False Discovery Rate Controller for Functional Brain Parcellation Using Resting-State fMRI

by

Adrian Kwok-Hang Wong

B.A.Sc., The University of British Columbia, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Biomedical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2016

© Adrian Kwok-Hang Wong, 2016
Abstract

Parcellation of brain imaging data is desired for proper neurological interpretation in resting-state functional magnetic resonance imaging (rs-fMRI) data. Some methods require specifying a number of parcels and using model selection to determine the number of parcels with rs-fMRI data. However, this generalization does not fit with all subjects in a given dataset. A method has been proposed using parametric formulas for the distribution of modularity in random networks to determine the statistical significance between parcels. In this thesis, we propose an agglomerative clustering algorithm using parametric formulas for the distribution of modularity in random networks, coupled with a false discovery rate (FDR) controller to parcellate rs-fMRI data. The proposed method controls the FDR to reduce the number of false positives and incorporates spatial information to ensure the regions are spatially contiguous. Simulations demonstrate that our proposed FDR-controlled agglomerative clustering algorithm yields more accurate results when compared with existing methods. We applied our proposed method to a rs-fMRI dataset and found that it obtained higher reproducibility compared to the Ward hierarchical clustering method. Lastly, we compared the normalized total connectivity degree of each region within the motor network between normal subjects and Parkinson’s disease (PD) subjects using sub-regions defined by our proposed method and the entire region. We found that PD subjects without medication had a significant increase in functional connectivity compared to normal subjects in the right primary motor cortex using our sub-regions within the right primary motor cortex, whereas this significant increase was not found using the entire right primary motor cortex. These sub-regions are of great interest in studying the differences in functional connectivity between different neurological diseases, which can be used as biomarkers and may provide insight in severity of the disease.
Preface

This thesis is based on one accepted conference paper:


The majority of this research, including literature review, model design, algorithm implementation, numerical simulations and draft writing, was conducted by the author of this thesis, with suggestions from Dr. Z. Jane Wang, Dr. Mehdi Moradi, and Dr. Martin J. McKeown. Dr. Martin J. McKeown and SunNee Tan helped with the data collection, result interpretation and paper revision. Dr. Z. Jane Wang helped greatly on revisions of the journal manuscript. Ethics approval was obtained for this research. This thesis is based on the data obtained for a Parkinson’s research study titled Making the Connection: Methods to Infer Functional Connectivity in Brain Studies (PI: M. Mckweon). This study was approved by the Clinical Research Ethics Board (CREB) of the University of British Columbia (UBC CREB number: H04-70177)
# Table of Contents

Abstract ................................................................. ii
Preface ................................................................. iii
Table of Contents ...................................................... iv
List of Tables ........................................................... vi
List of Figures ........................................................... vii
Glossary ................................................................. x
Acknowledgements ...................................................... xiv

Chapter 1: Introduction .................................................. 1
  1.1 Background ......................................................... 1
  1.2 Motivation .......................................................... 5
  1.3 Objective ........................................................... 8
  1.4 Contributions ....................................................... 8
  1.5 Thesis Outline ....................................................... 9

Chapter 2: Review of Related Work ................................. 10
  2.1 Overview of Functional Connectivity Analysis Using Resting-State fMRI ........................................... 10
  2.2 Functional Brain Parcellation Techniques for Resting-State fMRI ...................................................... 12

Chapter 3: Method ......................................................... 16
  3.1 Parcel Initialization ............................................... 16
  3.2 Statistical Significance between Parcels ....................... 18
  3.3 False Discovery Rate Controller ............................... 18
# TABLE OF CONTENTS

3.4 Accuracy of Representation .................................................. 20  
  3.4.1 Adjusted Rand Index .................................................. 20  
  3.4.2 Functional Connectivity Analysis ................................. 21  

Chapter 4: Results ................................................................. 22  
  4.1 Synthetic Dataset .......................................................... 22  
  4.2 fMRI Dataset ................................................................. 25  
    4.2.1 Data Acquisition .................................................... 25  
    4.2.2 Data Preprocessing ................................................ 25  
    4.2.3 fMRI Parcellation Results ....................................... 31  
    4.2.4 Functional Connectivity Results ............................... 41  

Chapter 5: Conclusion and Future Work .................................. 52  
  5.1 Discussions and Conclusions .......................................... 52  
  5.2 Future Work ............................................................... 53  

Bibliography ................................................................. 54
List of Tables

Table 3.1  Error control criteria for multiple testing with notation shown in Table 3.2. ........................................... 19
Table 3.2  The number of errors committed when testing m hypotheses. ......................................................... 19
Table 4.1  Our results when comparing our method with the Ward clustering method and the method proposed by Chang et al. method [8] using synthetic data. ....................... 24
Table 4.2  Demographic data of normal and PD subjects. ........ 25
Table 4.3  The cortical and subcortical ROIs extracted from the left hemisphere of the brain. ............................. 29
Table 4.4  The cortical and subcortical ROIs extracted from the right hemisphere of the brain. ............................. 30
Table 4.5  The left and right motor regions of the brain chosen for functional connectivity analysis. ..................... 41
## List of Figures

| Figure 1.1 | An example of a hemodynamic response function. | 3 |
| Figure 1.2 | Oxyhaemoglobin and deoxyhaemoglobin blood flow during rest and activation. | 3 |
| Figure 1.3 | BOLD signal mechanism for fMRI. | 4 |
| Figure 1.4 | fMRI scans collected over time from a fixed axial location of the brain. | 4 |
| Figure 1.5 | An illustration of the divisions of the basal ganglia. | 6 |
| Figure 1.6 | An illustration of the locations of the motor regions used for our functional connectivity analysis. | 7 |
| Figure 2.1 | Clusters within a network are compared against the expected null network. | 14 |
| Figure 4.1 | An example of the parcellation results from synthetic dataset 3 with an intra-correlation equal to 0.65 and an inter-correlation equal to 0.30. | 23 |
| Figure 4.2 | An illustration of a two-dimensional patch of nine voxels clustered into three regions. | 24 |
| Figure 4.3 | Examples of MRI scans collected from a normal subject. | 26 |
| Figure 4.4 | Examples of fMRI scans collected which spatially corresponds to the MRI scans in Figure 4.3. | 26 |
| Figure 4.5 | Examples of the gray matter mask which corresponds to the fMRI scans in Figure 4.4. | 27 |
| Figure 4.6 | Boxplots of the adjusted Rand index showing the reproducibility of the results using our method and the Ward agglomerative clustering using normal subjects. | 31 |
| Figure 4.7 | Boxplots of the adjusted Rand index showing the reproducibility of the results using our method and the Ward agglomerative clustering using PD subjects. | 32 |
FIGURE 4.8 The parcellation results using the proposed FDR-controlled agglomerative clustering algorithm on the left cerebellum on normal subjects. 33

FIGURE 4.9 The parcellation results using the proposed FDR-controlled agglomerative clustering algorithm on the left cerebellum on PD subjects. 34

FIGURE 4.10 The parcellation results using the proposed FDR-controlled agglomerative clustering algorithm on the left primary motor cortex on normal subjects. 35

FIGURE 4.11 The parcellation results using the proposed FDR-controlled agglomerative clustering algorithm on the left primary motor cortex on PD subjects. 36

FIGURE 4.12 Parcellation results for subject N003. 37

FIGURE 4.13 Parcellation results for subject N005. 38

FIGURE 4.14 Parcellation results for subject P001. 39

FIGURE 4.15 Parcellation results for subject P003. 40

FIGURE 4.16 The normalized total connectivity degree $\bar{\Gamma}$ of the left hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using ROIs defined by our pre-processing pipeline. 42

FIGURE 4.17 The normalized total connectivity degree $\bar{\Gamma}$ of the right hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using ROIs defined by our pre-processing pipeline. 43

FIGURE 4.18 The difference in the normalized total connectivity degree $\bar{\Gamma}$ in each motor-related brain region using rs-fMRI between normal and PD subjects using ROIs defined by our pre-processing pipeline. 43

FIGURE 4.19 The normalized total connectivity degree $\bar{\Gamma}$ of the left hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using nodes defined by the largest cluster of each region and the ROIs defined by our preprocessing pipeline. 45

FIGURE 4.20 The normalized total connectivity degree $\bar{\Gamma}$ of the right hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using nodes defined by the largest cluster of each region and the ROIs defined by our preprocessing pipeline. 45
Figure 4.21  The difference in the normalized total connectivity degree \( \bar{\Gamma} \) in each motor-related brain region using rs-fMRI between normal and PD subjects using nodes defined by the largest cluster of each region and the ROIs defined by our preprocessing pipeline. 46

Figure 4.22  Example transformations of small clusters into a seed cluster. 47

Figure 4.23  The normalized total connectivity degree \( \bar{\Gamma} \) of the left hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using nodes defined by the seeds and the ROIs defined by our preprocessing pipeline. 48

Figure 4.24  The normalized total connectivity degree \( \bar{\Gamma} \) of the right hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using nodes defined by the seeds and the ROIs defined by our preprocessing pipeline. 49

Figure 4.25  The difference in the normalized total connectivity degree \( \bar{\Gamma} \) in each motor-related brain region using rs-fMRI between normal and PD subjects using nodes defined by the seeds and the ROIs defined by our preprocessing pipeline. 49

Figure 4.26  Correlation between the UPDRS and normalized total connectivity degree \( \bar{\Gamma} \) at each node defined by the ROIs of each motor region in PD subjects. 50

Figure 4.27  Correlation between the UPDRS and normalized total connectivity degree \( \bar{\Gamma} \) at each node defined by the largest cluster of each motor region in PD subjects. 51

Figure 4.28  Correlation between the UPDRS and normalized total connectivity degree \( \bar{\Gamma} \) at each node defined by the seed cluster of each motor region in PD subjects. 51
Glossary

**BOLD**: Blood Oxygen Level Dependent. A method used in fMRI to observe different areas of the brain.

**Degree Distribution**: It measures the connectivity of a node with the rest of the nodes in a network.

**Diamagnetic**: These types of materials create an induced magnetic field in a direction opposite to an externally applied magnetic field, and are repelled by the applied magnetic field.

**DTI**: Diffusion Tensor Imaging. It enables mapping of the diffusion process of molecules in biological tissues, in-vivo and non-invasively.

