Multi-Modal Tensor Representations of Brain Networks

by

Göktekin Durusoy

B.S., Electronics Engineering, Uludağ University, 2010

M.S., Electrical Engineering, New York University, 2013

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Electrical and Electronics Engineering

Boğaziçi University

2021

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep sense of gratitude and profound respect to my supervisor Prof. Burak Acar who has helped and encouraged me at all stages of my research with great patience, motivation, enthusiasm, and immense care. This dissertation would not have been possible without his support, guidance, and persistent help.

I owe my gratitude to Prof. Ali Taylan Cemgil for his guidance about tensor factorization techniques, and to Prof. Murat Saraçlar and Assoc. Prof. Alkan Kabakçıoğlu for their constructive feedback and valuable discussions during this research. I would also like to thank Prof. Yasemin Gürsoy Özdemir for being in the thesis committee.

I would like to thank the faculty members and staff of Boğaziçi University and Namık Kemal University for their endless support, especially Lütfi Ulusoy and Serap Kılınç for providing a friendly environment at our workplace. In addition, I would like to acknowledge all Volumetric Analysis & Visualization Group (VAVlab) members for their friendship even if we could not be present in the lab in this challenging year. Moreover, I would like to give very special thanks to Demet Yüksel Dal, Abdullah Karaaslanlı, Zeynep Kahraman and Dr. Okan İrfanoğlu for their contributions to the preprocessing; and to Güneş Bayır, Gurur Gamgam for providing a machine learning perspective.

I acknowledge my deep sense of gratitude and indebtedness to all researchers who take part in the VAVlab BRAINet project at Istanbul Faculty of Medicine, Istanbul University. Their collaborative efforts made this thesis possible. Among them, I would like to thank especially Prof. Hakan Gürvit, and Zerrin Yıldırım for their contributions to the neurological discussions in this thesis.

The research for this thesis was in part financially supported by the Turkish Directorate of Strategy and Budget under the TAM Project number 2007K12-873, by TUBITAK-ARDEB 1003 Programme under Grant #114E053 and by Bogazici University Research Fund Grant #16862. I would also like to thank Prof. Cem Ersoy and TETAM family for their innovative comments, suggestions, and discussions throughout my research.

I would like to express my sincere thanks to all my friends, especially Can Zengin and Uğur Ertaş for their emotional & moral support, and even with their helpful comments and criticism over the thesis. I feel lucky to have them in my life.

I would like to show my deepest gratitude to my parents, in-laws for their continuous support, eternal love, and encouragement in my whole life. Special thanks goes to my sister, Göksenin, for everything she has done.

Last but not least, I would like to thank my beloved wife, Demet, for her unwavering moral & emotional support, and tolerance of the unremitting unsocial behavior caused by the hours and days that this thesis consumed. Love and a big hug to my son, Ahmet Salih, even if he was the biggest distraction. No worries, I will be at home soon.

ABSTRACT

Multi-Modal Tensor Representations of Brain Networks

Considering the economic, social and psychological burdens of Alzheimer's disease (AD), the most common form of dementia, it is essential to gain insight into the process and underlying mechanisms of the disease. Using structural and functional brain connectomes obtained by in-vivo MRI techniques as biomarkers is a promising approach. In this thesis, the B-Tensor structure that allows the representation of brain connectomes defined in structurally and functionally with a uni-modal and multi-modal fashion is presented. With the projection of structural connectomes onto known functional networks, patients with AD and healthy control group are distinguished in a 7-dimensional space with a separation performance of over 90%. In addition, with the uni-modal and multi-modal tensor factorization methods, 47 patients with different levels of AD, are diagnosed with an accuracy of 77% - 100% in a 5-dimensional space. The results show that the multi-modal factorization technique performs better than the uni-modal one by successfully fusing the structural and functional networks which offer complementary information. While the neurological evaluations of the obtained sub-networks are highly consistent with previous literature, new findings regarding the progression of the disease are also recommended.

ÖZET

Beyin Ağlarının Çok Modelli Tensör Gösterimleri

Demansın en yaygın şekli olan Alzheimer hastalığının ekonomik, sosyal ve psikolojik yükleri göz önünde bulundurulduğunda, hastalığın süreci ve altında yatan mekanizmaları hakkında fikir edinmek önemlidir. Biyobelirteçler olarak in-vivo MRI teknikleri ile elde edilen yapısal ve fonksiyonel beyin konnektomlarının kullanılması umut verici bir yaklaşımdır. Bu tezde, yapısal ve işlevsel olarak tanımlanan beyin konnektomlarının tek modlu ve çok modlu bir şekilde temsiline izin veren B-tensör yapısı sunulmuştur. Yapısal bağlantıların bilinen fonksiyonel ağlara projeksiyonu ile AD ve sağlıklı kontrol grubu, %90'ın üzerinde bir ayırma performansı ile 7 boyutlu bir uzayda ayırt edilmiştir. Aynı zamanda 47 kişiden oluşan farklı AD düzeylerine sahip hastalara tek modlu ve çok modlu tensör çarpanlarına ayırma yöntemleri ile 5 boyutlu uzayda %77-%100 doğrulukta tanı konulmaktadır. Tamamlayıcı bilgi sunan yapısal ve işlevsel ağları başarılı bir şekilde birleştirerek, çok-modlu tensör çarpanlarına ayırma tekniğinin tek-modludan daha iyi performans gösterdiği bulunmuştur. Elde edilen alt ağların nörolojik değerlendirmeleri literatürle son derece uyumlu olmakla birlikte, hastalığın ilerlemesine ilişkin yeni bulgular da önermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

A	Low level factor matrix related with patients
b	B-value characterizing the gradient pulses
B_{cm}	Composed multi-modal 4D B-Tensor
B_f	Uni-modal 3D functional B-Tensor
B_{fm}	Multi-modal 4D functional B-Tensor
B_m	Multi-modal 4D B-Tensor
B_s	Uni-modal 3D structural B-Tensor
B_{sm}	Multi-modal 4D structural B-Tensor
B_u	Uni-modal 3D B-Tensor
BC	Average betweenness centrality
BC_i	Betweenness centrality of node i
C	Number of classes
C_{av}	Average clustering coefficient
C_i	Clustering Coefficient of node i
$\hat{C}$	Clustering coefficient of a corresponding random graph
D	Low level factor matrix related with nodes
d_{ij}	Geodesic distance between node i, and j
d_q	CP scaling factor of q^{th} sub-network
E	Low level factor matrix related with modalities
$E_{max}(.)$	A function finding the eigenvector of the maximum eigenvalue
E_G	Global efficiency
e_k	End point of the k^{th} fiber
e^a_{ij}	Average FA-based structural edge association between parcels P_i and P_j
e^h_{ij}	Highest FA-based structural edge association between parcels P_i and P_j
e^l_{ij}	Lowest FA-based structural edge association between parcels P_i and P_j

e_{ij}^{m}	Median FA-based structural edge association between parcels P_i and P_j
e_{ij}^{Mc}	Maximum correlation of functional edge association between parcels P_i and P_j
e_{ij}^{n}	Normalized Weighted structural edge association between parcels P_i and P_j
e_{ij}^{Pc}	Partial correlation of functional edge association between parcels P_i and P_j
e_{ij}^{Pe}	Pearson correlation of functional edge association between parcels P_i and P_j
e_{ij}^{w}	Weighted structural edge association between parcels P_i and P_j
F	Low level factor matrix related with nodes
FA	Fractional Anisotropy
G	Known basis sub-networks
g_k	Magnetic field direction of k^{th} observation
k_i	Degree of node i
L	Average shortest path length
$\hat{L}$	Average shortest path length of a corresponding random graph
M	Compact & multi-modal representation of brain networks
N	Number of vertices in a graph
N_c	Number of patients in class c
n_{jk}	Number of shortest paths between node j, and k
$n_{jk}(i)$	Shortest paths between node j, and k that pass through node i
P_i	i^{th} parcel in brain
$PCC(u)$	Pearson correlation coefficient when one signal is shifted by u
Q	Number of sub-networks
r_j	j^{th} voxel position
SW	Small-worldness coefficient

S_k	Image intensity at a voxel when magnetic field is applied in g_k direction
S_0	Reference image intensity at a voxel
u	Shift parameter
U_f	Functional reconstruction coefficients
U_s	Structural reconstruction coefficients
u_q^c	Projection coefficient of q^{th} sub-network in class c
V	Set of vertices in a graph
V_i	Volume of parcel P_i
V_q	Basis q^{th} sub-network
v_q	Column vector related with the q^{th} sub-network
$V_{q,f}$	q^{th} sub-network of functional connectomes
$V_{q,s}$	q^{th} sub-network of structural connectomes
W_{ik}	Association between i^{th} parcel and k^{th} fiber
w_{ij}	Connectome weight between node i, and j
$\tilde{w_{ij}}$	Normalized connectome weight between node i, and j
w_q	Modality scaling parameter of q^{th} sub-network
w_q^*	Tied modality scaling parameter of q^{th} sub-network
$y_i(t)$	Summary signal at time point t of the parcel P_i
α	Density ratio value between fNET and sNET
Γ	Projection matrix
θ	Vector of density ratio
λ_k	k^{th} Lagrange multiplier
$\tilde{\lambda}_k$	k^{th} eigenvalue
$\hat{\lambda}$	Mean of eigenvalues
μ	Mean value
σ	Standard deviation
σ_i	Standard deviation of summary signal i
$(\Sigma^{-1})_{ij}$	i^{th} row and j^{th} column of inverse covariance matrix

