Acknowledgements

I thank my Ph.D. Thesis Supervisor, *Dr. Ganesh Bagler*, for initiating me into the area of 'network neuroscience' and guiding me throughout the course of the thesis. Apart from technical know-how of research, I have also learnt professional work ethics from him. I am thankful to the members of the Doctoral Committee, *Dr. Karunakar Kar*, *Dr. Sushmita Jha and Dr. Samanvita Pal* for their help.

I thank *Dr. Ambesh Dikshit* for his support, which helped me to overcome the tough times during the doctoral program. I thank *Dr. Snehlata Jaswal* for her insightful comments and encouragement.

My stay at the Institute was a wonderful experience because of my friends, *Megha Singh, Rakhi NK, Rakesh Kanji, Anuj K Bharti, Arun K Upadhayay and Bhuvnesh Rathore* for all the liveliness they infused into the non-academic part of the days at IIT Jodhpur.

Finally, I express my profound gratitude to my parents for providing me with infallible support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Rahul Badhwar
Ph.D. Student

List of Figures

Figures	*Title*	*page*
1.1	Evolution in structural complexity of nervous systems. (a) In primitive organisms such as a Hydra (a coelenterate), neurons are connected in an unorganized manner. (b) In nematodes such as *C. elegans*, the nervous system is more nuanced with specialized structures like nerve cords and ganglions.	1
1.2	Nervous systems seen as a meta-system comprising of different layer of sub-systems. Systems biological investigations conducted as part of this thesis invoke these towards answering questions addressing structure, function and control of brain networks.	2
2.1	The structure of a simple neuronal circuit required to respond to an external stimulus.	6
2.2	Different network motifs prevalent in biological systems. The feed-forward loop, bi-fan and biparallel are over-represented, whereas feedback loop is under-represented in gene regulatory networks and neuronal connectivity networks [R Milo et al., 2002]. Figure adapted from [Tran, Choi, and Zhang, 2013].	8
2.3	Illustrations explaining structural control in graphs. (a) Structurally controllable cacti structure where the driver node is represented in red. (b) Presence of dilations and inaccessible nodes makes the network structurally uncontrollable.	9
2.4	The labelled four node complete graph. Red edges represent animosity and green edges represent positive relationship. Figure adapted from Chapter 5 [Easley and Kleinberg, 2010].	11
2.5	Temporal and spatial resolution of different brain imaging techniques. Figure adapted from [Meyer-Lindenberg, 2010].	12
2.6	fMRI technique assumes causal connection between 'neural activity' and 'BOLD signal'.	12
3.1	*C. elegans* neuronal network: A network representation of *C. elegans* nervous system.	15
3.2	Strategy to construct degree distribution preserved control. The source nodes are shown in green, and the target nodes are shown in red. Other nodes of the network are shown in blue.	16
3.3	Classification of three node sub-graphs.	17
3.4	Hopcroft-Karp algorithm for maximum matching (adapted from ["Hopcroft–Karp algorithm," 2017]). (a) A directed graph represented as bipartite with *U* and *V* subsets. (b) Perform BFS starting at all the vertices in V without a match. Pick any unmatched leaf and go all the way back to a root using DFS. Match the leaf to the root. Repeat the process to find the next matching is observed. (c) Delete all the instances of 1 and a found in the trees. (d) Match b to 2 and delete b from the tree spanning from 3. (e) In this iteration, 1 is matched to a, 2 is matched to b, and 3, along with c, are left without a match. (f) Perform DFS once again from the unmatched leaf all the way to the root. (g) Thus produced the maximal matching between *U* and *V* in the graph. The yellow edge represent matching edge red hops represent depth first search steps.	20
3.5	Schematic representation of Augment matching algorithm [Pothen and Fan, 1990].	22
3.6	Generic framework for creating a functional network from 4D fMRI data.	23
4.1	Diagrammatic representation of *C. elegans* nervous system. The gastrointestinal tract lies in the middle of the body. Pharyngeal and circumferential ring neurons (yellow) are responsible for communication between dorsal nerve cord and ventral nerve cord.	26
4.2	The Gene Co-expression Network for CeNN. Each of the 116 neuronal groups is represented as a node. A weighted edge between any two neuronal groups represents the extent of gene co-expression. The shape and colour of the neuronal groups depict presence (Red Diamonds) or absence (Blue Circles) of driver neurons in them. The *C. elegans* co-expression based network is heterogeneous with 9 groups holding all the driver neurons, whereas rest of the 107 groups were devoid of driver neurons.	27
4.3	Role of degree in specifying driver nodes. (a) Fraction of driver nodes as found in the CeNN, and their corresponding random counterparts: Degree Distribution conserved control (DD) and Erdos-Reney graph (ER). (b) Fraction of driver neurons with low, medium and high degree in CeNN. These results are consistent with what is reported by Lui et. al [Y.-Y. Liu et al., 2011].	28
4.4	Distribution of driver neurons across different types of neurons: Sensory, Motor and Inter. The fraction of driver neurons was computed for each class separately.	29

