Biol*, 13(9): e1005697, 2017. doi:10.1371/journal.pcbi.1005697.

[251] Karl Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos Mag*, 50(302):157–175, 1900. doi:10.1080/14786440009463897.

[252] Hugh R Wilson and Jack D Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J*, 12(1):1–24, 1972.

[253] Alain Destexhe and Terrence J Sejnowski. The wilson–cowan model, 36 years later. *Biol Cybern*, 101(1):1–2, 2009.

[254] Yong He and Alan Evans. Graph theoretical modeling of brain connectivity. *Current Opin Neurol*, 23(4):341–350, 2010. doi:10.1097/WCO.0b013e32833aa567.

[255] Erzsébet Ravasz and Albert-László Barabási. Hierarchical organization in complex networks. *Phys Rev E*, 67(2):026112, 2003. doi:10.1103/PhysRevE.67.026112.

[256] Erzsébet Ravasz, Anna Lisa Somera, Dale A Mongru, Zoltán N Oltvai, and A-L Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–1555, 2002. doi:10.1126/science.1073374.

[257] Marta Sales-Pardo, Roger Guimera, André A Moreira, and Luís A Nunes Amaral. Extracting the hierarchical organization of complex systems. *Proc Natl Acad Sci, USA*, 104(39):15224–15229, 2007. doi:10.1073/pnas.0703740104.

[258] David C Van Essen and John HR Maunsell. Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci*, 6:370–375, 1983. doi:10.1016/0166-2236(83)90167-4.

[259] Victor AF Lamme, Hans Super, and Henk Spekreijse. Feedforward, horizontal, and feedback processing in the visual cortex. *Curr Opin Neurobiol*, 8(4):529–535, 1998. doi:10.1016/S0959-4388(98)80042-1.

[260] Nikola T Markov, Julien Vezoli, Pascal Chameau, Arnaud Falchier, René Quilodran, Cyril Huissoud, Camille Lamy, Pierre Misery, Pascale Giroud, Shimon Ullman, et al. Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J Comp Neurol*, 522(1):225–259, 2014. doi:10.1002/cne.23458.

[261] TA Coogan and Andreas Burkhalter. Conserved patterns of cortico-cortical connections define areal hierarchy in rat visual cortex. *Exp Brain Res*, 80(1):49–53, 1990. doi:10.1007/BF00228846.

[262] Rinaldo David D'Souza, Andrew Max Meier, Pawan Bista, Quanxin Wang, and Andreas Burkhalter. Recruitment of inhibition and excitation across mouse visual cortex depends on the hierarchy of interconnecting areas. *Elife*, 5:e19332, 2016. doi:10.7554/eLife.19332.

[263] Troy A Hackett, Lisa A de la Mothe, Corrie R Camalier, Arnaud Falchier, Peter Lakatos, Yoshinao Kajikawa, and Charles E Schroeder. Feedforward and feedback projections of caudal belt and parabelt areas of auditory cortex: refining the hierarchical model. *Front Neurosci*, 8:72, 2014. doi:10.3389/fnins.2014.00072.

[264] Moshe Abeles. *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge University Press, 1991. doi:10.1017/CBO9780511574566.

[265] Michael A Long, Dezhe Z Jin, and Michale S Fee. Support for a synaptic chain model of neuronal sequence generation. *Nature*, 468(7322):394–399, 2010. doi:10.1038/nature09514.

[266] Julie A Harris, Stefan Mihalas, Karla E Hirokawa, Jennifer D Whitesell, Hannah Choi, Amy Bernard, Phillip Bohn, Shiella Caldejon, Linzy Casal, Andrew Cho, et al.

Hierarchical organization of cortical and thalamic connectivity. *Nature*, 575(7781): 195–202, 2019. doi:10.1038/s41586-019-1716-z.

[267] Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Gorke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner. On modularity clustering. *IEEE Trans Knowl Data Eng*, 20(2):172–188, 2007. doi:10.1109/TKDE.2007.190689.

[268] Mechthild Stoer and Frank Wagner. A simple min-cut algorithm. *J ACM*, 44(4): 585–591, 1997. doi:10.1145/263867.263872.

[269] Lior Wolf, Chen Goldberg, Nathan Manor, Roded Sharan, and Eytan Ruppin. Gene expression in the rodent brain is associated with its regional connectivity. *PLoS Comput Biol*, 7(5):e1002040, 2011. doi:10.1371/journal.pcbi.1002040.

[270] Leon French and Paul Pavlidis. Relationships between gene expression and brain wiring in the adult rodent brain. *PLoS Comput Biol*, 7(1):e1001049, 2011. doi:10.1371/journal.pcbi.1001049.

[271] John A Lee and Michel Verleysen. *Nonlinear dimensionality reduction*. Springer Science & Business Media, 2007.

[272] Michael SA Graziano and Tyson N Aflalo. Mapping behavioral repertoire onto the cortex. *Neuron*, 56(2):239–251, 2007. doi:10.1016/j.neuron.2007.09.013.