**ECoG**: Electrocorticography. An invasive type of electrophysiological monitoring that uses electrodes placed directly on the exposed surface of the brain to record electrical activity from the cerebral cortex.

**EEG**: Electroencephalography. An electrophysiological monitoring method to record electrical activity of the brain.

**FDR**: False Discovery Rate. The expected percent of false predictions in the set of predictions.

**FN**: False Negative. A test result indicates that a condition failed, while it actually was successful.

**FP**: False Positive. A test result indicates that a condition failed, while it actually was not successful.

**FWER**: Family-Wise Error Rate. The probability of making one or more false discoveries among all hypotheses when performing multiple hypotheses tests.

Haemoglobin: A red protein responsible for transporting oxygen in the blood.

HMAT: Human Motor Area Template. A template used to determine motor regions of the brain.

ICA: Independent Component Analysis. A statistical method that separates an observed set of signal mixtures into a set of statistically independent component signals.

LDDMM: Large Deformation Diffeomorphic Metric Mapping. Algorithms used for mapping, manipulating, and transferring information in spatially distributed medical imagery based on diffeomorphic metric mapping.

MEG: Magnetoencephalography. A functional neuroimaging technique for mapping brain activity by recording magnetic fields produced by electrical currents occurring naturally in the brain, using very sensitive magnetometers.

Modularity: A measure of the degree to which a network is organized into a modular or community structure.

MRI: Magnetic Resonance Imaging. Uses magnetic fields and pulses of radio wave energy to create structural images of inside the body.

Network efficiency: It measures the ability of a network to transmit information at a global and local level.

Nodal centrality: It measures the importance of a node within a network.

NCUT: Normalized Cut. A spectral clustering technique which optimizes within cluster similarity and between cluster dissimilarity. [38]
**Glossary**

**Paramagnetic**: A form of magnetism whereby certain materials are attracted by an externally applied magnetic field, and form internal, induced magnetic fields in the direction of the applied magnetic field.

**PCA**: Principal Component Analysis. A statistical technique that transforms a large number of variables to a smaller number of variables while minimizing the loss of information.

**PET**: Positron Emission Tomography. A functional imaging technique that is used to observe metabolic processes in the body.

**PD**: Parkinson’s Disease. A degenerative disorder of the central nervous system mainly affecting the motor system.

**ReHo**: Regional Homogeneity. It measures the functional coherence of a given voxel with its nearest neighbors.

**ROI**: Region of Interest. Anatomical regions of cortical and subcortical areas of the brain.

**rs-fMRI**: Resting-state functional Magnetic Resonance Imaging. A fMRI study that determines brain connectivity while the subject is quietly awake and alert but does not engage in a specific cognitive or behavioral task.

**Small-world networks**: It characterizes the strength of interconnected sub-networks.

**Sensitivity**: Sensitivity measures the proportion measures the proportion of positives that are correctly identified.

**Specificity**: Specificity measures the proportion of negatives which are correctly identified in a classification test.

**SPM**: Statistical Parametric Mapping. It refers to the construction and assessment of spatially extended statistical processes used to test hypotheses about functional imaging data.

**SNR**: Signal-to-Noise Ratio. The ratio of signal power to noise power.
Glossary

**Talairach**: A brain atlas developed by a neurosurgeon named Jean Talairach.

**Task-based-fMRI**: A fMRI study to determine brain connectivity while the subject performs specific cognitive or behavioral tasks.

**TP**: True Positive. A test result indicates that a condition passed, while it actually was successful.

**TN**: True Negative. A test result indicates that a condition failed, while it actually was failed.

**UPDRS**: Unified Parkinson’s Disease Rating Scale. This scale is used to determine progression of Parkinson’s disease in a patient.
Acknowledgements

I would like to express my appreciation to my supervisors, Dr. Z. Jane Wang and Dr. Mehdi Moradi, who has supported me throughout my M.A.Sc. degree. I would like to thank you Dr. Mehdi Moradi for introducing me into the area of research during my undergraduate degree and nominating me for the Engineering in Scrubs (EiS) Graduate Student Award 2014. I am very grateful for the Biomedical Engineering department at UBC for awarding me with the EiS fellowship and provided me funding during the first year of my M.A.Sc. program. Lastly, I would like to thank Dr. Martin J. McKeown and his lab for providing the biomedical data for my research and insight into my results.
Chapter 1

Introduction

1.1 Background

The brain is the most complex organ in the human body. Due to its complexity, there has been lots of development in neuroimaging modalities such as magnetic resonance imaging (MRI), positron emission tomography (PET), electroencephalography (EEG), electrocorticography (ECoG), functional magnetic resonance imaging (fMRI), and magnetoencephalography (MEG) to observe and measure the activity and the anatomical changes in the brain. The above imagining modalities has been used to analyze the structural, functional, and effective connectivity of the brain.

Structural connectivity facilitates in identifying a network of anatomical links of the axon bundles (white matter) using MRI techniques such as diffusion tensor imaging (DTI). In contrast, functional connectivity identifies temporal correlations in neuronal cell bodies (gray matter) within the cortical and subcortical regions. Effective connectivity identifies the direct and causal relationships between cortical and subcortical regions of the brain. EEG, ECoG, fMRI, MEG and PET have been used to measure the functional connectivity and effective connectivity of the brain.

EEG is non-invasive, inexpensive and easily accessible compared to other modalities. EEG follows the internationally standardized 10-20 system to record brain activity where 21 electrodes are placed on the scalp to record electrical signals. These signals are correlated with each other to determine functional connectivity maps. EEG has high temporal resolution and low spatial resolution. Moreover, EEG only provides an overall activity between each section of the brain. ECoG is similar to EEG, except that ECoG is invasive. In ECoG, electrodes are placed directly on the surface of the brain to record the electrical activity. ECoG is not appropriate given the accuracy of non-noninvasive devices above. MEG detects the magnetic fields produced by electrical currents occurring naturally in the brain. MEG has been shown to have may be superior to EEG especially for functional connectivity studies. However, MEG studies are expensive and it is difficult to register MEG with structural MRI in comparison to EEG.
and MEG are used for connectivity analysis, whereas PET and fMRI can be used for both connectivity analysis and brain parcellation. In PET, radioactive tracers (radioactive medicine and natural chemicals) are injected into the subject’s body. The tracer is made such that it would react with a certain locations of the brain given the type of tracer used. fMRI collects 3 dimensional (3D) volumes of the brain over time by measuring the blood oxygen level in the neurons. fMRI provides superior spatial resolution, but suffers from low temporal resolution. Despite this drawback, higher spatial resolution can lead to new discoveries. Studies have discovered new findings in the functional connections of specific brain regions and local networks, as well as insights in the overall organization in the brain using fMRI. [43]

In 1948, Seymour Kety and Carl Schmidt showed that oxygen metabolism and blood flow in the brain proved to be regulated by brain activities. [24] In 1990, one of the significance of this study was shown when Seiji Ogawa and Ken Kwong developed an fMRI technique and showed the potential to visualize neural activity. [32] Neural activity is represented by a haemodynamic response function (HRF) as shown in Figure 1.1. Neuronal activity increases when there is an increased demand for oxygen in the neurons. Oxygen is delivered to neurons by haemoglobin in capillary red blood cells. Haemoglobin is diamagnetic when oxygenated, but paramagnetic when deoxygenated as illustrated in Figure 1.2. This creates a difference in magnetic fields, which leads to small differences in the MR signal of blood depending on the degree of oxygenation. A diagram highlighting the effects of oxygenation in neurons is shown in Figure 1.3. This form of MRI is known as blood oxygenation level dependent (BOLD) contrast imaging.

fMRI scans the head to obtain 3D volumes of the brain over time. These scans contain voxels which stores information on the neuronal activity of the brain at a specific time point. These voxels are used to form BOLD fMRI signals as shown in Figure 1.4. In Figure 1.4, two BOLD time-series are shown in green and blue which are collected from two neighboring voxels. Note the similarity in the time-series between two neighboring voxels.

BOLD signals are composed of HRF impulses, hardware noise and physiological noise. Hardware noises are caused by the fMRI machine while it is in use. Thermal noise is caused by heat increases, which increases the movement of electrons and distorts the current in the fMRI machine. Another source of variability in the BOLD signal is caused by superconducting magnetic field drifting over time which is known as scanner drift. Physiological noise is due to motion of the subject’s head, respiration and heartbeat.

Functional connectivity between brain regions can be determined using BOLD signals from either resting-state fMRI (rs-fMRI) or task-based fMRI.
1.1. Background

Figure 1.1: An example of a hemodynamic response function. This figure is from [41].

Figure 1.2: Oxyhaemoglobin and deoxyhaemoglobin blood flow during rest and activation. This figure is from [10].
1.1. Background

Figure 1.3: BOLD signal mechanism for fMRI.

Figure 1.4: fMRI scans collected over time from a fixed axial location of the brain. The Y-axis is the strength of the BOLD signal.
1.2. Motivation

In task-based fMRI, researchers seek to identify areas in the brain that are related to a sensory stimulus, motor act, or cognitive process. Task-based fMRI requires a specific experimental design and patient will have a varying levels of activity in the brain. Rs-fMRI can reveal synchronous activities between different brain regions that occur in the absence of a task or stimulus, and it may include spontaneous low frequencies fluctuations (<0.1 Hz) that can be used to determine functional regions of the brain. [43] One benefit of using rs-fMRI is the potential to reduce confounding factors like inter-individual variability in task compliance and/or performance during fMRI acquisition. [16] Thus, rs-fMRI is preferred for functional connectivity study in some cases. Identification of functional regions in the brain facilitates discovery of previously unknown connections and components which provide insight into brain function of people with neurological diseases such as Parkinson’s disease (PD). [17, 26]

1.2 Motivation

An estimate of over seven million people worldwide are living with PD.¹ PD is a progressive neurodegenerative disorder characterized by severe motor symptoms including uncontrollable tremor, postural imbalance, slowness of movement and rigidity. PD is a chronic and slowly progressive disease that usually affects men and woman over the age of 50. A person begins to notice the effects of PD when s/he loses a significant amount of dopamine-producing neurons in the substantia nigra pars compacta, which results in a drastic depletion of dopamine in the basal ganglia. Figure 1.5 shows the location of the basal ganglia. A scale commonly used to assess the progression of PD is the United Parkinson’s Disease Rating Scale (UPDRS). The UPDRS is split into seven parts and measures the following:

i evaluates the mentation, behavior, and mood;

ii self-evaluation of the activities of daily life (ADL) including speech, swallowing, handwriting, dressing, hygiene, falling, salivating, turning in bed, walking, and cutting food;

iii monitored motor evaluation;

iv complications of therapy;

v clinician-scored;

vi Hoehn and Yahr staging of movement symptoms;

1.2. Motivation

Extensive neuroimaging studies have been performed, and commonly found hypoactivation in the supplementary motor area (SMA) and putamen. Hyperactivation in the cerebellum, premotor area (PMA), and parietal cortex during movements in PD compared to normal subject. [33, 36, 46, 48] Recently, it has been shown that the putamen region could be subdivided into two regions. [20, 50] We postulate that other regions in the brain could be subdivided into sub-regions instead of the regions shown in Figure 1.6. By finding consistent functional sub-regions, we can isolate the regions that are affected most by PD.