4.5 Distribution of driver neurons across other phenotypic features (location and span). (a) Location of neurons within the body of the organism. (b) Span of neurons in accordance with the length of axons. 30

4.6 Visualization of driver neurons in CeNN and distribution of driver neurons across the body. The neurons are arranged in accordance with Cartesian coordinates presented within the body. This figure clearly shows presence of driver neurons in the mid-ventral region of the organism. 30

4.7 Nine neuronal groups (clusters) obtained starting from GCN using affinity propagation algorithm [Morris et al., 2011]. The three largest clusters contain driver nodes (red diamonds) hinting at possible role they play. Cluster three forms a perfect clique with highest number of unique genes expressed within a cluster. 31

4.8 Phenotypic distribution of neurons in the clusters of GCN in accordance with, (a) functional types, (b) location of the body, (c) Span of the neuronal axon. 32

5.1 Network structure of *C. elegans* nervous system. Functionally relevant driver neurons (34 nodes highlighted in white) were identified with maximum matching criterion. Beyond explaining the small world nature, saturation of feed forward motifs and observed number driver neurons, the distance constrained synaptic plasticity model accurately identifies specific driver neurons. 36

5.2 Statistics for unidirectional three node motifs depicting over-representation of feed forward motifs. The Z-Score was computed in comparison to 100 instances of random controls (ER) of CeNN. 37

5.3 The 1D ring model, with neurons linked for maximizing number of feed forward motifs, was rewired with increasing probability of synaptic rewiring. Starting with an asymptotic model (with 277 nodes and 8 out-going edges) saturated with FFMs, synaptic rewiring was emulated with probability p. The model exhibits a spectrum of topological variations between extreme regularity and randomness. The figure shows an illustration for 10 nodes and 2 outgoing edges. See Figure C.1 of Annexure C for another illustration with larger network size. 38

5.4 Response of 1D ring model with changing probability of synaptic rewiring was measured in terms of average clustering coefficient ($\bar{C}$), characteristic path-length (L), number of FFMs (n_{FFM}) and number of driver nodes (n_D). For intermediate values of p, the model exhibits small world phenomenon as well as FFM saturation, but cannot account for controllability ($n_D = 0 \; \forall \; p$). All parameters were normalized with respect to the initial ring graph ($p = 0$). Error bars represent standard deviation over 100 instances. Please see Figure C.2, Figure C.3 and Figure C.4 in Annexure C for non-normalized data. 39

5.5 Empirical distance constrained synaptic connectivity pattern observed in *C. elegans* neuronal wiring. The number of synapses that connect neurons at distance d follows a power law pattern with an exponent of $\alpha = 2.02$ ($p - value = 0.92$) [Clauset, Shalizi, and Newman, 2009]. 40