LIST OF ACRONYMS/ABBREVIATIONS

3D	Three Dimensional
4D	Four Dimensional
Aβ	Amyloid beta
AC	Anterior Commissure
AD	Alzheimer's Disease
ADD	Activity Dependent Degeneration
AFC	Atomic Functional Connectome
ALS	Alternating Least Squares
BOLD	Blood Oxygen Level Dependent
B-Tensor	Brain Tensor
CDR	Clinical Dementia Rating
CFI-S	Cognitive Function Instrument Subject
CFI-SP	Cognitive Function Instrument Study Partner
CI	Cue Index
CP	Canonical Polyadic
CSF	Cerebrospinal Fluid
DAN	Dorsal Attention Network
DICCCOL	Dense Individualized and Common Connectivity-based Cortical Landmarks
DMN	Default Mode Network
dMRI	Diffusion Magnetic Resonance Imaging
DT	Diffusion Tensor
DWI	Diffusion Weighted Imaging
DW-MRI	Diffusion Weighted Magnetic Resonance Imaging
EEG	Electroencephalography
EPI	Echo Planar Imaging
FA	Fractional Anisotropy
FC-SRT	Free and Cued Selective Reminding Test

FDR	False Discovery Rate
FFE	Fast Field Echo
fMRI	Functional Magnetic Resonance Imaging
fNET	Functional Network
FWER	Family-Wise Error Rate
GD	Gradient Descent
GEVD	Generalized Eigenvalue Decomposition
GM	Gray Matter
HCP	Human Connectome Project
ICA	Independent Component Analysis
IQ	Intelligence Quotient
KDE	Kernel Density Estimation
LDA	Linear Discriminant Analysis
LOOCV	Leave One Out Cross Validation
MANOVA	Multivariate Analysis of Variance
MB-GD	Mini Batch Gradient Descent
MCI	Mild Cognitive Impairment
MGLM	Multivariate General Linear Model
MRI	Magnetic Resonance Imaging
NBS	Network Based Statistic
NFT	Neurofibrillary Tangles
NIA-AA	National Institute on Aging/Alzheimer's Association
NMWF	Non-negative Multi-way Factorization
NODDI	Neurite Orientation Dispersion and Density Imaging
PARAFAC	Parallel Factor Analysis
PC	Posterior Commissure
PCA	Principal Component Analysis
PCC	Posterior Cingulate Cortex
PGSE	Pulse Gradient Spin Echo
RK4	4^{th} order Runge-Kutta
rs-fMRI	Resting State Functional Magnetic Resonance Imaging

SCI	Subjective Cognitive Impairment
SCP	Sparse Connectivity Patterns
SGD	Stochastic Gradient Descent
sNET	Structural Network
SPC	Spatial Pairwise Clustering
SPL	Superior Parietal Lobule
SVD	Singular Value Decomposition
SVD-RG	Singular Value Decomposition Region Growing
SVM	Support Vector Machine
T1b	T1w with Skull Stripped
T1w	T1-weighted
T2b	T2w with Skull Stripped
T2w	T2-weighted
TFE	Turbo Field Echo
TFR	Total Free Recall
TN-PCA	Tensor Network Principal Component Analysis
WM	White Matter
VAN	Ventral Attention Network

1. INTRODUCTION

With the increase in the elderly population, neurodegenerative diseases have started to be investigated more by researchers. While the world population over the age of 65 was 703 million by 2019, this number is expected to double in 2050. Consequently, this will cause an increase in Alzheimer's Disease(AD), which is one of the age-related diseases [1].

AD should not be confused with normal aging and dementia. Although severe memory loss is one of the key characteristics of AD, it is not one of the normal aging symptoms. In normal aging, hair, weight, height and muscle loss may occur gradually, or skin and bones may become sensitive. In addition, there may be a decline in hearing, vision and metabolic rate of the body. Although there may be some memory problems, such as taking longer to learn or remembering information, a cognitive loss that affects daily life is not seen in normal aging. Dementia, on the other hand, is defined as a significant loss of cognitive abilities that can seriously affect social life. Dementia, unlike AD, is not a specific disease name, it is the umbrella term describing all diseases that are accompanied by impairment of memory and mental abilities. It can be caused by various diseases that damage brain cells, and AD is just the most common form of dementia [1–4].

Symptoms in AD can be divided into two categories which are clinical and pathological. Neurologists try to understand the clinical symptoms by applying tests to see whether individuals experience any problems with memory, visual, or even language. Pathological symptoms represent the changes inside of the brain. Two hallmarks of AD have been widely studied: Amyloid beta (Aβ) and p-tau [5]. With respect to the amyloid cascade hypothesis, abnormal accumulations inside the cerebrospinal fluid affect tau aggregation which cause brain cells to die and finally lead to memory loss. Moreover, Aβ and p-tau in patients with AD are targeted in clinical trials which have unsuccessful results. This is probably because the accumulation of tau and Aβ begin

years before the cognitive symptoms,that is to say, the pathological symptoms begin before the clinical diagnostic of a patient. Therefore, any contribution to diagnose patients with early AD will be beneficial which is one of the contribution of this thesis.

The increase in the prevalence of AD worldwide has necessitated the development of non-invasive bio-markers. Thanks to the magnetic resonance imaging (MRI) technology, researchers have the opportunity of noninvasive diagnosis. Functional (fNET) and structural network (sNET) models (also called connectome) can be constructed by using functional MRI (fMRI) and diffusion MRI (dMRI) modalities. Connnectome is a map of neural systems in human brains into networks which is utilized to explain the underlying function and structure of the brain. There are two things important in a graph representation as can be seen in Figure 1.1, one is nodes/regions and the other one is associations/edges between them.

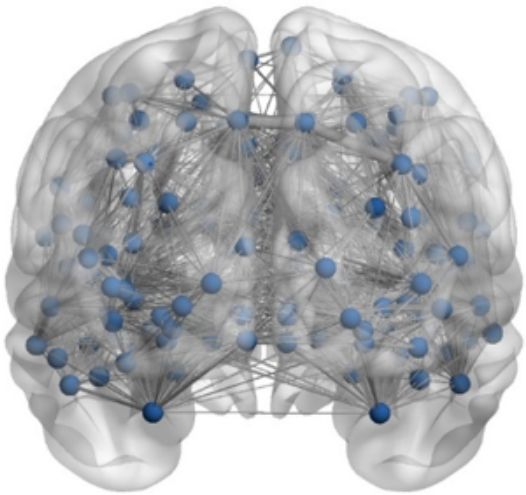

Figure 1.1. Graph representation of the brain. Blue spheres represent nodes/regions and gray lines illustrates connections/associations between regions [6].

Information about how the brain works and the underlying causes and the progress of neurodegenerative diseases can be obtained by analyzing the network models. The idea of representing human brain with networks supports the connectionist point of view which states the underlying functioning of brain is related with the changes of

connections between regions. In the connectionist point of view, a cognitive deficit in brain is related with a set of disconnection processes between brain regions as opposed to one processing module [7]. Although the causes of AD and the mechanism behind it are not fully understood, it is increasingly promising to define it as a disconnection syndrome in which the connections between cortical brain regions are disrupted [7, 8]. Therefore understanding the spectrum of dementia would benefit immensely from identifying disrupted functional/structural associations between cortical regions.

Brain is a complex structure and its functioning is still an open research area. Theories targeting cognition of brain are still a hot debate and what we wonder are deep enough such as understanding the underlying structure and the progress of AD. Moreover, the biggest data used in these studies do not consist of more than few hundreds patients. For these reasons, interpretation and intuition of the models become important. In this thesis, mathematical models are generated to computationally examine patients with AD. Data obtained from different imaging modalities are constructing a relation between each other, and tensors help us to use them together. The models in this thesis are constructed by considering their interpretation and not specific to AD, but also have the ability to shed light on other neurodegenerative diseases. Hence, the other contributions of this thesis are the mathematical models which can be utilized to address neurological point of view in a disease.

Then, the question is whether the proposed structure of the mathematical models do show any relations with the clinical diagnoses. This is a standard classification problem. Hence, performances of the formulated models are obtained via solving a classification problem defined on the spectrum of AD which is divided into stages clinically, such as subjective cognitive impairment (SCI), mild cognitive impairment (MCI) and AD. The proposed mathematical models provide feature vectors which can be rephrased as minimum dimensional brain connectome representations in terms of the discovered basis networks. Now the second question arises about the discovered basis networks: what do they tell about the disease? Discriminative behavior of these networks have the ability to provide information about disease continuum. Hence,

strongly associated regions are discussed with a neurological perspective.