[273] Maurizio Gentilucci, Leonardo Fogassi, Giuseppe Luppino, Maurizio Matelli, Rosolino Camarda, and Giacommo Rizzolatti. Functional organization of inferior area 6 in the macaque monkey. *Exp Brain Res*, 71(3):475–490, 1988. doi:10.1007/BF00248741.

[274] S Thomas Carmichael, M-C Clugnet, and Joseph L Price. Central olfactory connections in the macaque monkey. *J Comp Neurol*, 346(3):403–434, 1994. doi:10.1002/cne.903460306.

[275] P Apicella, T Ljungberg, E Scarnati, and W Schultz. Responses to reward in monkey dorsal and ventral striatum. *Exp Brain Res*, 85(3):491–500, 1991. doi:10.1007/bf00231732.

[276] Christine Baleydier and Francois Mauguiere. The duality of the cingulate gyrus in monkey. neuroanatomical study and functional hypothesis. *Brain*, 103(3):525–554, 1980. doi:10.1093/brain/103.3.525.

[277] Henry C Evrard, Thomas Forro, and Nikos K Logothetis. Von economo neurons in the anterior insula of the macaque monkey. *Neuron*, 74(3):482–489, 2012. doi:10.1016/j.neuron.2012.03.003.

[278] Jeffrey D Schall. On the role of frontal eye field in guiding attention and saccades. *Vis Res*, 44(12):1453–1467, 2004. doi:10.1016/j.visres.2003.10.025.

[279] Tullio Manzoni, Fiorenzo Conti, and Mara Fabri. Callosal projections from area sii to si in monkeys: anatomical organization and comparison with association projections. *J Comp Neurol*, 252(2):245–263, 1986. doi:10.1002/cne.902520208.

[280] Richard A Andersen, C Asanuma, G Essick, and RM Siegel. Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule. *Journal of Comparative Neurology*, 296(1):65–113, 1990. doi:10.1002/cne.902960106.

[281] Roberto Caminiti, Stefano Ferraina, and Paul B Johnson. The sources of visual information to the primate frontal lobe: a novel role for the superior parietal lobule. *Cereb Cortex*, 6(3):319–328, 1996. doi:10.1093/cercor/6.3.319.

[282] Elisabeth A Murray, Timothy J Bussey, and Lisa M Saksida. Visual perception and memory: a new view of medial temporal lobe function in primates and rodents. *Annu Rev Neurosci*, 30:99–122, 2007. doi:10.1146/annurev.neuro.29.051605.113046.

[283] Ludise Malkova and Mortimer Mishkin. One-trial memory for object-place associations after separate lesions of hippocampus and posterior parahippocampal region

in the monkey. *J Neurosci*, 23(5):1956–1965, 2003. doi:10.1523/JNEUROSCI.23-05-01956.2003.

[284] Nobuhisa Matsumura, Hisao Nishijo, Ryoi Tamura, Satoshi Eifuku, Shunro Endo, and Taketoshi Ono. Spatial-and task-dependent neuronal responses during real and virtual translocation in the monkey hippocampal formation. *J Neurosci*, 19(6):2381–2393, 1999. doi:10.1523/JNEUROSCI.19-06-02381.1999.

[285] Hristos S Courellis, Samuel U Nummela, Michael Metke, Geoffrey W Diehl, Robert Bussell, Gert Cauwenberghs, and Cory T Miller. Spatial encoding in primate hippocampus during free navigation. *PLoS Biol*, 17(12), 2019. doi:10.1371/journal.pbio.3000546.

[286] Michael J Jutras and Elizabeth A Buffalo. Recognition memory signals in the macaque hippocampus. *Proc Natl Acad Sci USA*, 107(1):401–406, 2010. doi:10.1073/pnas.0908378107.

[287] Michael Petrides. Monitoring of selections of visual stimuli and the primate frontal cortex. *Proc R Soc Lond, Ser B*, 246(1317):293–298, 1991. doi:10.1098/rspb.1991.0157.

[288] Michael Petrides. Impairments on nonspatial self-ordered and externally ordered working memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the monkey. *J Neurosci*, 15(1):359–375, 1995. doi:10.1523/JNEUROSCI.15-01-00359.1995.

[289] A Morel, PE Garraghty, and JH Kaas. Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J Comp Neurol*, 335(3): 437–459, 1993. doi:10.1002/cne.903350312.

[290] Lizabeth M Romanski and Bruno B Averbeck. The primate cortical auditory system and neural representation of conspecific vocalizations. *Annu Rev Neurosci*, 32:315–346, 2009. doi:10.1146/annurev.neuro.051508.135431.

[291] Nick E Barraclough, Dengke Xiao, Chris I Baker, Mike W Oram, and David I Perrett. Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cogn Neurosci*, 17(3):377–391, 2005. doi:10.1162/0898929053279586.

[292] Neil J Smelser and Paul B Baltes. *International encyclopedia of the social & behavioral sciences*, volume 11. Elsevier Amsterdam, 2001.

[293] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967. doi:10.1007/BF02289588.

[294] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Phys Rev E*, 78(4):046110, 2008. doi:10.1103/PhysRevE.78.046110.

[295] Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: a comparative analysis. *Phys Rev E*, 80(5):056117, 2009. doi:10.1103/PhysRevE.80.056117.