![Figure 1.5: An illustration of the divisions of the basal ganglia. This figure is from [5].](image)

There have been studies to determine functional brain parcellation [13, 37, 42]. However, these techniques requires model selection techniques to determine the optimal number of parcels. Even if model selection techniques
Figure 1.6: An illustration of the location of the motor regions used for our functional connectivity analysis. The numbers represent the axial slice location.
used, there is still a discrepancy on the number of parcels depending on the model that is used. [42] Chang et al. developed a significance-based modularity partitioning algorithm that uses statistical criteria to define the number of parcels. This algorithm was used to determine networks based on the mean time-series of ROIs which is a not meaningful if the study is to determine whether regions could be subdivided into smaller regions. Moreover, the significance-based modularity partitioning algorithm is a top-down divisive clustering algorithm to partition the network which can be computationally intensive for a large network size \( (N \geq 20000) \). [8]

1.3 Objective

The objective of the research in this thesis is to use graph theory and statistical methods to develop a functional brain parcellation algorithm using rs-fMRI data. Our approach is to use an agglomerative clustering algorithm that uses parametric formulas for the distribution of modularity, coupled with a false discovery rate (FDR) controller. We believe that using our agglomerative clustering algorithm, we can find similar and different functional regions between normal and PD subjects. Moreover, we will use the parcels to determine the differences in functional connectivity between normal and PD subjects.

1.4 Contributions

The major contributions of this work are:

- We extended the work in [8] by using an agglomerative clustering algorithm and controlled the FDR, whereas [8] used a divisive clustering method and controlled the family-wise error rate to parcellate rs-fMRI data using the mean of value of each region.

- We applied our FDR-controlled agglomerative clustering algorithm to a fMRI dataset and found there is a significant increase in the normalized total connectivity degree \( \bar{\Gamma} \) in the right primary motor cortex in PD subjects compared to normal subjects using subregions within the right primary motor cortex, whereas this significant increase was not found using the entire right primary motor cortex.

- We found that certain regions of the motor network had an increase in correlation between the normalized total connectivity degree \( \bar{\Gamma} \) and
the UPDRS in different sub-regions of the brain for PD subjects when compared to the entire region.

1.5 Thesis Outline

In this chapter, we presented an overview of neuroimaging modalities and PD. We presented our research problem and motivated the significance in finding functional sub-regions. Also, we listed the contributions of this thesis. The subsequent chapters are organized as follows:

- Chapter 2 provides an overview of methods and techniques used to evaluate functional connectivity using rs-fMRI data between normal and PD subjects. Functional brain parcellation methods are reviewed in more detail.

- Chapter 3 describes our proposed FDR-controlled agglomerative clustering algorithm.

- Chapter 4 shows our results using the proposed FDR-controlled agglomerative clustering algorithm with other methods when parcellating synthetic and rs-fMRI data.

- Finally, Chapter 5 provides a summary of the thesis, the limitations of the proposed framework and future works.
Chapter 2

Review of Related Work

This chapter provides a general overview of model-based and data-driven methods in functional connectivity studies with the focus on functional brain parcellation using rs-fMRI data. Moreover, we will present findings in functional connectivity analysis in PD studies. We review numerous studies including seed models, independent component analysis, graph theory, and clustering methods. The reader is encouraged to consult survey papers by Cole et al. [11], Margulies et al. [29], and Prodoehl et al. [34] for further exploration of the field.

2.1 Overview of Functional Connectivity Analysis Using Resting-State fMRI

Biswal et al. were the first to demonstrate functional connectivity mapping using rs-fMRI data. They were able to show that a seed region in the left motor cortex replicated patterns of motor task activation. [3] Ever since this study, neuroscientists have been investigating the relationship between resting-state connectivity and functional organization of the brain. Two main categories used to determine functional connectivity are model-based methods and data-driven methods. Both categories have their own unique advantages and disadvantages.

Model-based methods are most commonly used because they are easy to implement and interpret. However, model-based methods require strong prior neuroscience knowledge or experience to determine the functional connectivity between different regions. Model-based methods use regions of interests (ROIs) known as “seeds”, which are correlated with other ROIs or voxels to determine whether there are any underlying relationships between different regions of the brain. Anatomical atlases are one way to define ROIs, which are represented by its mean time-series. [19, 28, 48] Wu et al. [48] used the pre-supplementary motor area (P-SMA) and primary motor cortex (M1) as seeds and a stronger connectivity in PD subjects compared to normal subjects between these regions. They also found weaker connectivity in
PD subjects compared to normal between the P-SMA and the following regions: putamen, insula, premotor cortex, and inferior parietal lobule (IPL). These findings are consistent with deficits in networks supporting movement planning and initiation in PD. Despite the popularity, there are drawbacks in using model-based methods. Functional connectivity is highly seed-dependent, which can only determine the connectivity between the seed and other ROIs or voxels. Secondly, a lot of focus is spent on relating data with prior knowledge and neglecting other parts or functions of the brain. Therefore, data-driven methods might be more useful in determining unexpected relationships between brain regions.

Data-driven methods do not require any prior neuroscience knowledge to determine the connectivity of the brain and does not require pre-defining a seed region. A wide range of data-driven methods have been used including clustering analysis, independent components analysis (ICA), graph theory, and regional homogeneity (ReHo). Clustering techniques such as spectral clustering and hierarchical clustering have been widely used on rs-fMRI datasets to determine functional connectivity patterns and functional brain parcellation. Shi and Malik developed the normalized cut (NCUT) spectral clustering algorithm, which optimizes within cluster similarity and between cluster dissimilarity. [38] NCUT has been applied by Craddock et al. [13], Shen et al. [37], and van den Heuvel [44] for their parcellation algorithms. In hierarchical clustering, singleton clusters or seed clusters are grown into parcels through an iterative process in which neighboring vertices are attached to a parcel. [4, 12, 15] A drawback to using clustering analysis is that a specific number of cluster needs to be define by using modeling selection techniques such as using goodness-of-fit or Bayesian information criterion [42]. Model selection is proven difficult given the subject variability and identifying structure in noisy data.

ICA decomposes rs-fMRI data into independent spatial or temporal components. ICA was first introduced to a task-based fMRI dataset by McKeeown et al. [30], which has since been used on rs-fMRI. ICA based whole brain pacellation has been successful. [39] However, there are drawback to using ICA. It is difficult to choose the number of clusters. Also, ICA assumes the components (fMRI time-series) are independent. If these conditions are not met, the parcellations are not as significant. [6] Therefore, ICA is mostly used as a preprocessing technique to remove physiological noise and noise caused by head motion.

Graph theory defines a set of nodes which can be ROIs or voxels and defines a set of edges to be the relationship between the nodes. A mathematical model of functional connectivity can be defined using graph theory.
and network properties such as degree distribution, network efficiency, nodal centrality, and modularity. [45] Graph theory has been applied using ROIs or voxels as nodes. For nodes based on ROIs, each node is represented by its mean time-series. The time-series are then correlated with one another to determine whether there are any functional connections and small world networks\(^2\). Wu et al. [47] used a form of degree distribution and found that PD subjects without medication showed significantly decreased functional connectivity in the supplementary motor area (SMA), left dorsolateral prefrontal cortex (DLPFC), and left putamen, and significantly increased connectivity in the left cerebellum, left M1, and left parietal cortex compared with normal subjects. Chang et al. proposed a technique to decompose a graph while maximizing its modularity by modeling the distribution of modularity in random networks. [8] They were able to find the default mode network, a motor-sensory subnetwork, and a visual-related subnetwork using an online rs-fMRI datasets.

ReHo analysis measures the similarity of time-series of each voxel with its nearest neighbors. ReHo analysis focuses on short distance functional connectivity. The issue with ReHo is the inconsistency between studies. [9, 34] These studies were able to find reduce activity in the putamen and increase activity in the cerebellum, medial frontal gyrus, and middle temporal gyrus in PD subjects compared with controls. The differences are the activity in the inferior parietal lobule.

### 2.2 Functional Brain Parcellation Techniques for Resting-State fMRI

We reviewed state-of-the-art parcellation methods on functional brain parcellation. Studies in spatially constrained parcellation and hierarchical clustering has grown during the recent years. This is due to the potential for robust parcels across different subjects.

A network can be represented by a graph \( G = (V,E) \), with vertices \( V = \{v_1, v_2, \ldots, v_N\} \) and edges \( E = \{e(i,j) : v_i, v_j \in V\} \). The vertex \( v_i \) represents node \( i \), and the edge \( e(i,j) \) represents the connection strength between node \( i \) and node \( j \). The connection strength is characterized by a weight \( w(i,j) \), which is usually a function of the Pearson’s correlation coefficient. The normalized cut (NCUT) is a type of spectral clustering that normalizes the cut cost by the sum of weights on all edges connecting voxels.