Finally, the overall objective of this thesis is to develop methods to discover and assess the disruptions in functional and structural associations between cortical regions throughout the disease spectrum by using fMRI and dMRI as non-invasive techniques. To sum up, main contributions of this thesis can be listed as: (i) to discover a minimal dimensional discriminative connectome space spanned by basis networks, (ii) to get minimum dimensional brain connectome representations in terms of the discovered basis networks, (iii) to assess the discriminative power of the minimum dimensional connectome representations throughout the spectrum of dementia.

The rest of this thesis is organized as follows: Section II describes state of the art connectome analysis which is categorized into graph theoretical approaches, statistical analysis and factorization based techniques. Section III first gives the idea of generating brain networks, and then discusses the factorization of these by using different models. Outcomes of different factorization techniques with real data can be found in Section IV. Discussion of the results is presented in Section V from both neurological and data science point of view. Section VI concludes the research and offers future perspective.

2. STATE-OF-ART: CONNECTOME ANALYSIS

2.1. Graph Theoretical Approaches

A network, or equivalently a graph, is a mathematical structure composed of a set of nodes and directed/undirected edges between these nodes. In brain networks at the macroscopic scale, those nodes represent cortical regions and the edges are identified as a measure of associations (connectivity) between these regions. Early graph theoretical approaches mostly compare the differences of topological network properties between groups [9]. Node degree, clustering coefficient, path length, small-worldness, and global connectivity are the mostly utilized features which are used to separate AD from healthy controls [10].

When a given network is compared with another reference network or different types of networks, it will be useful to use global properties such as clustering coefficients, average path length. In short, examined network is considered as a whole when global features are used, and hence the local differences cannot be observed. Typically, the global features have high sensitivity but low specificity with respect to a certain disease. On the other hand, when the properties of individual sub-components are studied, local features such as local clustering coefficient and shortest path become important [11].

Clustering coefficient is defined for one node as the ratio of sum of its direct neighbors' geometric mean to all possible connections, and it measures how well its neighbors are connected to each other. Mostly, it can be considered as a resilience power of a network when a random error occurs. If one of the edges are lost, other edges that connect the node (neighbors) still exist. Onnela et al. [12] expressed clustering coefficient as:

$$C_i = \frac{2}{k_i(k_i-1)} \sum_{j,k \in V} (\tilde{w_{ij}} \tilde{w_{jk}} \tilde{w_{ik}})^{1/3} \ , \ C_{av} = \frac{1}{N} \sum_{i \in V} C_i \tag{2.1}$$

where $\tilde{w_{ij}}$ represents the normalized connection between node i and j which is obtained by dividing the weights to the maximum weight in the graph, V is the set of vertices in the graph, k_i is the degree of node i. C_{av} is the average clustering coefficient and N is the total number of vertices. There are some conflicting results in literature which state that average normalized clustering coefficients are decreased/increased in AD cases [13–16]. In addition, some researchers could not find any differences for the average normalized clustering coefficients between AD and healthy controls [17–20].

Average shortest path length, also known as characteristic path length, is defined as the average of minimum number of edges that connect any two nodes of the graph, and measures how well information is transmitted by the graph, and formulated as:

$$L = \frac{1}{N(N-1)} \sum_{i,j \in V, i \neq j} d_{ij} \tag{2.2}$$

by Watts and Strogatz [21, 22] where d_{ij} is geodesic distance between nodes i and j which is reciprocal of connectivity weight between those nodes, N is the total number of vertices. This formulation could be problematic when there is no connection between any two nodes, which forces the distance to be infinite. Therefore, as an alternative approach, using the harmonic mean of geodesic distance is proposed by Newman [23], and it is called global efficiency which can be written as:

$$E_G = \frac{1}{N(N-1)} \sum_{i,j \in V, i \neq j} \frac{1}{d_{ij}} \tag{2.3}$$

In many studies, average shortest path length is found to be increased for AD patients which can be interpreted as connection loss [10, 13–15, 18, 24]. However, there are also existing studies that show decreased average path length for AD patients [17, 25]. Additionally, Supekar et al. could not find any significant difference for global efficiency between AD and healthy controls [16].

Small-world property of a graph is defined as the ratio of average clustering coefficient to the average shortest path length which are normalized with respect to a random graph, and measures the balance between functional integration and segregation [26]. Using the above definitions of clustering coefficient and shortest path length, small-worldness can be defined as:

$$SW = \frac{C_{av}/\hat{C}}{L/\hat{L}} \tag{2.4}$$

where $\hat{C}$ and $\hat{L}$ represent the average clustering coefficient, and the average shortest path length of corresponding random graph, respectively [27]. Many researchers stated that the small-world graph property has been observed for both AD patients and healthy controls which proves that those graphs are different from a corresponding random graphs [16,18,24,25,28]. However, there are studies that claim both an increase and a decrease in small-worldness in AD patients with respect to control groups [16, 24, 25, 28].

Betweenness centrality of node i is defined as the ratio of the number of shortest paths that passes through node i to the number of shortest paths, and aggregated over all node pairs in a graph, and measures the importance of a node in a graph [29]. Freeman defined it as:

$$BC_i = \sum_{j \neq k, j,k \in V} \frac{n_{jk}(i)}{n_{jk}} \ , \ BC = \frac{1}{N} \sum_{i \in V} BC_i \tag{2.5}$$

where $n_{jk}(i)$ represents the shortest paths between node j, and k that run through a node i, n_{jk} is the number of shortest paths that connect node j, and k. In addition, BC represents the average betweenness centrality of a graph. It has been widely used to identify hubs in a graph [13, 15, 18, 30–32]. It has been found that temporal, insula, cingulate and parietal regions are affected by AD, for a review, see [10].

K-core decomposition of a graph is recursively calculated for a given k by extracting nodes and corresponding edges if node degree is less than k [33]. For instance,

to calculate 2-core decomposition of an undirected graph, first 1-core decomposition should be applied to graph such that all the nodes and corresponding edges are removed if their node degree is less than 1. Then remaining nodes with degree lower than 2 should be removed from the graph as well as corresponding edges. Highly interconnected nodes are found by using k-core decomposition. Daianu et al. implement k-core decomposition algorithm to obtain sub-graphs, and each k is interpreted as a threshold constraint for node degree [34]. Above-mentioned features are used in those sub-graphs to differentiate AD vs mild cognitive impairment (MCI) and healthy controls to obtain changes when the disease progresses. They have found that normalized characteristic path length, and global efficiency are decreased while normalized clustering coefficient and small-worldness are increased for AD patients, relative to controls. In another study, it has been preferred using rich club with a k-core algorithm on the highly connected graphs to assess the effects of AD [35].

In addition to above graph features, there are several other graph properties such as modularity, assortativity, participation coefficient, local efficiency, global vulnerability [10, 36–40]. Even though graph features show important properties of brain connectomes, none of them can develop an insight about disrupted connections which is a characteristic of AD as a disconnection syndrome [10]. To sum up, having relatively high sensitivity and low specificity is the vulnerable part of the graph theoretical approaches which provides very limited insight about the underlying disease progress [41].

2.2. Statistical Analysis

After constructing networks called connectome, their interaction with higher-order cognitive processes or their effects in terms of diagnosing the disease should be examined such as cognitive/behavioral abilities, personality traits, task performance or disease severity, and prognosis [42–44]. These examinations are often based on performing statistical analysis using data from a population-wide or an individuals over time (longitudinal) [43]. Many statistical analyses are suitable for the use of connectomes derived from functional or structural associations. The main problem

arises in the interpretation of the obtained results. Functional connections can often be positive and negative, and often weighted connections are used instead of binary ones. On the other hand, structural networks are nonnegative [45]. In addition, even though we can think of structural associations as links/paths through which information is transmitted, it is not possible to interpret the same logic for functional networks [46].

Commonly used statistical methods perform univariate analyses to reveal the differences between groups on each random variable by assuming each network edge as an independent random variable [43]. Examples of such univariate approaches are t-test, f-test and regression analysis. In this way, researchers can identify relationships for categorical or dimensional variables and edge weights [47]. However, such approaches require a lot of correction to control the number of false positives. False discovery rate, one of the standard correction techniques, may not give correct results since it does not consider the relationship between edges [48]. Therefore, the weakness of these approaches is that they do not consider the dependency between edges, that is to say, they do not include an information about the structure of connectome groups [47].

Network Based Statistics (NBS) is proposed to include edge dependency by Zalesky et al. [49]. They developed a method to find statistically significant edges that differ between classes. At first, they perform two sample t-test at each edge independently with the null hypothesis that two populations come from the same distributions with equal means. In order to construct a set of suprathreshold edges, t-statistic at each edge is thresholded. Connected components of suprathresholded edges which is also referred as observed components, are identified with a bread first search [50] and the size of the largest component is stored. This procedure is implemented independently M times and each time group labels are randomly assigned to each member. Therefore, the null distribution of maximal component size is obtained. Later on by using the original data, a p-value is assigned for each observed component with size k by dividing the number of permutations where observed components' size is greater than k to total number of permutations (M). As an alternative to NBS approach, Spatial Pairwise Clustering (SPC) is proposed with the same idea except identifying the

connected components which is obtain by considering clusters as pairwise suprathresholded edges [51]. Performing these methods by using all edges in brain connectome rather than using the connections in few regions of interest, and the nonparametric nature of them can be seen as advantages. However, it is worth noting that finding a statistically significant edge depends highly on a good threshold value of the t-statistic.