5.6 Response of distance constrained models of CeNN (DCR and DCP) with increasing constraint (α) measured in terms of (a) average clustering coefficient ($\bar{C}$), (b) characteristic path-length (L), (c) number of FFMs (n_{FFM}), and (d) number of driver nodes (n_D). The lower the α more heterogenous are the synaptic lengths (larger proportion of long range synapses). For $\alpha = 0$, DCR and DCP models converge to ER and DD controls, respectively. For asymptotic limits of $\alpha < \infty$ both the models converge to the Cartesian model, a regular model with saturation of FFMs coupled with high clustering but devoid of driver nodes. While the small world nature (reflected in high clustering and low path-length) and FFM saturation is realized by both DCR and DCP models, DCP model stands out in reproducing all key features of CeNN for $0.6 < \alpha < 1.8$ (highlighted with gray background). Please see Figure C.5, Table C.6 and Table C.7 of Annexure C for data associated with this figure. 42

5.7 Accuracy of identification of specific driver neurons with changing distance constraint exponent. The DCR model, with random synaptic connectivity pattern, fared poorly. The performance of DCP model was consistently better than that of the DCR model indicating the critical role played by the distance constraint in specifying the control of the neuronal network. For optimum distance constraint ($\alpha = 0.6$) DCP model provides the best match with the reality ($F1 \; score = 0.77$), better than what could be accounted for by only neuronal connectivity (DD control; indicated with a dashed line). The spectrum of distance constraint regime for which DCP model is closest to CeNN ($0.6 \leq \alpha \leq 1.8$) is highlighted with gray background. 43

6.1	Motif tuning algorithm. Strategy implemented to (a) increase the number of feed-forward motifs and to (b) decrease the number of feed-forward motifs.	46
6.2	Feed-forward motifs are significantly over-represented in CeNN, followed by feedback motifs, as measured in terms of Z-score against a background of random networks.	48
6.3	Asymmetric controllability response (enumerated with the number of driver neurons, D_n) of *C. elegans* neuronal network with monotonic increase/decrease in number of FFMs (n_{FFM}). The random control (ER), on the other hand, does not show any change in D_n. This implies that while CeNN exhibits robust control response to systematic increase in FFMs, it is extremely sensitive to systematic depletion of FFMs. Dashed lines represent the starting points for the models. Error bars indicate standard deviation over 1000 instances.	49
6.4	Change in n_{FFM} per rewiring for different strategies. DC and MTA+ show a positive Δn_{FFM} whereas random and MTA- were presented with negative Δn_{FFM}. This implies that MTA+ and distance constrained rewiring lead to robust response.	50
6.5	Robust control response of CeNN under distance constrained rewiring. (a) With increasing extent of rewiring, the number of FFMs are preserved. The stronger the constraint, the better is the response. (b) Correspondingly, the number of driver nodes is preserved with distance constrained rewiring. (c) With increasing power law exponent the average change in number of FFMs (Δn_{FFM}) is diminished.	51
7.1	Visualization of structural balance on the basis of positive and negative edges. Green edges represents a positive relationship, red edges represents a negative relationship between two nodes.	53
7.2	Diagrammatic depiction of signed graph creation.	55
7.3	Visualization of signed graph. (a) C. elegans neuronal network signed graph; green edges represents positive (excitatory) edge red edge represents negative (inhibitory) edge. (b) Positively signed subgraph. (c) Negatively signed subgraph.	57
8.1	Strategy implemented for modeling 4-D fMRI data as brain functional network and its topological characterization.	62
8.2	Detailed procedure implemented for pre-processing of raw fMRI data, creation of brain functional network of independent components and its characterization using graph theoretical metrics.	64
8.3	Normalized distance matrices and their corresponding weighted BFNs for a representative healthy control (40018) and schizophrenic patient (40009). In the distance matrices lighter gray colours represent lower distances thus higher correlation. The weighted BFNs with 30 components (nodes) show nodes sizes scaled to 'clustering coefficient'. The larger the node, higher is its clustering coefficient. The networks were visualized using Cytoscape 3.2.0 [Shannon et al., 2003].	65
8.4	Percentage contribution of features with combined discrimination accuracy above 50% in binary and weighted BFNs (Table D.1 of Annexure D). 'Others' represents remaining parameters in both types of BFNs.	66
8.5	Best accuracy obtained for classification between BFN models of healthy and schizophrenia brain data with increased order of feature combinations. The error bars represent standard error for 100 experiments of 10-cross fold tests.	67
C.1	Regular graph for maximum number of Feed forward motifs with graph growth. Edges are directed and point from nodes to nearest neighbours along with neighbour of neighbours and so on. This representation is for $n = 20$ and $k = 4$.	81
C.2	Response of average clustering coefficient to rewiring.	82
C.3	Response of characteristic path-length to rewiring.	82
C.4	Response of number of feed forward motifs to rewiring.	83
C.5	Degree distributions of CeNN, its controls and distance constrained models.	83