\(^2\)characterized by strongly interconnected sub-networks
2.2. Functional Brain Parcellation Techniques for Resting-State fMRI

in a partition to every voxel in the graph. A graph $G$ can be bi-partition into set $A_1$ and set $A_2$ where $A_1 \cup A_2 = V$ and $A_1 \cap A_2 = \emptyset$ while maximizing the similarity within set $A_1$ and set $A_2$ and minimizing the dissimilarity between set $A_1$ and set $A_2$. The cost function for NCUT algorithm is given by Eq. (2.1),

$$\phi(A_1, A_2) = \sum_{i \in A_1, j \in A_2} w(i, j) + \sum_{i \in A_1, j \in V} w(i, j) + \sum_{i \in A_2, j \in V} w(i, j).$$  (2.1)

Craddock et al. [13] found clusters with spatially coherent regions of homogeneous functional connectivity by adding spatial constrains to NCUT clustering algorithm. Another extension of NCUT is a multiclass spectral clustering, which has be proposed by Yu et al. [49]. Shen et al. extended NCUT by implementing a K-class spectral clustering algorithm which uses groupwise optimization to ensure functional homogeneity within each cluster and that the clusters are consistent across subjects. [37] The cost function of the groupwise optimization is given by Eq. (2.2),

$$\phi(Y, R_s) = 2NM - 2 * tr \left( Y \sum_s (X_sR_s)^T \right),$$  (2.2)

where $tr(.)$ denotes the trace$^3$ of a matrix. Let $S = \{1, 2, \ldots, M\}$ denote the set of graphs and $N$ voxels in each graph to be partitioned. The matrix $Y$ is a binary matrix with $K$ columns, where each column defines the membership of the $k$th cluster. The matrix $X_s$ is the first $K$ eigenvectors of the normalized weight matrix $W_s$ ranked by the eigenvalues $(1 = \lambda_0 > \lambda_1 > \cdots > \lambda_K)$, $W_s = D_s^{-\frac{1}{2}} \tilde{W}_s D_s^{-\frac{1}{2}}$, where $D_s$ is the diagonal matrix of degrees $D_s(i, i) = \sum_j w_s(i, j)$. Each row of the input data $X_s$ is normalized to have unit norm. The matrix $R_s$ is a generalized rotation matrix in this high-dimensional space such that $R_s^T R_s = I_K$, where $I_K$ is an identity matrix.

Thirion et al. showed that Ward hierarchical agglomerative clustering algorithm performed better than spectral and k-means clustering with regards to reproducibility and accuracy. [42] For Ward’s algorithm, all voxels are initialized to singleton clusters. At each iteration, clusters are merged based on minimizing the cost function in Eq. (2.3). Two clusters are merged if the resulting cluster minimizes the sum of squared differences of the fMRI signal within all clusters. At each step, the cost function is given by Eq. (2.3) for cluster $c_1$ and cluster $c_2$.

$^3tr(A) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^n a_{ii}$
2.2. Functional Brain Parcellation Techniques for Resting-State fMRI

Figure 2.1: Clusters within a network are compared against a null network. If a merge within a network exists, the within-module connection $A_{ij}$ is weaker than their expected value $E_{ij}$ and vice versa. This figure is adapted from [8].

$$\phi(c_1, c_2) = \frac{|c_1||c_2|}{|c_1| + |c_2|} ||\langle Y_{c_1} \rangle - \langle Y_{c_2} \rangle||^2_2,$$

where $\langle Y_c \rangle$ is the mean time-series of cluster $c$. To account for spatial information, only neighboring clusters were merged.

Clusters have been determined by maximizing modularity in a network. Newman introduced modularity, which measures the network structure by comparing it with the expect null network (i.e. a network with the same number of vertices and with no underlying structure). [31] If a partition within a network exists, then the edge strength within the clusters are weaker compared to the expected null network. If a merge within a network exists, then the edge strength within the clusters should be stronger than that of the expected null network. This is illustrated in Figure 2.1. Thus, the cost function is given by Eq. (2.4),

$$\phi = E - A,$$

where $A$ is the adjacency matrix and $E$ is the expected null network. Traditionally, $E$ was proposed by Newman [31] with elements shown in Eq. (2.5).
2.2. Functional Brain Parcellation Techniques for Resting-State fMRI

\[ E_{ij} = \frac{k_i k_j}{2m}, \quad (2.5) \]

where the degree vector \( k \) has elements \( k_i = \sum_j A_{ij} \) (i.e., \( k_i \) equals to the sum of the connection strength associated with voxel \( i \)). The total sum of all the edges of the network is given by \( m = \frac{1}{2} \sum_i k_i \).

Chang et al. recently proposed an expected null model defined in Eq. (2.6), which has been shown in enhance parcel detection and it allows for negative edges (\( A_{ij} < 0 \)) for a given network. [7] The conventional null model implicitly assumes self-loops (connections from voxels to themselves, \( E_{ii} \neq 0 \)) which are often topologically invalid and meaningless for the majority of brain networks.

\[ E_{ij} = E(A_{ij}) = \begin{cases} \frac{k_i + k_j}{N-2} - \frac{2m}{(N-1)(N-2)}, & \text{if } i \neq j, \\ 0, & \text{otherwise}, \end{cases} \quad (2.6) \]

The algorithm developed by Chang et al. was used to find the default mode network, motor-sensory subnetwork and visual-related subnetwork using the mean time-series of ROIs.
Chapter 3
Method

The proposed FDR-controlled agglomerative clustering algorithm for functional brain parcellation contains two major steps. The first step is to initialize the parcels with highly correlated rs-fMRI time-series (voxels) within 54 anatomical regions defined in Table 4.3 and Table 4.4. Next, we used our FDR Controller to iteratively determine the parcels within each anatomical regions using significance-based modularity partitioning algorithm\(^4\). The proposed framework is shown in Algorithm 1. The subsequent subsections will elaborate on the individual steps of the FDR-controlled agglomerative clustering algorithm.

3.1 Parcel Initialization

Let rs-fMRI data be denoted by a matrix \( X = [x_1, x_2, \ldots, x_N] \), where \( x_i \) represents a rs-fMRI time-series at voxel \( i \). \( X \) has size of \( T \times N \), where \( T \) represents the length of the time-points, and \( N \) represents the number of voxels in fMRI data. The connection and dependency relationships between the voxels can be represented by an undirected weighted graph \( G = (V, E) \), with vertices (or nodes) \( V = \{v_1, v_2, \ldots, v_N\} \) and edges \( E = \{e(i, j) : v_i, v_j \in V\} \). The vertex \( v_i \) represents voxel \( i \), and the edge \( e(i, j) \) represents the connection strength between voxel \( i \) and voxel \( j \). The graph \( G \) is represented by a network with an adjacency matrix \( A \) of size \( N \times N \), where the element \( A_{ij} \) indicates the connection strength between voxel \( i \) and voxel \( j \). The matrix \( A \) can be defined differently based on the dependency relationships between voxels. In this thesis, we defined the element \( A_{ij} \) to be the Pearson’s correlation coefficient between voxel \( i \) and voxel \( j \),

\[
A(x_i, x_j) = \frac{\sum_{l=0}^{T} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j)}{(T - 1)\sigma_{x_i}\sigma_{x_j}},
\]

where \( \bar{x} \) is the mean and \( \sigma_x \) is the standard deviation of a rs-fMRI time-series.

\(^4\)http://mcgovern.mit.edu/technology/meg-lab/software
3.1. Parcel Initialization

Sub-regions of the brain are ensured to be spatially contiguous by merging clusters that are closer to each other. The matrix $D$ with element $D_{ij}$ is defined to be the minimum Euclidean distance between cluster $C_i$ and cluster $C_j$, as expressed in Eq. (3.2),

$$D_{ij} = \min_{k_1 \in C_i, k_2 \in C_j} d(C_i(k_1), C_j(k_2)), \quad (3.2)$$

where the function $d(\cdot)$ calculates the Euclidean distance between voxels in cluster $C_i$ and cluster $C_j$ with indices $k_1$ and $k_2$ respectively. We also define the Gaussian adjusted distance matrix $D_G$, where its element $D_{Gi,j}$ is defined as mapping the element $D_{ij}$ to a normalize Gaussian function with $\mu = 1$ and $\sigma = 1$, as in Eq. (3.3).

$$D_{Gi,j} = e^{-\frac{(D_{ij} - \mu)^2}{2\sigma}} \quad (3.3)$$

We assumed that the anatomical ↔ functional relations are consistent across individuals, parcellation of the brain into functional regions can be done by using anatomical templates (and/or probabilistic atlases) to define anatomical and hence functional regions. The parcels are initialized by merging highly correlated voxels within each cortical and subcortical regions.
3.2 Statistical Significance between Parcels

Voxels are merged into parcels if the voxels satisfy the following condition in Eq. (3.4),

\[ A \ast D \geq t \]  

(3.4)

where \( t = 0.5 \) is the threshold value. The threshold value was set to 0.5 to ensure the parcels are spatially contiguous and to exclude negative and weak correlations.

3.2 Statistical Significance between Parcels

The modularity of the network is measured using the expected null network defined in Eq. (2.6). The modularity matrix is denoted by \( B \) with element \( B_{ij} \), where \( B_{ij} = A_{ij} - E_{ij} \). The modularity value is close to the maximum eigenvalue of matrix \( B \). We tested whether the maximum eigenvalue of matrix \( B \) of the original network is higher than the distribution of eigenvalues of \( B_{\text{Random}} \) of random networks. The distribution of the largest eigenvalue has been shown to be related to a linear transformation of the Tracy-Widom distribution. Using this, Chang et. al. derived the parameters for the distribution of modularity in random networks (interested readers are encouraged to refer to [8] for their derivation). Using these parameters, we compared the modularity of functional brain regions with the modularity in random networks. We denote the matrix \( p \) with element \( p_{ij} \) to be the statistical significance of cluster \( C_i \) and cluster \( C_j \) when comparing the largest eigenvalue of \( B \) against the distribution of eigenvalues of \( B_{\text{Random}} \).

3.3 False Discovery Rate Controller

One of the differences between the significance-based modularity partitioning algorithm\(^5\) and our FDR-controlled agglomerative clustering algorithm is controlling the rate of false discoveries rather than controlling the tradition type one error. We need to account for false discoveries (type-I error) when testing multiple \( m \) independent hypotheses simultaneously. For this reason, FDR is more reasonable in some real applications such as bioinformatics and neuroimaging since it directly related with the uncertainty of the discovered positive results. The FDR is one of the important error control criteria for multiple testing (see Table 3.1).

\(^5\)http://mcgovern.mit.edu/technology/meg-lab/software
3.3. False Discovery Rate Controller

Table 3.1: Error control criteria for multiple testing with notation shown in Table 3.2.