Meskaldji et al. proposed to assign summary statistics to the set of regions in order to identify statistically different subsets of connectomes between two populations which is grouped according to their intelligence quotients (IQs) [52]. First of all, they split connectomes into subsets by dividing the brain into meaningful lobes and found summary statistics for each defined subset. Later, based on these summary statistics, multiple corrections were made with respect to either Family-Wise Error Rate (FWER) or False Discovery Rate (FDR). However, the performance of this method is mostly based on the selection of the subsets [53].

Another study proposed by Kim et al. reveals that group differences between AD and healthy controls can be obtained via individual edge weights by performing multivariate general linear model (MGLM) [54]. In order to obtain multi-resolution features at each graph edge, they have applied line graph transform to connectomes and wavelet operations to these line graphs. Line graph representation of an undirected graph transforms the edges of a graph into vertices, and these vertices are connected to each other if corresponding edges of the graph share common vertex. P-values related with the MGLM are further corrected with multiple comparisons.

All of these statistical approaches have an important role in the analysis of detecting differences between groups of brain networks. In addition, once a subset is found to be statistically significant with the methods mentioned above, it can be said in which group edges are weak or strong. Statistical approaches are non-generative (non-constructive) and are based on a posterior analysis of observations. However, factorization is a generative (constructive) approach that constructs structures that show a difference.

2.3. Factorization Based Analysis

Discovering sub-networks which are composed of some regions in brain has gained popularity among researchers to obtain insight into neurodegenerative diseases. Decomposition/factorization based approaches assume that the observed data can be factorized as a multiplication of vectors/matrices or even tensors. Previous studies adopted matrix factorization methods, such as Principal Component Analysis (PCA) [55–57], Independent Component Analysis (ICA) [58–60], Non-negative Matrix Factorization (NMF) [61, 62] to obtain sub-networks. However those methods are applied directly to the time-series of blood-oxygen-level-dependent (BOLD) signal. These approaches mostly identifies motor, visual, and default mode networks etc. (representing functional activity) which are not necessarily discriminative between groups. Therefore, several studies discovering discriminative sub-networks applied factorization based approaches to the connectome data [63].

Atomic Functional Connectomes (AFC) is proposed to discover discriminative sub-networks between MCI and healthy controls by Jinli et al. [64]. Rather than using BOLD signals of every voxel in resting state fMRI (rs-FMRI), they prefer to use cortical landmarks named Dense Individualized and Common Connectivity-based Cortical Landmarks (DICCCOL) system to construct brain connectomes. Since connectivity matrices are symmetric, they use only the upper triangle part and vectorize it for concatenating purposes. Later on, non-negative matrix factorization is separately applied to two groups [65]. Since this is an optimization problem, every run ended up with different basis sub-networks (AFCs), and interpretation of them are made based on their average after several runs.

Eavani et al. proposed Sparse Connectivity Patterns (SCP) to discover discriminative sub-networks between children and young adults [66]. They construct joint optimization framework consists of matrix factorization and support vector machine (SVM) terms and add sparsity constraints to make results stable. In addition, relative weight of different terms also assessed by using values between 0 and 1. In another

study, SCPs are found to characterize functional connectivity while representing the inter-subject variability. They have found functionally (anti-)correlated regions, and compare their results with ICA approaches [67]. Both studies solve optimization problem via projected gradient method [68], and find global set of SCPs for corresponding data.

Zille et al. [69] further investigate SCPs and expand this study to obtain class specific SCPs by using fNETs of children and young adults. They performed coupled matrix factorization with common SCP, and used fused Lasso penalty. Shared and differential components are defined by thresholding the absolute differences of SCPs obtained from different classes. Later on, each differential SCP can be used as basis sub-networks and individual coefficients are found by projecting connectomes onto basis sub-network space. Regularization parameter of fused term can be problematic when there are more than two classes, hence careful consideration is necessary. In another study, connectomes are constructed with sliding window method [70]. Consequently, by using more than one connectome per subject, dynamic SCPs are found using slightly different optimization which forces additional connectome factorization coupled with the same SCPs. Found SCPs are further studied to obtain their discriminative behavior by using corresponding individual coefficients [71].

Durusoy et al. [72] proposed brain tensor (B-Tensor) representation to uncover discriminative behavior of anatomically defined sub-networks between AD and healthy controls by using a cohort of sNETs. They prefer to use tensor representation which enables them to include different connectome constructions. Salience sub-network is found to be the most discriminative one. However, fixing the basis sub-network is one of the limitations in this study.

Zhang et al. [73] used 3^{rd} order tensor representing a group of healthy subjects' sNETs to discover the relationship between structure of the brain and personal traits. They have applied tensor-network PCA (TN-PCA) decomposition which factorizes a group of sNETs into individual reconstruction coefficients and common sub-network

basis while posing an orthogonality to basis sub-networks. Regression analysis is also made by using individual reconstruction coefficients with their age and gender which act as a baseline. In another study, TN-PCA decomposition is used by assigning relative weights to different groups which will be more general optimization when classes are imbalanced [74]. Discriminative sub-networks have been found between AD, MCI and subjective cognitive impairment (SCI) patients by utilizing their fNETs, and sNETs. Further, multi-modality effect of brain connectomes on classification is researched by using parameter tying.

There are other studies that use factorization based approaches such as Non-negative Multi-way Factorization (NMWF) [75,76], Parallel Factor Analysis (PARAFAC) [77–79], sparse tensor decomposition [80,81], Tucker decomposition [82], multi-paradigm sparse tensor decomposition [83], structured sparse representation [84] which are applied to electroencephalography (EEG) and/or (task-)fMRI, dMRI data analysis. However, those studies are out of the scope of this thesis and we recommend readers to check those papers to gain insight about different research areas in neuroimaging analysis. In this dissertation, B-Tensor factorization approaches will be discussed in detail. In our previous work [72], known set of sub-networks are examined using only sNETs of AD and healthy controls; while in another study, Durusoy et al. have successfully combined sNETs and fNETs [74]. Hence, the proposed algorithm is appropriate to examine different modalities in terms of discovering discriminant sub-networks. Moreover, minimum dimensional basis sub-networks as well as corresponding individual reconstruction coefficients are discovered and their discriminative power is examined.

3. B-TENSOR MODEL

3.1. Brain Connectome Construction

Resting state fMRI, dMRI, T1-weighted (T1w) and T2-weighted (T2w) images obtained from patients should be co-registered in the common space with equal resolution to generate fNET and sNET which will be later used for cross-subject comparisons and multi-modal analysis of the brain connectivity. The connectome construction steps are presented in detail in the following sections.

3.1.1. Preprocessing

The preprocessing of the data was made using the well-known FSL [85], FreeSurfer [86] and Tortoise [87] programs. FreeSurfer is used to obtain atlas parcellation, diffusion-weighted magnetic resonance imaging (DWI, DW-MRI, dMRI) data is preprocessed via Tortoise, and FSL is utilized for both preprocessing of fMRI data, registering and masking purposes. In the next step, co-registered data were visualized and/or manually controlled by expert neurologists before network generation, and then poorly registered data were removed. Consequently, a fully automatic preprocessing has been applied to all the data without any manual correction. Preprocessing pipeline can be seen in Figure 3.1.

In FreeSurfer section of the preprocessing pipeline, the purple box in the Figure 3.1, T1w and T2w data are used as inputs, and the final outputs of this box are the processed T1w, T2w with/without skull and corresponding Destrieux Atlas parcellation. This section can be divided into two main processes which are described by Glasser et al. as pre-freesurfer (Figure 3.2) and freesurfer (Figure 3.3) pipelines [89]. Utilized scripts for this section are based on the Human Connectome Project (HCP) pipeline, which can be found on Github [90].