List of Tables

Table	*Title*	*page*
4.1	List of gene obtained within unique set of genes after gene ontological enrichment of each clusters.	33
5.1	Topological properties of CeNN and its controls.	37
7.1	Neurotransmitter list and their actions performed in *C. elegans* neuronal system. Data collected from WormAtlas [Altun, Z.F., Herndon, L.A., Crocker, C., Wolkow, C.A., Lints, R. and Hall, 2016].	54
7.2	Graph theoretical properties of CeNN and signed subgraphs.	56
A.1	Neurons of *C. elegans*. These neurons could be divided into somatic (277; white boxes), pharyngeal (20; yellow boxes) and unconnected (5; grey boxes). For studies presented in Chapter 4, somatic as well as pharyngeal neurons (297) were used. For the remaining studies, presented in Chapter 5, Chapter 6, and Chapter 7, somatic neurons were considered.	71
B.1	Unique genes expressed in each neuronal group. 116 neuronal groups and 433 non redundant genes.	73
C.6	Differential topological properties of distance constraint random model at different values of α starting from an ER control.	84
C.7	Differential properties of distance constraint synaptic plasticity model at different values of α starting from DD control.	85
C.8	List of 34 driver neurons identified.	85
D.1	Network based features of each subject that will be used for classification between two classes.	87

List of Symbols

Symbol	*Description*
$A_{t\mu}$	Amplitude of component μ at time t
B_i	Balance Index
$\bar{C}$	Average clustering coefficient
C_b	Balanced cycles
$\overline{C_{reg}}$	Average Clustering coefficient of a regular graph
C_{ub}	Unbalanced cycles
$C_{\mu s}$	Magnitude of component μ at space point s
D_n	Driver Neurons
E	Edges
G	*C. elegans* genes
G_e	Excitatory genes
G_i	Inhibitory genes
G_n	Genes expressed only in neurons
L	Characteristic path-length
L_{reg}	Characteristic path-length of regular graph
M	Matching
n_D	Number of driver neurons
n_{FFM}	Number of feed forward motifs
Δn_{FFM}	Change in number of feed forward motifs
S_p	Potential of synapse
U_g	Unique genes
V	Vertex
X_{ts}	Raw scan intensity at time t and space point s

List of Abbreviations

Abbreviation	*Full form*
1D	One dimensional
2D	Two dimensional
4D	Four dimentional
AFNI	Analysis of Functional NeuroImages
BFN	Brain Functional Networks
BiNGO	Biological Networks Gene Ontology Tool
BOLD	Blood Oxygen Level Dependent
CCF	Cross-Correlation Function
CeNN	*C. elegans* Neuronal Network
COBRE	Centers of Biomedical Research Excellence
DCP	Distance Constraint Synaptic Plasticity
DCR	Distance Constraint Random
DD	Degree Distribution Conserved
DNC	Dorsal Nerve Cord
DSM	Diagnostic and Statistical Manual of Mental Disorders
EEG	Electro Encephalo Gram
ER	Erdős–Rényi random control
FFM	Feed Forward Motif
fMRI	functional Magnetic Resonance Imaging
FN	False Negative
FP	False Positive
FSL	fMRI Software Library
GCN	Gene Co-expression Network
GO	Gene Ontology
ICA	Independent Component Analysis
MEG	Magnetoencephalography
MTA	Motif Tuning Algorithm
PET	Positron Emission Tomography
R-fMRI	Resting-state fMRI
SVM	Support Vector Machine
TE	Echo Time
TN	True Negative
TP	True Positive
TR	Repetition Time
VNC	Ventral Nerve Cord