<table>
<thead>
<tr>
<th>Types of Error Rate Control</th>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>False Discovery Rate</td>
<td>FDR</td>
<td>$E\left(\frac{FP}{R}\right)$</td>
</tr>
<tr>
<td>Family Wise Error Rate</td>
<td>FWER</td>
<td>$P(FP \geq 1)$</td>
</tr>
<tr>
<td>Type I Error Rate (False Positive Rate)</td>
<td>$\alpha$</td>
<td>$E\left(\frac{FP}{m_0}\right)$</td>
</tr>
<tr>
<td>Type II Error Rate (False Negative Rate)</td>
<td>$\beta$</td>
<td>$E\left(\frac{FP}{m_0}\right)$</td>
</tr>
<tr>
<td>Sensitivity (True Positive Rate)</td>
<td>$1 - \beta$</td>
<td>$E\left(\frac{TP}{m_0}\right)$</td>
</tr>
<tr>
<td>Specificity (True Negative Rate)</td>
<td>$1 - \alpha$</td>
<td>$E\left(\frac{TN}{m - m_0}\right)$</td>
</tr>
</tbody>
</table>

Table 3.2: The number of errors committed when testing $m$ hypotheses.

<table>
<thead>
<tr>
<th></th>
<th>Null hypothesis is True (H0)</th>
<th>Alternative hypothesis is True (H1)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declared significant</td>
<td>False positive (FP)</td>
<td>True positive (TP)</td>
<td>$R$</td>
</tr>
<tr>
<td>Declared non-significant</td>
<td>True negative (TN)</td>
<td>False negative (FN)</td>
<td>$m - R$</td>
</tr>
<tr>
<td>Total</td>
<td>$m_0$</td>
<td>$m - m_0$</td>
<td>$m$</td>
</tr>
</tbody>
</table>

The FDR is defined as the expected ratio of falsely discovered positive hypothesis to all those discovered as defined in Eq. (3.5),

$$Q = \begin{cases} E\left(\frac{FP}{R}\right), & FP + R > 0, \\ 0, & FP + R = 0, \end{cases}$$  

(3.5)

where FP is the number of false positives and R is the number of hypothesis declared significant. The number of false discoveries can be controlled by using the linear step-up Benjamini-Hochberg (BH) procedure [2]. The BH procedure is outlined as follows:

1. Sort the adjusted $p$-values from $H$ hypothesis tests in ascending order: $p_1 \leq p_2 \leq \cdots \leq p_H$;
2. Find the largest $k$ such that $p_k \leq \frac{k}{H} q$;
3. Reject hypotheses 1, 2, \ldots , and $k$. 

19
3.4. Accuracy of Representation

The minimum FDR threshold at which a hypothesis is rejected is defined as the \( q \)-value. In our method, we set \( q = 0.05 \) and followed the BH procedure to determine which parcels to merge.

3.4 Accuracy of Representation

There is no standard in evaluating parcellations using rs-fMRI data since there is no known gold standard in anatomical and functional segments of the brain. Therefore, we needed to indirectly validate our results using the adjusted Rand index. We also calculated the normalized total connectivity degree \( \Gamma \) between each motor region to determine if our parcellations could be used to find differences between normal and PD subjects using rs-fMRI.

3.4.1 Adjusted Rand Index

There are a lot of commonly used methods to compare parcellation results including the Dice index, but it does not apply in our framework. Two parcellations does not usually reflect the same number of parcels. Hence, we have chosen the adjusted Rand index \cite{22} to compute the mismatch between two parcellation results. \cite{21} In our framework, two parcellations do not necessarily contain the same number of parcels. Also, the parcel indices do not reflect their spatial proximity.

Let’s denote \( U = i \) and \( V = j \) to be two parcellation results where \( i \) and \( j \) are parcels of \( U \) and \( V \) respectively. The mismatch matrix \( (n_{ij}) \) contains one element \( n_{ij} \) for each parcel \( i \) of parcellation \( U \) and each parcel \( j \) of parcellation \( V \) that is equal to the number of nodes that are shared by these two parcels:

\[
n_{ij} = |i \cap j|
\]

The adjusted Rand index defined in Eq. (3.6),

\[
aRI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[ \sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right]/\binom{n}{2}}{\frac{1}{2} \left[ \sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[ \sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right]/\binom{n}{2}},
\]

(3.6)

where \( n_{ij} \) represents the number of mismatch in parcel \( i \) of parcellation \( U \) and parcel \( j \) in parcellation \( V \), \( a_i = \sum_j n_{ij}, b_j = \sum_i n_{ij} \) and \( n = a_i = b_j. \)

The adjusted Rand index is between -1 and 1. \( aRI = 0 \) means that the parcels are independent, \( aRI = 1 \) indicates the clusters are identical.
3.4.2 Functional Connectivity Analysis

We used a network model based on graph theory developed by Jiang et al. [23] and used in [47] to compared functional connectivity between normal and PD subjects. The connectivity degree $\eta_{ij}$ between node $i$ and node $j$ is defined in Eq. (3.7),

$$\eta_{ij} = e^{\xi d_{ij}}, \quad (3.7)$$

where $\xi$ is a real positive constant ($\xi$ is a subjective selection and discussed by [27] and is set to fixed to $\xi = 2$) and $d_{ij}$ is the distance between the two nodes. The element $d_{ij}$ is a hyperbolic correlation measure which is defined in Eq. (3.8),

$$d_{ij} = \frac{1}{1 + c_{ij}}, \quad (3.8)$$

where $c_{ij}$ represents the Pearson’s correlation coefficient between node $i$ and node $j$. The total connectivity degree $\Gamma_i$ at node $i$ in a graph is the sum of all the connectivity degrees between node $i$ and all other nodes as shown in Eq. (3.9). The total connectivity degree $\Gamma_i$ at node $i$ is the amount of information receiving from the particular network. Hence, the larger the total connectivity degree $\Gamma_i$ at node $i$ means that this node is more functionally connected to other regions in the network. The total connectivity degree $\Gamma$ takes into account n-to-1 connectivity using 1-to-1 connectivity measures. Thus, it is possible to find the changes of the total functional connectivity degree for some regions by detecting in different brain activity states. The normalized total connectivity $\bar{\Gamma}_i$ at node $i$ is normalized to account for the varying number of voxels in each brain region as shown in Eq. (3.10).

$$\Gamma_i = \sum_{j=1}^{n} \eta_{ij}, \quad (3.9)$$

$$\bar{\Gamma}_i = \frac{\Gamma_i}{\sum_{j=1}^{n} \Gamma_j} \quad (3.10)$$
Chapter 4

Results

In this chapter, we first show a comparison between the our FDR-controlled agglomerative clustering algorithm with Ward agglomerative clustering method and the method proposed by Chang et. al. [8] using synthetic datasets. The proposed method outperformed the other two methods. The proposed approach was then applied to a rs-fMRI dataset and functional connectivity results subregions of the brain are shown.

4.1 Synthetic Dataset

We needed to ensure that our synthetic datasets would match those of rs-fMRI data. Our synthetic datasets contains time-series generated from a Gaussian distribution with mean zero and unit variance. The synthetic dataset contains 27 parcels, each with its own color, and has a size of $9 \times 9 \times 9 \times 500$ as shown in Figure 4.1. Each voxels are placed on a three dimensional grid with position $p(x, y, z)$. Each parcel has 27 time-series (voxels), and each time-series contains 500 time-points. The time-series are generated with strong intra-cluster correlations (in the range of 0.50-1.00) and weaker inter-cluster correlations (in the range of 0.00-0.30). Figure 4.2 illustrates a simple example of nine voxels with three respective clusters to show the correlations between the voxels, which is similar to our synthetic data. The intra-cluster and inter-cluster correlations are varied in 60 repeated experiments to test the robustness of the parcellation methods. The first set consisted of inter correlations to be 0.00-0.10. The second set consisted of inter correlations to be 0.10-0.20. The third set consisted of inter correlations to be 0.20-0.30. The Cholesky factorization\(^6\) is used to generate the time series with the correlation values.

In our case, parcellation $U$ represents the ground truth of our synthetic data, and parcellation $V$ represents the parcellation results using our proposed parcellation method, our implementation of Ward agglomerative clus-

\(^6\)R^*R=A, where the upper triangular matrix R from the diagonal and upper triangle of matrix A
4.1. Synthetic Dataset

Figure 4.1: An example of the parcellation results from synthetic dataset 3 with an intra-correlation equal to 0.65 and an inter-correlation equal to 0.30.

(a) Ground truth  (b) Results using FDR Controller

(c) Results using Chang et. al. [11]  (d) Results using Ward

Figure 4.1: An example of the parcellation results from synthetic dataset 3 with an intra-correlation equal to 0.65 and an inter-correlation equal to 0.30.
4.1. Synthetic Dataset

Figure 4.2: An illustration of a two-dimensional patch of nine voxels clustered into three regions: white, purple, and red. This translates into a $9 \times 9$ connectivity matrix with a 0.9 intra-cluster correlation and a 0.1 inter-cluster correlation.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th></th>
<th>4</th>
<th>5</th>
<th>6</th>
<th></th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
</table>

$$
\begin{bmatrix}
1 & 0.9 & 0.9 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\
0.9 & 1 & 0.9 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\
0.9 & 0.9 & 1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\
0.1 & 0.1 & 0.1 & 1 & 0.9 & 0.1 & 0.9 & 0.9 & 0.1 \\
0.1 & 0.1 & 0.1 & 0.9 & 1 & 0.1 & 0.9 & 0.9 & 0.1 \\
0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 1 & 0.1 & 0.1 & 0.9 \\
0.1 & 0.1 & 0.1 & 0.9 & 0.9 & 0.1 & 1 & 0.9 & 0.1 \\
0.1 & 0.1 & 0.1 & 0.9 & 0.9 & 0.1 & 0.9 & 1 & 0.1 \\
0.1 & 0.1 & 0.1 & 0.1 & 0.9 & 0.1 & 0.1 & 0.1 & 1
\end{bmatrix}
$$

To apply our dataset to the method proposed by Chang et al. [8], we calculated the Pearson’s correlation coefficients between each time series. This will form a correlation matrix which is the only input needed for their significance-based modularity partitioning algorithm.

The average value of the adjusted Rand index for all 60 experiments are shown in Table 4.1. It is clear that the proposed method outperforms the other two methods when tested on all three synthetic datasets.