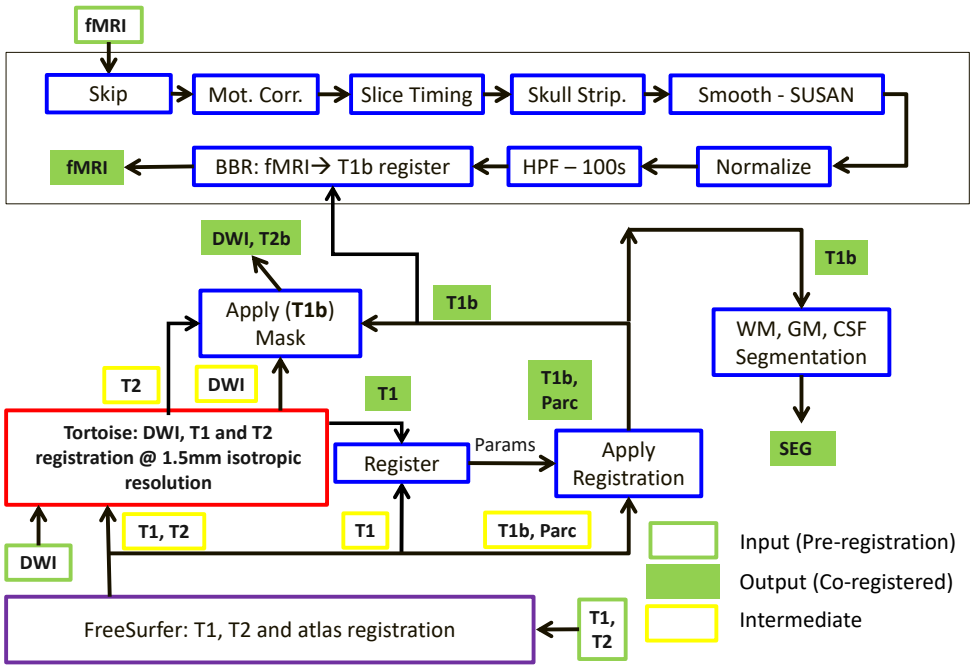

Figure 3.1. Preprocessing Pipeline [88]. Red, purple and blue boxes represent the usage of the Tortoise, FreeSurfer and FSL programs, respectively. T1/T2 and T1b/T2b represent T1w/T2w images and skull stripped versions, respectively. Parc stands for Destrieux Atlas parcellation.

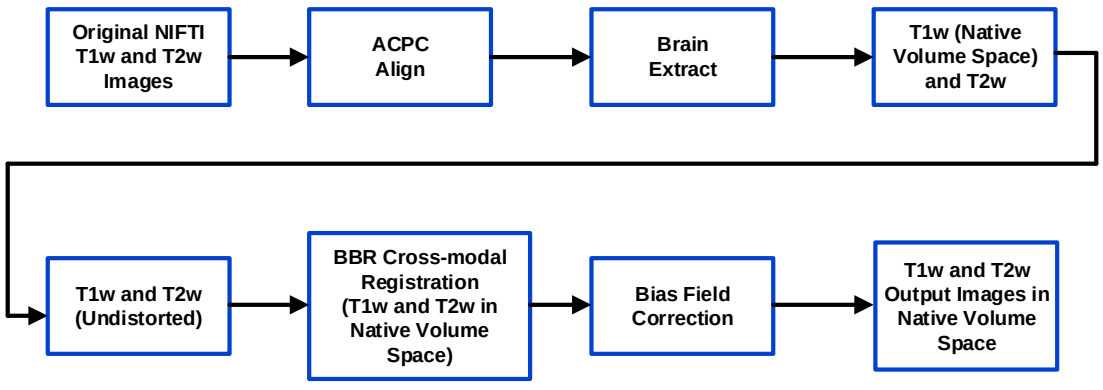

Figure 3.2. Pre-FreeSurfer Pipeline Block Diagram [89]

In pre-freesurfer pipeline, original T1w and T2w images are first aligned with respect to the Anterior Commissure (AC) and the Posterior Commissure (PC) to get the images roughly in the same orientation. After this step, brain extraction is performed, and the skull stripped T1w image transformed to native volume space which is the best approximation of the subject's physical brain with respect to individual's brain size and shape. Undistorted T2w image is cross-modally registered to the T1w image (in native volume space) with boundary based registration [91]. Bias field is obtained by square root of the multiplication of T1w and T2w images in native volume space, and both images are normalized with the smoothed bias field to obtain the final outputs of the pre-freesurfer pipeline.

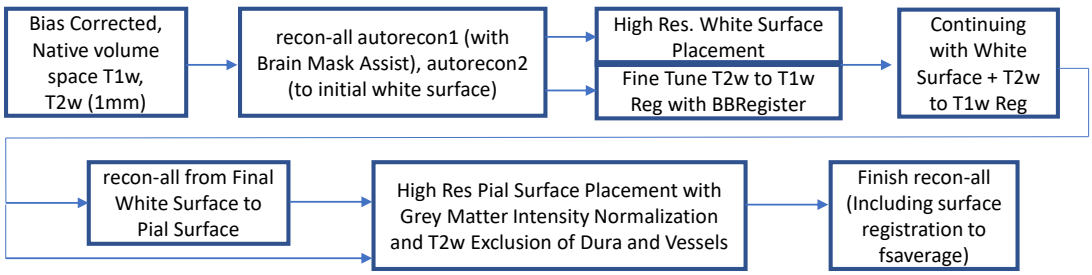

Figure 3.3. FreeSurfer pipeline block diagram [89].

In the freesurfer pipeline (depicted in Figure 3.3), bias corrected T1w and T2w images in native volume space are used as inputs to recon-all process which has several steps consisting of autorecon1, autorecon2, and autorecon3 [92]. The brain mask obtained beforehand with the pre-freesurfer pipeline is used in autorecon1 step, and process continues until autorecon2 (white surface border determination). White surface border is found via T1w image, and T2w is also registered to T1w via BBRegister algorithm of FreeSurfer. T1w/T2w outputs will later be used as inputs for Tortoise program. Later, recon-all process runs until pial surface detection step. Pial surfaces are obtained via improved algorithm that uses high resolution T1w and T2w, and dura and vessels are removed via T2w image. Finally, recon-all process continues until the end.

In Tortoise program, the red box in Figure 3.1, mainly two steps are used for structural data processing called DIFF PREP and DIFF CALC softwares which are used for image resampling, motion distortion correction, re-orientation to a common space, and error analysis. The resolutions of the final output images after this process become 1.5 mm. for DWI, T1w and T2w. In addition, chosen options of Tortoise settings are added in the Appendix, Table B.1.

FSL program is used to obtain registration parameters by using 1mm T1w (output of Freesurfer) and 1.5 mm T1w (output of Tortoise). Found parameters are also applied to T1w with skull stripped (T1b), and Destrieux atlas parcellation via python script that can be found in Appendix B.2. The final version of T1b then used for white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) segmentation, see Appendix B.3. T1b also utilized for extracting brain in T2, shown in Figure 3.1 as T2w with skull stripped (T2b), and in DWI data (output of Tortoise), check Appendix B.4. In addition, T1b is necessary for registration of fMRI data which is the last step of fMRI data analysis.

The data analysis of fMRI is also made by using FSL program, the upper part of Figure 3.1. First 5 volumes are extracted to prevent the effect of unstable volumes. Later on, motion correction is applied to fix the spatial space. Slice timing correction is made for temporal space, since the slices of fMRI volume are acquired at different times. Later on, brain is extracted from skull, and spatial smoothing is applied to reduce noise while keeping valid activation. Normalizing each fMRI volume is made such that each of them have the same mean value. In order to remove low frequency artifacts, high pass temporal filtering is applied, and in the final step, fMRI data is registered to skull stripped T1b data by using boundary based registration. Settings of fMRI data preprocessing are also added to Appendix Table B.2 as well as the corresponding script, see Appendix B.5.

3.1.2. Connectome Node Definitions

One of the steps to construct brain connectomes is the definition of network nodes, that is to say, a set of brain areas are used to divide brain into regions to research its function and structure [93]. Using a voxel-based representation is less efficient, and very noisy which makes network analysis more difficult [94]. Hence, parcellation of the brain into set of distinct regions is considered as an alternative approach which can be either based on cytoarchitecture, structural or functional information [95].

Brain has a folded appearance which consists of gyri and sulci (singular: gyrus and sulcus) which are shaped like hills and valleys, respectively. Precise locations of them is important to define regions, and it requires anatomical expertise as well as time consuming unless automated. Destrieux et al. proposed fully automatic labeling which divides the brain into 74 regions per hemisphere by using surface-based parcellation [96] based on sulco-gyral pattern of the brain [97]. In this study, we have used Destrieux Atlas for node definitions of brain connectomes. In addition, Yeo et al. proposed coarse parcellation by using clustering approach to define functionally distinct cortical regions via rs-fMRI of 1000 healthy subjects [98]. They have found that 7 regions/networks which are also in consistent with the other resting state literature, namely the visual, the somatomotor and auditory, the dorsal attention, the salience, the limbic, the frontoparietal, and the default mode networks. Names of anatomical regions of Destrieux atlas, their indices, and corresponding 7 functional sub-networks defined by Yeo are given in the Appendix, see Table B.3.

3.1.3. Structural Networks

Structural components such as cell membranes, macromolecules, and fibers act as obstacles in brain, and they yield water molecules to diffuse anisotropically. DWI gives information about diffusivity of water molecules by applying magnetic field along different gradient directions. The anisotropic behavior of water molecules can be modelled via anisotropic three dimensional (3D) Gaussian distribution which can be character-

ized by its covariance matrix, also known as diffusion tensor (DT). DT is a symmetric 3x3 real matrix with six degree of freedoms. Therefore, DWI images obtained at least 6 different gradient directions, and the reference image gathered without applying any magnetic field are necessary (at least) in order to obtain the DT [99]. It can be calculated by using the following Stejskal-Tanner equation [100]:

$$S_k = S_0 e^{-b g_k^T D g_k} \tag{3.1}$$

where S_0 is the reference image intensity at the voxel when $b = 0$, and S_k represents the image intensity at the voxel when magnetic field is applied in the direction of g_k, D is the diffusion tensor, k refers to the observation number, and b is the b-value also known as b-factor that characterizes the gradient pulses [101]. In order to find DT, a system of linear equations should be solved, see [102] for details.

DT matrix can be represented by its eigenvalues and corresponding eigenvectors, therefore eigenvalue decomposition is applied to DT in order to obtain the statistical information about diffusion in every voxel such that mean diffusivity, degree of anisotropy, and fiber orientation [103]. Degree of anisotropy can be measured with the following formula:

$$FA = \frac{\sqrt{3\left((\tilde{\lambda}_1 - \hat{\lambda})^2 + (\tilde{\lambda}_2 - \hat{\lambda})^2 + (\tilde{\lambda}_3 - \hat{\lambda})^2\right)}}{\sqrt{2\left(\tilde{\lambda}_1^2 + \tilde{\lambda}_2^2 + \tilde{\lambda}_3^2\right)}} \tag{3.2}$$

where $\tilde{\lambda}_{1,2,3}$ represent eigenvalues of DT, and $\hat{\lambda}$ represents the mean of eigenvalues. This measure is also known as fractional anisotropy (FA). In addition, eigenvector which corresponds the maximum eigenvalue is used to obtain fibers in brain via tractography algorithms since it represents the direction of diffusion at each voxel.

The 4^{th}-order Runge Kutta (RK4) deterministic tractography algorithm is used to tract fibers using the diffusion directions at each voxel. Randomly positioned 30

seed points at each voxel in the gray matter and the white matter with FA$\geq$ 0.15 are used to initiate RK4, and each step fibers are propagated with the step size 0.7 mm. until any of the stopping criteria is met such that arriving the edge of the brain, having a direction less than a given threshold degree, and reaching the voxel with FA$<$ 0.15. Maximum curvature degree and minimum fiber length are set to 35° and 20 mm. [104].

Obtained fibers are used to compute the associations between i^{th} parcel and k^{th} fiber by imposing a 3D Gaussian sphere to the end points of fibers which can be formulated as :

$$W_{ik} = \sum_{\substack{r_j \in P_i \\ |r_j - e_k| \leq 2\sigma}} \frac{1}{(2\pi)^{3/2}\sigma^3} \exp(\frac{-|r_j - e_k|^2}{2\sigma^2}) \tag{3.3}$$

where e_k represents the end point of the k^{th} fiber, r_j is the j^{th} voxel position inside i^{th} parcel P_i. The optimum value of σ has been found 0.155 for 30 seeds per voxel and is determined by using the kernel density estimation (KDE) methodology described in [105]. To construct sNETs, 6 different edge definitions related on fiber counts and FA are used. Edge weights between parcels/nodes P_i and P_j are defined as:

$$e^w_{ij} = \sum_k W_{ik}W_{jk} \tag{3.4}$$

$$e^n_{ij} = \frac{2}{V_i + V_j} \sum_k W_{ik}W_{jk} \tag{3.5}$$

$$e^l_{ij} = \frac{2}{V_i + V_j} \sum_k \left[W_{ik}W_{jk} \min_s(FA(f_k(s)))\right] \tag{3.6}$$

$$e^a_{ij} = \frac{2}{V_i + V_j} \sum_k \left[W_{ik}W_{jk} \operatorname{avg}_s(FA(f_k(s)))\right] \tag{3.7}$$

$$e^m_{ij} = \frac{2}{V_i + V_j} \sum_k \left[W_{ik}W_{jk} \operatorname{median}_s(FA(f_k(s)))\right] \tag{3.8}$$

$$e^h_{ij} = \frac{2}{V_i + V_j} \sum_k \left[W_{ik}W_{jk} \max_s(FA(f_k(s)))\right] \tag{3.9}$$

where V_i and V_j represent volumes in mm^3, f_k is k^{th} fiber, $\min_s()$, $\text{avg}_s()$, $\text{median}_s()$ and $\max_s()$ are minimum, average, median and max operators over fibers' FA values arbitrarily parameterized by s. Superscripts of edge weights are chosen to briefly express its formulation such that w represents weighted, n stands for normalized weighted (fiber counts related), and l, a, m, h are lowest, average, median, and highest (FA-based).

3.1.4. Functional Networks

Regions in the brain are defined in the Section 3.1.2, and these regions should be represented by a signal to obtain relationships between them. However, rs-fMRI data consists of the BOLD signals at each voxel and this information is needed to be used to obtain region-specific signals, also called summary signals. Three different methodologies have been used in this thesis to calculate summary signals of every region: averaging, singular value decomposition (SVD) and SVD with region growing (SVD-RG).

In averaging approach, as the name suggests, a summary signal of a parcel can be found by averaging the BOLD signals of every voxel in the parcel. In SVD approach, a matrix is constructed first such that columns represent voxels in a parcel and rows specify time, and then SVD is applied. The first left singular vector corresponded to highest singular value is used as summary signal since it is the vector that spans the temporal space mostly. The third approach, SVD-RG, is based on the frequency analysis in a parcel such that SVD is applied to the power spectrum of bold signals to obtain the dominant frequency pattern which is the first left singular vector corresponded to maximum singular value. In order to find a seed point for the region growing algorithm, the voxel is chosen that shows the same frequency characteristic with the dominant frequency vector by calculating the difference using l^2-norm. The new region is grown starting from the seed voxel until reaching the predetermined voxel number such as 150 by using the distances of the voxels to the seed voxel on the condition that each voxel should fall within the same region [106]. Later on, SVD is applied to BOLD signals of the new region same as the SVD approach to obtain the summary signal of a parcel.

Note that the used BOLD signals' means are zero.

Summary signals of every parcel in the brain are used to find the associations between parcels. Three different measurements are defined to obtain edge weights of fNETs between parcels/nodes P_i and P_j as:

$$PCC(u) = \frac{\sum_{t=1}^{T-u}[y_i(t+u)-\bar{y_i}][y_j(t)-\bar{y_j}]}{\sigma_i\sigma_j}, \quad u \in [-T,T] \tag{3.10}$$

$$e_{ij}^{Pe} = PCC(0) \tag{3.11}$$

$$e_{ij}^{Mc} = PCC\left(\underset{u\in[-T,T]}{argmax}\,|PCC(u)|\right) \tag{3.12}$$

$$e_{ij}^{Pc} = -\frac{(\Sigma^{-1})_{ij}}{\sqrt{(\Sigma^{-1})_{ii}(\Sigma^{-1})_{jj}}} \tag{3.13}$$

where $y_i(t)$ represents the summary signal at the time point t of the parcel P_i, $\bar{y_i}$ is the mean of the summary signals in parcel P_i, σ_i is the standard deviation of summary signal belonging parcel P_i, u is a shift parameter, (Σ^{-1}) is the inverse covariance matrix, also known as precision matrix, and its i^{th} row and j^{th} column is represented as $(\Sigma^{-1})_{ij}$.

3.2. B-Tensor Factorization

Basically, all networks have two different components which are composed of the nodes in a system and the relationship between these nodes. In this thesis, these nodes (elements of the system) are the brain parcels which divide the brain in spatial domain into different set of regions, and the relationship is connectivity weight which is either derived by using fMRI or dMRI, in case of fNETs and sNETs, respectively.

A complete population of uni-modal (fNET or sNET) or multi-modal (fNET and sNET) brain connectomes can be represented in a 3D or four dimensional (4D) multidimensional array, respectively, that is termed as a tensor. Throughout this thesis, I will call these 3D and 4D tensor representations of population-wide brain connectomes

as "B-Tensor" and denote it with B.

To make the concepts clear, subscripts are added to the B with respect to the modality sources. If a tensor is uni-modal which contains only one modality, it is described by its modality source such as B_f and B_s represent a population of fNETs and sNETs, respectively. Moreover, if a tensor is multi-modal which contains more than one modality (connectome definitions), it is represented via modality source and a letter "m" indicating multi such as B_{fm}, B_{sm} , and B_{cm} tensors contain a population of different modalities using only fNETs, sNETs, and composed of them. All multi-modal tensors are 4D ($\in \mathbb{R}^{I\times J\times K\times R}$), and first two dimensions which are equal represent node indices, 3^{rd} dimension shows patients, and the last one indicates modalities. Moreover, the 3D uni-model B-Tensor is obtained by dropping the 4th dimension (R) that represents the modality. Hence, the uni-modal B-Tensors have the 3D ($\in \mathbb{R}^{I\times J\times K}$) structure. In general, B_u and B_m terms are used to indicate uni-modal 3D B-Tensors which can be either B_f or B_s, and multi-modal 4D B-Tensors which can be either B_{fm}, or B_{sm}, respectively. Superscripts are only used to define B-Tensors labels e.g. B_u^c, B_{fm}^c.