Table 4.1: Our results when comparing our method with the Ward clustering method and the method proposed by Chang et al. method [8] using synthetic data. The values represent the average adjusted Rand index of all 60 experiments in each dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Data</th>
<th>Syn 1</th>
<th>Syn 2</th>
<th>Syn 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our method</td>
<td></td>
<td>0.9185</td>
<td>0.8726</td>
<td>0.8398</td>
</tr>
<tr>
<td>Ward clustering method</td>
<td></td>
<td>0.7782</td>
<td>0.7816</td>
<td>0.7595</td>
</tr>
<tr>
<td>Chang et al. method [8]</td>
<td></td>
<td>0.5441</td>
<td>0.5322</td>
<td>0.5528</td>
</tr>
</tbody>
</table>
Table 4.2: Demographic data of normal and PD subjects. Four demographics for normal subjects were not recorded.

<table>
<thead>
<tr>
<th></th>
<th>Normal subjects</th>
<th>Parkinson’s disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>2/3</td>
<td>8/3</td>
</tr>
<tr>
<td>Age (years)</td>
<td>59 ± 4.5</td>
<td>60.0 ± 10.4</td>
</tr>
<tr>
<td>Handedness (Left/Right)</td>
<td>0/5</td>
<td>0/11</td>
</tr>
<tr>
<td>UPDSR</td>
<td>N/A</td>
<td>29 ± 9.9</td>
</tr>
</tbody>
</table>

4.2 fMRI Dataset

4.2.1 Data Acquisition

Nine healthy control subjects and eleven PD subjects were recruited from the Pacific Parkinson’s Research Center at the University of British Columbia (UBC). All experiments were approved by the Ethics Board at UBC, and all subjects provided informed consent prior to participating in the experiment. The subjects laid on their back with eyes closed for several minutes before data was acquired. The PD subjects were not on medication. The relevant demographics are shown in Table 4.2.

The subjects were scanned using a 3-Tesla Philips Gyroscan Intera (Philips Medical Systems, Netherlands) equipped with a head coil was used to collect rs-fMRI data. BOLD contrast echo-planar imaging (EPI) T2*-weighted images were taken with the following specifications: repetition time TR = 1985 ms; echo time TE = 37 ms; flip angle = 90°; field of view (FOV) = 240.00 mm; Image of size 128 × 128 pixels; pixel size of 1.9 × 1.9 mm². There are 36 axial slices, each slice has a thickness of 3-mm and a gap thickness of 1-mm as shown in Figure 4.3. The duration of each functional run was four minutes during which we obtained the axial slices. The fMRI scans are shown in Figure 4.4. The FOV was set to include the cerebellum ventrally and also include the dorsal surface of the brain.

4.2.2 Data Preprocessing

We applied our processing pipeline on the data collected using a 3-Tesla Philips Gyroscan Intera using various open source software. Our preprocessing pipeline which performed SPM⁷ for slice timing correction, isotropic reslicing, SPM-based motion correction and registration.

⁷http://www.fil.ion.ucl.ac.uk/spm/software/spm8/
4.2. fMRI Dataset

Figure 4.3: Examples of MRI scans collected from a normal subject.

Figure 4.4: Examples of fMRI scans collected which spatially corresponds to the MRI scans in Figure 4.3.
4.2. fMRI Dataset

Figure 4.5: Examples of the gray matter mask which corresponds to the fMRI scans in Figure 4.4.

In order to minimize registration error due to inter-subject variability of brain regions, particularly of subcortical brain regions, we refrained from warping brain images to a standard space such as Talairach template or MNI template\(^8\). Instead, we used an ROI-based co-registration of fMRI data to structural images. 54 ROIs were derived from an open source software called Freesurfer\(^9\). Freesurfer uses a high-dimensional registration process by utilizing probabilistic atlases for cortical and subcortical labeling. The labels are then propagated back to the subject’s native space by relying on the template label and the subject’s transformed voxel location obtained from T1-weighted structural scans. The voxels selected from our fMRI data was by using a gray matter mask as shown in Figure 4.5.

From our experience, this ROI-based segmentation method is superior to manually-drawn ROIs and voxel-based method as it minimizes registration error particularly in subcortical brain regions. We also defined ROIs based on the Human Motor Area Template (HMAT)\(^{10}\) labels to define motor regions of the brain. Table 4.3 and Table 4.4 contains a list of all ROIs used for our research.

For highly accurate structural analysis of subcortical structures, we have in the past used Large Deformation Diffeomorphic Metric Mapping (LD-DMM) \(^{[14]}\). However, we using he Freesurfer segmentation on high-resolution structural scans (\(1 \times 1 \times 1 \text{ mm}^3\)) to provide ROI-based binary masks to the relatively low-resolution fMRI scans (\(3 \times 3 \times 3 \text{ mm}^3\)). We have found the Freesufer segmentations well-suited for this purpose. The extracted ROIs are

---

\(^8\)http://www.nil.wustl.edu/labs/kevin/man/answers/mnispace.html
\(^9\)http://freesurfer.net/
\(^{10}\)http://fmlab.org/
visually checked by experienced neurologists if needed. The registration results are manually verified by overlaying the cortical and subcortical regions onto the mean fMRI image and matching the folds. The time course was extracted the fMRI time series that corresponded to those cortical/subcortical labels. The rs-fMRI time-series was filtered using a Butterworth bandpass filter (0.02 Hz – 0.1 Hz) and global signal regressed. [37]
### 4.2. fMRI Dataset

Table 4.3: The cortical and subcortical ROIs extracted from the left hemisphere of the brain. The left column are the ROIs and the right column correspond to Freesurfer and HMAT labels. The HMAT labels are bolded.

<table>
<thead>
<tr>
<th>ROI</th>
<th>Freesurfer and HMAT labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Dorsal lateral prefrontal cortex</td>
<td>ctx_lh_G_front_middle</td>
</tr>
<tr>
<td>Left Ventral lateral prefrontal cortex</td>
<td>ctx_lh_G_front_inf-Opercicular, ctx_lh_G_front_inf-Triangul</td>
</tr>
<tr>
<td>Left Insula</td>
<td>ctx-lh-insula</td>
</tr>
<tr>
<td>Left Superior temporal lobe</td>
<td>ctx-lh-supiortemporal</td>
</tr>
<tr>
<td>Left Middle temporal lobe</td>
<td>ctx-lh-middletemporal</td>
</tr>
<tr>
<td>Left Inferior temporal lobe</td>
<td>ctx-lh-supramarginal, ctx_lh_G_pariet_inf-Angular, ctx_lh_G_pariet_inf-Supramar</td>
</tr>
<tr>
<td>Left Parahippocampal cortex</td>
<td>ctx-lh-parahippocampal</td>
</tr>
<tr>
<td>Left Hippocampus</td>
<td>Left-Hippocampus</td>
</tr>
<tr>
<td>Left Somatosensory cortex</td>
<td>ctx-lh-postcentral, L_S1</td>
</tr>
<tr>
<td>Left Superior parietal lobe</td>
<td>ctx-lh-supiortparietal</td>
</tr>
<tr>
<td>Left Inferior parietal lobe</td>
<td>ctx-lh-supramarginal, ctx_lh_G_pariet_inf-Angular, ctx_lh_G_pariet_inf-Supramar</td>
</tr>
<tr>
<td>Left Occipital-parietal cortex</td>
<td>ctx_lh_S_parieto_occipital, ctx_lh_G_occipital_sup</td>
</tr>
<tr>
<td>Left Lateral occipital lobe</td>
<td>ctx-lh-lateraloccipital</td>
</tr>
<tr>
<td>Left Anterior cingulate cortex</td>
<td>ctx-lh-caudalanteriorcingulate</td>
</tr>
<tr>
<td>Left Posterior cingulate cortex</td>
<td>ctx-lh-posteriorcingulate</td>
</tr>
<tr>
<td>Left Precuneus</td>
<td>ctx-lh-precuneus</td>
</tr>
<tr>
<td>Left Orbitofrontal gyrus</td>
<td>ctx-lh-medialorbitofrontal, ctx-lh-lateralorbitofrontal</td>
</tr>
<tr>
<td>Left Cerebellum</td>
<td>Left-Cerebellum-Cortex</td>
</tr>
<tr>
<td>Left Primary motor cortex</td>
<td>L_M1</td>
</tr>
<tr>
<td>Left Supplementary motor area</td>
<td>L_SMA_proper</td>
</tr>
<tr>
<td>Left Pre-supplementary motor area</td>
<td>L_Pre_SMA</td>
</tr>
<tr>
<td>Left Dorsal premotor area</td>
<td>L_PMd</td>
</tr>
<tr>
<td>Left Ventral premotor area</td>
<td>L_PMd</td>
</tr>
<tr>
<td>Left Thalamus</td>
<td>Left-Thalamus- Proper</td>
</tr>
<tr>
<td>Left Caudate</td>
<td>Left-Caudate</td>
</tr>
<tr>
<td>Left Putamen</td>
<td>Left-Putamen</td>
</tr>
<tr>
<td>Left Pallidum</td>
<td>Left-Pallidum</td>
</tr>
</tbody>
</table>
Table 4.4: The cortical and subcortical ROIs extracted from the right hemisphere of the brain. The left column are the ROIs and the right column correspond to Freesurfer and HMAT labels. The HMAT labels are bolded.