3.2.1. Projections onto Known Factors (Sub-networks)

Multi-modal 4D B-Tensor, B_m, can be decomposed into two components as follows:

$$B_m \approx \sum_{q=1}^{Q} M_{k,r,q} G_{i,j,q} \tag{3.14}$$

where G represents the known basis networks invariant to modality. In addition, M represents the projection of each subject's (k) specific connectome modality (r) onto a given (q) basis network, and Q is the number of known network components. M is the compact & multi-modal representation of brain networks.

Decomposing B_m into its factor matrices can be formulated as following:

$$B_m \approx \sum_{q=1}^{Q} d_q A_{k,q} F_{i,q} D_{j,q} E_{r,q} \ , \text{ where } d_q = 1 \ \ \forall q = 1, \cdots, Q \tag{3.15}$$

where $B_m \in \mathbb{R}^{I \times J \times K \times R}$, $A \in \mathbb{R}^{K \times Q}$, $F \in \mathbb{R}^{I \times Q}$, $D \in \mathbb{R}^{J \times Q}$, $E \in \mathbb{R}^{R \times Q}$ and $d \in \mathbb{R}^{Q}$. Assuming that $\Lambda \in \mathbb{R}^{Q \times Q \times Q \times Q}$ is a tensor with ones(d_qs) along the superdiagonal, this decomposition can be expressed via n-mode product as written in Appendix A.3:

$$B_m \approx \Lambda \times_1 F \times_2 D \times_3 A \times_4 E \tag{3.16}$$

Equation 3.15 is an example of the Canonical Polyadic (CP) decomposition, see Appendix C. There are several ways of implementing CP decomposition such as the alternating least squares (ALS), the gradient descent(GD), and the mini-batch GD (MB-GD). In order to decompose B_m tensor into the M and G factors, low-level factor matrices such as F, D, A, E are found via CP decomposition algorithms mentioned above, and the low-level factor matrices are combined to obtain M and G tensors. In all algorithms, the factor matrices are initialized at first.

Initialization is made in two ways: by assigning random numbers to factor matrices or by using the generalized eigenvalue decomposition (GEVD) [107], more can be found in [108]. The GEVD algorithm utilizes tucker decomposition to reduce the size of the tensor to $Q \times Q \times 2$ dimensions by using the core tensor. These two matrices are used to estimate first two factor matrices by using the generalized Schur decomposition, also known as the QZ algorithm, and other factor matrices can be found by using calculated factor matrices with a least square approach.

Factor matrices when G is not known can be found by applying above-mentioned algorithms which are described in the Appendix C.1, C.2 and C.3 for ALS, GD and MB-GD, respectively. However, when G is known, algorithms should be changed with respect to a fixed G tensor. The relationships between low-level factor matrices and corresponding M and G tensors are described in Proposition 3.1. Moreover, by using

these relations, how fixed G is embedded in the algorithms is discussed below.

Proposition 3.1. *Let* $B_m \in \mathbb{R}^{I\times J\times K\times R}$,$A \in \mathbb{R}^{K\times Q}$, $F \in \mathbb{R}^{I\times Q}$, $D \in \mathbb{R}^{J\times Q}$, $E \in \mathbb{R}^{R\times Q}$, *and* $O \in \mathbb{R}^{Q\times Q}$ *identity matrix. The relationships between* B_m *and* M, G *can be found as:*

(i) $B_m = \Lambda \times_1 F \times_2 D \times_3 A \times_4 E$,
where $\Lambda \in \mathbb{R}^{Q\times Q\times Q\times Q}$ *with ones along the superdiagonal.*

(ii) $M_{k,r,q} = \sum_{s=1}^{S} A_{k,s}E_{r,s}\sqrt{\Lambda_{q,s}}$, *and* $G_{i,j,q} = \sum_{s=1}^{S} F_{i,q}D_{j,q}\sqrt{\Lambda_{q,s}}$ *where* $\Lambda_{q,s} = \begin{cases} 0 & q \neq s \\ 1 & q = s \end{cases}$

(iii) $\begin{aligned} M &= \Lambda \times_1 A \times_2 E \times_3 O \\ G &= \Lambda \times_1 F \times_2 D \times_3 O \end{aligned}$ *where* $\Lambda \in \mathbb{R}^{Q\times Q\times Q}$ *with ones along the superdiagonal.*

(iv) $\begin{aligned} & M = \Lambda \times_1 A \times_2 E \times_3 O && && G = \Lambda \times_1 F \times_2 D \times_3 O \\ & M = O\Lambda_{(3)}(E \otimes A)^\top && \textit{and} && G = O\Lambda_{(3)}(D \otimes F)^\top \\ & M_{(3)} = (E \odot A)^\top && && G_{(3)} = (D \odot F)^\top \end{aligned}$

(v) $\begin{aligned} & B_m = \Lambda \times_1 F \times_2 D \times_3 A \times_4 E \\ & B_{m_{(3,4\times 1,2)}} = (E \otimes A)\Lambda_{qq\times qq}(D \otimes F)^\top \\ & B_{m_{(3,4\times 1,2)}} = (E \odot A)(D \odot F)^\top \\ & B_{m_{(3,4\times 1,2)}} = {M_{(3)}}^\top G_{(3)} \end{aligned}$

Note that one can get a solution via matrix decomposition as a result of this formula :

$$B_{m_{(3,4\times 1,2)}} = {M_{(3)}}^\top G_{(3)} \tag{3.17}$$

In the Equation 3.17, B_m and G are known, and M is unknown. Therefore, the problem of finding M is a least square minimization problem which can be solved by finding the inverse of $G_{(3)}$ which may not be possible all the time. Therefore, the QR decomposition is applied to decompose $G_{(3)} = \hat{Q}\hat{R}$ where $\hat{Q}$ is an orthogonal matrix

($\hat{Q}^\top \hat{Q} = I$) and $\hat{R}$ is an upper triangular matrix. In this regard, M can be found as:

$$G_{(3)} = \hat{Q}\hat{R} \tag{3.18}$$

$$G_{(3)}^{-1} = \hat{R}^{-1}\hat{Q}^{-1} = \hat{R}^{-1}\hat{Q}^\top \tag{3.19}$$

$$M_{(3)}^\top = B_{m_{(3,4\times 1,2)}}\hat{R}^{-1}\hat{Q}^\top \tag{3.20}$$

The idea behind the ALS is that each factor matrix is solved in a turn, when the other factor matrices are fixed. Although F and D are unknown, we have the knowledge of G. Therefore, in each iteration finding A and E should be enough. By using the n-mode matricization explained in Appendix A.2 and Proposition A.3, following equations can be written with respect to Equation 3.16:

$$B_{m_{(3)}} = A\Lambda_{(3)}(E \otimes D \otimes F)^T \quad \text{and} \quad B_{m_{(4)}} = E\Lambda_{(4)}(D \otimes F \otimes A)^T$$

Because of the structure of Λ (e.i., being superdiagonal with ones), Kronecker products of the factor matrices turn into Khatri-Rao product as shown in the next formula:

$$B_{m_{(3)}} = A(E \odot D \odot F)^T \quad \text{and} \quad B_{m_{(4)}} = E(D \odot F \odot A)^T$$

$G_{(3)}^\top$ is found in Proposition 3.1 as $G_{(3)}^\top = (D \odot F)$. Therefore, if $D \odot F$ is substituted with $G_{(3)}^\top$, it can be considered as the first factor matrix in the decomposition. Pseudocode of the ALS algorithm via embedded G is given in Figure 3.4.

As it has been done in ALS algorithm, we can assume $G_{(3)}^\top$ as the first factor matrix. Finding the gradient of B_m without factorizing G can be formulated as follows:

$$L(F, D, A, E) = ||B_{m_{(3)}} - A(E \odot D \odot F)^\top||_F^2 \tag{3.21}$$

$$\frac{\partial L(F, D, A, E)}{\partial A} = -2\left(B_{m_{(3)}} - A(E \odot D \odot F)^\top\right)(E \odot D \odot F) \tag{3.22}$$

INPUT: $B_m \in \mathbb{R}^{I\times J\times K\times R}, G \in \mathbb{R}^{I\times J\times Q}$

OUTPUT: $A \in \mathbb{R}^{K\times Q}, E \in \mathbb{R}^{R\times Q}$

initialize $A \in \mathbb{R}^{K\times Q}$ and $E \in \mathbb{R}^{R\times Q}$

repeat

$V \leftarrow (E^\top E) * (G_{(3)}G_{(3)}^\top)$

$A \leftarrow B_{m_{(3)}}\Big(E \odot G_{(3)}^\top\Big)V^\dagger$

$V \leftarrow (G_{(3)}G_{(3)}^\top) * (A^\top A)$

$E \leftarrow B_{m_{(3)}}\Big(G_{(3)}^\top \odot A\Big)V^\dagger$

until some convergence criteria is satisfied

Figure 3.4. ALS with fixed G algorithm.