<table>
<thead>
<tr>
<th>ROI</th>
<th>Freesurfer and HMAT labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right Dorsal lateral prefrontal cortex</td>
<td>ctx_rh_G_front_middle\nctx_rh_G_front_inf-Opercual,\nctx_rh_G_front_inf-Orbital,\nctx_rh_G_front_inf-Triangul</td>
</tr>
<tr>
<td>Right Ventral lateral prefrontal cortex</td>
<td>ctx_rh_insula\nctx_rh_superiortemporal\nctx_rh_middletemporal\nctx_rh_supramarginal,\nctx_rh_G_pariet_inf-Angular,\nctx_rh_G_pariet_inf-Supramar\nctx_rh_parahippocampal\nctx_rh_postcentral,\nR_S1</td>
</tr>
<tr>
<td>Right Insula</td>
<td>R_S1\nctx_rh_superiorparietal\nctx_rh_supramarginal,\nctx_rh_G_pariet_inf-Angular,\nctx_rh_G_pariet_inf-Supramar\nctx_rh_S_parieto_occipital,\ncx_rh_G_occipital_sup,</td>
</tr>
<tr>
<td>Right Superior parietal lobe</td>
<td>cx_rh_lateraloccipital\nctx_rh_caudalanteriorcingulate\nctx_rh_posteriorcingulate\nctx_rh_precuneus\nctx_rh_medialorbitofrontal,\ncx_rh_lateralorbitofrontal</td>
</tr>
<tr>
<td>Right Inferior parietal lobe</td>
<td></td>
</tr>
<tr>
<td>Right Occipital-parietal cortex</td>
<td>R_S1\nctx_rh_superiorparietal\nctx_rh_supramarginal,\nctx_rh_G_pariet_inf-Angular,\nctx_rh_G_pariet_inf-Supramar,\ncx_rh_S_parieto_occipital,\ncx_rh_G_occipital_sup,</td>
</tr>
<tr>
<td>Right Lateral occipital lobe</td>
<td></td>
</tr>
<tr>
<td>Right Anterior cingulate cortex</td>
<td></td>
</tr>
<tr>
<td>Right Posterior cingulate cortex</td>
<td></td>
</tr>
<tr>
<td>Right Precuneus</td>
<td></td>
</tr>
<tr>
<td>Right Orbitofrontal gyrus</td>
<td></td>
</tr>
<tr>
<td>Right Cerebellum</td>
<td>R_M1\nR_SMA_proper\nR_Pre_SMA\nR_PMd\nR_PMv</td>
</tr>
<tr>
<td>Right Primary motor cortex</td>
<td></td>
</tr>
<tr>
<td>Right Supplementary motor area</td>
<td></td>
</tr>
<tr>
<td>Right Pre- supplementary motor area</td>
<td></td>
</tr>
<tr>
<td>Right Dorsal premotor area</td>
<td></td>
</tr>
<tr>
<td>Right Ventral premotor area</td>
<td></td>
</tr>
<tr>
<td>Right Thalamus</td>
<td>Right-Thalamus-Proper\nRight-Caudate\nRight-Putamen\nRight-Pallidum</td>
</tr>
<tr>
<td>Right Caudate</td>
<td></td>
</tr>
<tr>
<td>Right Putamen</td>
<td></td>
</tr>
<tr>
<td>Right Pallidum</td>
<td></td>
</tr>
</tbody>
</table>
4.2.3 fMRI Parcellation Results

We continued to test the robustness of our parcellation algorithm with the Ward’s agglomerative clustering method. The time-points of each individual subject are split into two parts. The first part contains the first half of the time-points for that a dataset, and the second part contains the second half of the time-points for the same dataset. We did not use the method by Chang et al. [8] given the poor parcellation results on our synthetic datasets and the computational time required to parcellate the fMRI data. Figure 4.6 and Figure 4.7 shows the reproducibility for our FDR-controlled agglomerative clustering algorithm and the Ward agglomerative clustering algorithm using the adjusted Rand indices. It shows that our FDR-controlled agglomerative clustering algorithm produces higher reproducibility compared to the Ward hierarchical clustering method. The variance in the adjusted Rand index is larger for PD subjects compared to normal subjects. This could be due to decrease in functional connectivity in PD subjects.

![Boxplots of the adjusted Rand index showing the reproducibility of the results using our method and the Ward agglomerative clustering using normal subjects.](image)

Figure 4.6: Boxplots of the adjusted Rand index showing the reproducibility of the results using our method and the Ward agglomerative clustering using normal subjects.
There are some consistency in the parcellations between brain regions across normal and PD subjects. For instance, in Figure 4.8 and Figure 4.9 show the posterior lobe of the left cerebellum is one cluster, but the anterior lobe consists of multiple clusters for normal and PD subjects. Also, in Figure 4.10 and Figure 4.11, the inferior portion of the left primary motor cortex is shown to be one cluster, but the superior and lateral regions were not consider to be a cluster. The largest cluster is indicated by the orange points. The parcellation of each region for all subjects shown in Table 4.3 and Table 4.4 are available online\textsuperscript{11}.

There are consistencies in the parcellation between normal and PD subjects. Examples of the combined parcellated regions of normal subjects are shown in Figure 4.12 and Figure 4.13. Examples of the combined parcellated regions of PD subjects are shown in Figure 4.14 and Figure 4.15. Note the color represents only the cluster index. Most of the superior portions of the regions from Table 4.3 and Table 4.4 were had multiple communities.

\textsuperscript{11}https://www.dropbox.com/sh/ytqcmnaw0to1a98/AACl7stmddt8_sw9R6pH3Lota?dl=0
4.2. fMRI Dataset

Figure 4.8: The parcellation results using the proposed FDR-controlled agglomerative clustering algorithm on the left cerebellum on normal subjects. The orange points indicate the largest cluster.
Figure 4.9: The parcellation results using the proposed FDR-controlled agglomerative clustering algorithm on the left cerebellum on PD subjects. The orange points indicate the largest cluster.
4.2. fMRI Dataset

Figure 4.10: The parcellation results using the proposed FDR-controlled agglomerative clustering algorithm on the left primary motor cortex on normal subjects. The orange points indicate the largest cluster.
Figure 4.11: The parcellation results using the proposed FDR-controlled agglomerative clustering algorithm on the left primary motor cortex on PD subjects. The orange points indicate the largest cluster.
4.2. fMRI Dataset

Figure 4.12: Parcellation results for subject N003. The numbers increase from inferior to superior for axial (A) slices. The numbers increase from posterior to anterior for coronal (C) slices. The numbers increase from right to left for sagittal (S) slices.
4.2. fMRI Dataset

Figure 4.13: Parcellation results on subject N005. The numbers increase from inferior to superior for axial (A) slices. The numbers increase from posterior to anterior for coronal (C) slices. The numbers increase from right to left for sagittal (S) slices.
Figure 4.14: Parcellation results on subject P001. The numbers increase from inferior to superior for axial (A) slices. The numbers increase from posterior to anterior for coronal (C) slices. The numbers increase from right to left for saggital (S) slices. The numbers increase from inferior to superior for axial (A) slices. The numbers increase from posterior to anterior for coronal (C) slices. The numbers increase from right to left for saggital (S) slices.
4.2. fMRI Dataset

Figure 4.15: Parcellation results on subject P003. The numbers increase from inferior to superior for axial (A) slices. The numbers increase from posterior to anterior for coronal (C) slices. The numbers increase from right to left for saggital (S) slices. The numbers increase from inferior to superior for axial (A) slices. The numbers increase from posterior to anterior for coronal (C) slices. The numbers increase from right to left for saggital (S) slices.
4.2.4 Functional Connectivity Results

The 32 brain regions selected as the motor network for our functional connectivity analysis are shown in Table 4.5 with their associated abbreviations. The regions and clusters within the regions are defined as nodes in our analysis. Some brain regions including the left primary motor cortex (M1), the left superior parietal lobe (SPL), the left inferior parietal lobe (IPL) and the left supplementary motor area (SMA) have a cluster that appears consistent across subjects for both normal and PD subjects. The other clusters within those regions have singleton clusters and clusters with less than five voxels. We decided to compare whether these smaller clusters had any functional connectivity with other regions of the brain. So, the smaller clusters were merged into one cluster as shown in Figure 4.22.

Table 4.5: The left (L) and right (R) motor regions of the brain chosen for functional connectivity analysis.

<table>
<thead>
<tr>
<th>Brain region</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>L/R Dorsal lateral prefrontal cortex</td>
<td>DLPFC</td>
</tr>
<tr>
<td>L/R Superior parietal lobe</td>
<td>SPL</td>
</tr>
<tr>
<td>L/R Inferior parietal lobe</td>
<td>IPL</td>
</tr>
<tr>
<td>L/R Occipital-parietal cortex</td>
<td>OPC</td>
</tr>
<tr>
<td>L/R Anterior cingulate cortex</td>
<td>ACC</td>
</tr>
<tr>
<td>L/R Posterior cingulate cortex</td>
<td>PCC</td>
</tr>
<tr>
<td>L/R Cerebellum</td>
<td>Cb</td>
</tr>
<tr>
<td>L/R Primary motor cortex</td>
<td>M1</td>
</tr>
<tr>
<td>L/R Supplementary motor area</td>
<td>SMA</td>
</tr>
<tr>
<td>L/R Pre-supplementary motor area</td>
<td>P-SMA</td>
</tr>
<tr>
<td>L/R Dorsal premotor area</td>
<td>PMd</td>
</tr>
<tr>
<td>L/R Ventral premotor area</td>
<td>PMv</td>
</tr>
<tr>
<td>L/R Thalamus</td>
<td>Th</td>
</tr>
<tr>
<td>L/R Caudate</td>
<td>Ca</td>
</tr>
<tr>
<td>L/R Putamen</td>
<td>Pu</td>
</tr>
<tr>
<td>L/R Pallidum</td>
<td>Pa</td>
</tr>
</tbody>
</table>

We conducted three studies to compare the normalized total connectivity degree $\bar{\Gamma}$ within each of the following:

1. each ROI defined by our preprocessing pipeline;

2. the largest cluster defined by our FDR-controlled agglomerative clustering algorithm and the ROIs defined by our preprocessing pipeline;
3. the small clusters defined by our FDR-controlled agglomerative clustering algorithm and the ROIs defined by our preprocessing pipeline.

For the first test, the normalized total connectivity degree $\bar{\Gamma}$ between each ROI node are shown in Figure 4.16 and Figure 4.17. The difference in the normalized total connectivity degree $\bar{\Gamma}$ of each ROI node between normal and PD subjects is shown in Figure 4.18. We used a two-sample, two sided t-test to compare the results between normal and PD subjects. A strict Bonferroni correction (i.e., $p = 0.05/32 \approx 0.0015$ as threshold) to correct for comparison between different nodes. For nodes defined by the ROI, we found that PD subjects had significantly decreased functional connectivity in the right putamen ($p < 0.0015$). In addition, the functional connectivity had a trend toward decrease in the left SPL ($p = 0.038$), left IPL ($p = 0.031$), and a relative increase in the right M1 ($p = 0.035$) with PD subjects compared to normal subjects.

![Figure 4.16: The normalized total connectivity degree $\bar{\Gamma}$ of the left hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using ROIs defined by our pre-processing pipeline. The brain regions with larger $\bar{\Gamma}$ are considered as more important nodes in the motor network.](image-url)
4.2. fMRI Dataset

Figure 4.17: The normalized total connectivity degree $\bar{\Gamma}$ of the right hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using ROIs defined by our pre-processing pipeline. The brain regions with larger $\bar{\Gamma}$ are considered as more important nodes in the motor network.

Figure 4.18: The difference in the normalized total connectivity degree $\bar{\Gamma}$ in each motor-related brain region using rs-fMRI between normal and PD subjects using ROIs defined by our pre-processing pipeline. Data are shown as the normalized total connectivity degree $\bar{\Gamma}$ of normal subjects minus PD subjects in each brain region. *$p < 0.05$, two-sample t-test; **$p < 0.01$, two-sample t-test; ***$p < 0.0015$, two-sample t-test.
4.2. fMRI Dataset

For the second test, the normalized total connectivity degree $\overline{\Gamma}$ between the largest cluster defined by our FDR-controlled agglomerative clustering algorithm and the ROIs defined by our preprocessing pipeline are shown in Figure 4.19 and Figure 4.20. We used the Grubb's test to determine which clusters should be excluded if the clusters are too small (i.e. left cerebellum in P009) or too big. Cluster sizes are considered to be outliers if the inequality in Eq. (4.1) holds true.

$$Z_{critical} = \frac{N - 1}{\sqrt{N}} \sqrt{\frac{t^2_{\alpha/(2N), N-2}}{N - 2 + t^2_{\alpha/(2N), N-2}}} < \frac{|Y_i - \overline{Y}|}{s},$$  \hspace{1cm} (4.1)

where $N$ denotes the number of subjects, $t_{\alpha/(2N), N-2}$ denotes the critical value of the t-distribution with $N - 2$ degrees of freedom and a significance level of $\alpha/(2N)$, $Y_i$ is the number of voxels in a cluster of a region, the $\overline{Y}$ and $s$ are the mean and standard deviation of $Y$. Clusters are not part of our analysis if $|Y_i - \overline{Y}| > 2.21$ for normal subjects ($N = 9$) and $|Y_i - \overline{Y}| > 2.34$ for PD subjects ($N = 11$).

The difference in the normalized total connectivity degree $\overline{\Gamma}$ for the largest cluster node is shown in Figure 4.21. For nodes defined by the largest cluster, we found that PD subjects at off state trended toward decrease in the left DLPFC ($p = 0.005$), left SPL ($p = 0.008$), left putamen ($p = 0.01$), right SPL ($p = 0.008$), and right putamen ($p = 0.005$) compared to normal subjects.
4.2. fMRI Dataset

Figure 4.19: The normalized total connectivity degree $\bar{\Gamma}$ of the left hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using nodes defined by the largest cluster of each region and the ROIs defined by our preprocessing pipeline. The brain regions with larger $\bar{\Gamma}$ are considered as more important nodes in the motor network.

Figure 4.20: The normalized total connectivity degree $\bar{\Gamma}$ of the right hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using nodes defined by the largest cluster of each region and the ROIs defined by our preprocessing pipeline. The brain regions with larger $\bar{\Gamma}$ are considered as more important nodes in the motor network.
4.2. fMRI Dataset

Figure 4.21: The difference in the normalized total connectivity degree $\bar{\Gamma}$ in each motor-related brain region using rs-fMRI between normal and PD subjects using nodes defined by the largest cluster of each region and the ROIs defined by our preprocessing pipeline. Data are shown as the normalized total connectivity degree $\bar{\Gamma}$ of normal subjects minus PD subjects for each node. **$p < 0.01$, two-sample t-test.

For the last test, we defined the seeds to be small and functional clusters of each region (i.e. clusters are not defined by a large ROI) of each subject. An example of how these seeds are formed is shown in Figure 4.22.
4.2. fMRI Dataset

Figure 4.22: Example transformations of small clusters into a seed cluster. The largest cluster is shown in orange and the seed is shown in yellow.

The normalized total connectivity degree $\bar{\Gamma}$ between the seed and the ROIs defined by our preprocessing pipeline are shown in Figure 4.23 and Figure 4.24. The difference in the connectivity for ROI node is shown in Figure 4.25. For nodes defined by the seed, we found that PD subjects at off state had significantly decreased functional connectivity in the right M1 ($p < 0.0015$). In addition, the functional connectivity had a trend toward
4.2. fMRI Dataset

decrease in the left SPL ($p = 0.038$), right occipital-parietal cortex (OPC) ($p = 0.018$), right putamen ($p = 0.0019$) and a relative increase in the left SMA ($p = 0.0071$), right M1 ($p = 0.035$), right SMA ($p = 0.037$), right thalamus ($p = 0.035$) compared to normal subjects.

Figure 4.23: The normalized total connectivity degree $\bar{\Gamma}$ of the left hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using nodes defined by the seeds and the ROIs defined by our pre-processing pipeline. The brain regions with larger $\bar{\Gamma}$ are considered as more important nodes in the motor network.
4.2. fMRI Dataset

Figure 4.24: The normalized total connectivity degree $\bar{\Gamma}$ of the right hemisphere of motor-related brain region using rs-fMRI across normal and PD subjects using seeds using nodes defined by the seeds and the ROIs defined by our preprocessing pipeline. The brain regions with larger $\bar{\Gamma}$ are considered as more important nodes in the motor network.

Figure 4.25: The difference in the normalized total connectivity degree $\bar{\Gamma}$ in each motor-related brain region using rs-fMRI between normal and PD subjects using nodes defined by the seeds and the ROIs defined by our preprocessing pipeline. Data are shown as the normalized total connectivity degree $\bar{\Gamma}$ of normal subjects minus PD subjects for each node. *$p < 0.05$, two-sample t-test; **$p < 0.01$, two-sample t-test; ***$p < 0.0015$, two-sample t-test.
4.2. fMRI Dataset

Correlation analysis of normalized total connectivity degree $\bar{\Gamma}$ at each node against the UPDRS motor score in PD subjects shows negative correlation in the left posterior cingulate cortex (PCC) ($\rho = -0.72$) and right PCC ($\rho = -0.49$) and positive correlations in the right P-SMA ($\rho = 0.45$) using ROIs as nodes. The left PCC ($\rho = -0.76$) using the largest clusters of each region as nodes, left PCC ($\rho = -0.50$) and right PCC ($\rho = -0.44$) and positive correlations in the right P-SMA ($\rho = 0.53$) and right Pa ($\rho = 0.62$) using seeds as nodes. The correlations between other motor regions and the UPDRS score are shown in Figure 4.26, Figure 4.27, and Figure 4.28.

Figure 4.26: Correlation between the UPDRS and normalized total connectivity degree $\bar{\Gamma}$ at each node defined by the ROIs of each motor region in PD subjects.
4.2. fMRI Dataset

Figure 4.27: Correlation between the UPDRS and normalized total connectivity degree $\Gamma$ at each node defined by the largest cluster of each motor region in PD subjects.

Figure 4.28: Correlation between the UPDRS and normalized total connectivity degree $\Gamma$ at each node defined by the seed cluster of each motor region in PD subjects.
Chapter 5

Conclusion and Future Work

5.1 Discussions and Conclusions

In this thesis, we presented a framework for functional brain parcellation using rs-fMRI data with the application of exploring functional connectivity using those parcels. The framework builds on the significance-based modularity partitioning algorithm proposed by Chang et al. [8] The proposed framework initializes parcels with highly correlated neighboring voxels in each region. The parcels are then merged such that if the distribution of parcels in the network is significantly greater than that of random networks. The process is stopped when $p < 0.05$ (FDR corrected).

We tested the FDR-controlled agglomerative clustering algorithm on synthetic datasets and compared it with two other state-of-art methods. It is shown that the our FDR-controlled agglomerative clustering algorithm yields better parcellations compared the method proposed by Chang et al. [8] and Ward hierarchical agglomerative clustering algorithm. In particular, the FDR-controlled agglomerative clustering algorithm is able to parcellate even when there are lower intra-correlation and higher inter-correlation. Afterwards, we applied our FDR-controlled agglomerative clustering algorithm to a fMRI dataset with nine healthy normal subjects and eleven PD subjects. The parcellation results typically have one large cluster and many small clusters in certain regions of the brain such as the bilateral cerebellum and the bilateral M1.

The normalized total connectivity degree $\bar{\Gamma}$ between different nodes of the brain reflects how well the nodes are connected to one another in a network. Thus, if a region has a high degree of connectivity, it means that the region has highly functional connections with multiple other brain regions, and is thus important in the network. We were able to find significant increase in right M1 between normal and PD subjects when using seed nodes compared to using ROIs as nodes. However, we were not able to find a significant decrease in the right putamen using seed nodes compared to using ROIs as nodes, whereas there is a significant decrease shown using ROIs nodes. This is consistent with the study by Wu et al. [47] except that they found
a significant decreased in functional connectivity in the SMA, left DLPFC and left putamen, and a significant increased in functional connectivity in the left cerebellum, and left M1 compared to normal subjects. Even though not significant, we found a trend that there is a decrease in the left DLPFC, bilateral SPL, right IPL, putamen indicating that these regions have less functionally connected with other motor regions. In contrast, the increased of functional connectivity in M1 and PCC, suggests that these areas might be more important role in the motor network in PD as compared to controls. We did not find an increase in functional connectivity in the cerebellum and did not find a decrease in functional connectivity in the SMA between normal and PD subjects. Lastly, we also found that the functional connectivity in PCC is negatively correlated with the UPDRS.

### 5.2 Future Work

The framework in this thesis could be improved with modifications in the pre-processing pipeline and the FDR-controlled agglomerative clustering algorithm. The fMRI dataset was co-registered with structural images and then propagated back to the subject’s native space instead of using anatomical template such as Talairach or MNI. If the fMRI datasets were registered with anatomical templates, then this would allow for group-level parcellation. This would allow for more consistent functional connectivity analysis using rs-fMRI. The FDR-controlled agglomerative clustering algorithm produced clusters with less than 5 voxels. This may not be significant or meaningful in functional connectivity analysis. We suggest to initialized a certain number of clusters instead of clustering highly correlated fMRI time-series. Afterwards, apply the FDR-controlled agglomerative clustering algorithm. This might produce different clusters between normal and PD subjects.

The proposed FDR-controlled agglomerative clustering algorithm can be applied to other datasets to determine functional brain regions. These datasets include task-based fMRI to see if there are any differences in regions related to certain tasks between normal and PD subjects. Moreover, the framework could be applied to other neurological diseases such as Alzheimer’s disease and Autism to see if there are any functional regions that are consistent and different between subjects.
Bibliography