Furthermore, if $(D \odot F)$ is substituted with $G_{(3)}^\top$ as

$$\frac{\partial L(F,D,A,E)}{\partial A} = -2\left(B_{m_{(3)}} - A(E \odot G_{(3)}^\top)^\top\right)(E \odot G_{(3)}^\top) \tag{3.23}$$

$$\frac{\partial L(F,D,A,E)}{\partial E} = -2\left(B_{m_{(4)}} - E(A \odot G_{(3)}^\top)^\top\right)(A \odot G_{(3)}^\top) \tag{3.24}$$

The GD algorithm (see Figure 3.5) demonstrates how the integration of tensor G changes the procedure.

INPUT: $B_m \in \mathbb{R}^{K\times I\times J\times R}, G \in \mathbb{R}^{I\times J\times Q}, \eta$

OUTPUT: $A \in \mathbb{R}^{K\times Q}, E \in \mathbb{R}^{R\times Q}$

initialize $A \in \mathbb{R}^{K\times Q}$ and $E \in \mathbb{R}^{R\times Q}$

repeat

$Grad \leftarrow \left(B_{m_{(3)}} - A(E \odot G_{(3)}^\top)^\top\right)(E \odot G_{(3)}^\top)$

$A \leftarrow A + \eta Grad$

$Grad \leftarrow \left(B_{m_{(4)}} - E(A \odot G_{(3)}^\top)^\top\right)(A \odot G_{(3)}^\top)$

$E \leftarrow E + \eta Grad$

until some convergence criteria is satisfied

Figure 3.5. GD with fixed G algorithm.

The idea behind the stochastic gradient descent (SGD) is that randomly selected indices of factor matrices are updated at each iteration (one randomized point at a time) which forces us to reach the value of F and D at a randomized position. The MB-GD algorithm calculates the gradient using neither the whole data nor a single data, but instead finds the gradient over a batch. Hence, a predefined batch size is enough to calculate the gradient. In every iteration, gradients are calculated for all randomly ordered batches. In addition, without the knowledge of F and D, it is still possible to embed G in the following algorithm, see Figure 3.6. Rows of F and D which will be used in the MB-GD algorithm are needed to be known. Although the row values of F and D are unknown, G can be applied to the formula since $G_{i,j,:}$ is the hadamard product of i^{th} row of F and j^{th} row of D.

INPUT: $B_m \in \mathbb{R}^{I\times J\times K\times R}, G \in \mathbb{R}^{I\times J\times Q},\ batch_{size} > 0$, M is the elements of B

OUTPUT: $A \in \mathbb{R}^{K\times Q}, E \in \mathbb{R}^{R\times Q}$

initialize $A \in \mathbb{R}^{K\times Q}$ and $E \in \mathbb{R}^{R\times Q}$
B_{ind} = shuffle.index(B_m)
$Batch = getbatches(B_{ind}, batch_{size})$
repeat
  Batch = random.shuffle(Batch)
  for $n = 1, \cdots, M/batch_{size}$ **do**
    $(\hat{B}, \hat{A}, \hat{E}, \hat{G}) = getbatch_n(B_m, A, E, G_{(3)}^{\top})$
    $Grad \leftarrow \left(\hat{B}_{(3)} - \hat{A}(\hat{E} \odot \hat{G}_{(3)}^{\top})^{\top}\right)(\hat{E} \odot \hat{G}_{(3)}^{\top})$
    $A \leftarrow A + \eta Grad$
    $(\hat{A}) = getbatch_n(A)$
    $Grad \leftarrow \left(\hat{B}_{(4)} - \hat{E}(\hat{A} \odot \hat{G}_{(3)}^{\top})^{\top}\right)(\hat{A} \odot \hat{G}_{(3)}^{\top})$
    $E \leftarrow E + \eta Grad$
  end for
until some convergence criteria is satisfied

Figure 3.6. MB-GD with fixed G algorithm.

3.2.2. Uni-Modal B-Tensor Factorization

In this section, B_u, the uni-model 3D B-Tensor decomposed into different components to find low dimensional connectome representations and orthogonal basis sub-networks as they have been obtained by M and G in Section 3.2.1, respectively. However, in this model, basis sub-networks are not known. This optimization problem can be written as:

$$\begin{aligned} &\min_{d_q, v_q, u_q^c} \left\{ \sum_{c=1}^{C} \frac{1}{N_c} \left\| B_u^c - \sum_{q=1}^{Q} d_q (v_q \circ v_q \circ u_q^c) \right\|_2^2 \right\} \\ &\text{s.t.} \ \sum_{c=1}^{C} ||u_q^c||_2^2 = 1; \ v_q^\top v_j = \delta_{qj} \end{aligned} \tag{3.25}$$

where C is the number of classes in the dataset, N_c is the number of patients in class c, Q represents the number of orthogonal basis sub-networks, $d_q \in \mathbb{R}$ is the q^{th} sub-network's scaling factor, $v_q \circ v_q$ indicates the q^{th} basis sub-network where $v_q \in \mathbb{R}^I$ (note that I is equal to J), and $u_q^c \in \mathbb{R}^{N_c}$ is the corresponding projection coefficient. This optimization model handles the class imbalance problem by applying an appropriate normalization, see [73] for equal size classes.

This model is a CP decomposition, and the minimization problem can also be written as maximization [109]:

$$\begin{aligned} &\max_{v_q, u_q^c} \sum_{c=1}^{C} \frac{1}{N_c} \left(B_u^c \times_1 (\Gamma_{q-1} v_q) \times_2 (\Gamma_{q-1} v_q) \times_3 u_q^c \right) \\ &\text{s.t.} \ \sum_{c=1}^{C} ||u_q^c||_2^2 = 1; \ v_q^\top v_j = \delta_{jq} \end{aligned} \tag{3.26}$$

where $\Gamma_{q-1} = P_{q-1} P_{q-1}^\top$ is the projection matrix with $P_{q-1} = [v_1, \cdots, v_{q-1}]$ imposing the Gram-Schmidt orthogonality process.

In order to provide parsimony for the notation, subscripts of v_q and u_q^c are

dropped. The optimization problem stated in Equation 3.26 can be solved using the Lagrange multipliers as:

$$L(v, u^c) = \sum_{c=1}^{C} \frac{1}{N_c} B_u^c \times_1 v \times_2 v \times_3 u^c - \lambda_1 (\sum_{c=1}^{C} (u^c)^\top (u^c) - 1) - \lambda_2 (v^\top v - 1) \quad (3.27)$$

Taking the partial derivative of L w.r.t. u^c,

$$\frac{\partial L}{\partial u^c} = \frac{1}{N_c} (B_u^c \times_1 v \times_2 v) - 2\lambda_1 (u^c)^\top = 0 \quad (3.28)$$

$$\Rightarrow u^c = \frac{(B_u^c \times_1 v \times_2 v)^\top}{N_c 2\lambda_1} \quad (3.29)$$

$$\lambda_1 \left(\sum_{c=1}^{C} (u^c)^T (u^c) - 1 \right) = 0 \Rightarrow \sum_{c=1}^{C} ||u^c||_2^2 = 1$$

$$\Rightarrow \sum_{c=1}^{C} \left|\left| \frac{(B_u^c \times_1 v \times_2 v)^\top}{N_c 2\lambda_1} \right|\right|_2^2 = 1$$

$$\Rightarrow \lambda_1 = \frac{1}{2} \sqrt{\sum_{c=1}^{C} \left|\left| \frac{(B_u^c \times_1 v \times_2 v)^\top}{N_c} \right|\right|_2^2} \quad (3.30)$$

Equations 3.29 and 3.30 give,

$$u^c = \frac{(B_u^c \times_1 v \times_2 v)^\top}{N_c \sqrt{\sum_{c=1}^{C} \left|\left| \frac{(B_u^c \times_1 v \times_2 v)^\top}{N_c} \right|\right|_2^2}} \quad (3.31)$$

Taking the partial derivative of Equation 3.27 w.r.t. v,

$$\frac{\partial L}{\partial v} = 2 \sum_{c=1}^{C} \frac{1}{N_c} v^\top (B_u^c \times_3 u^c) - 2\lambda_2 v^\top = 0 \quad (3.32)$$

$$\Rightarrow \sum_{c=1}^{C} \frac{1}{N_c} (B_u^c \times_3 u^c)^\top v = \lambda_2 v \quad (3.33)$$

This formulation is an eigenvalue decomposition where λ_2 is the eigenvalue of corresponding eigenvector v. Since Equation 3.26 can be written without the orthogonality

constraint as $\max_{v,u^c} \sum_{c=1}^{C} \frac{1}{N_c} v^\top (B_u^c \times_3 u^c)^\top v$; eigenvector, corresponding with the maximum eigenvalue, maximizes the Equation 3.26. Finally, v can be found as:

$$v = E_{max} \left(\sum_{c=1}^{C} \frac{1}{N_c} (B_u^c \times_3 u^c)^\top \right) \tag{3.34}$$

where $E_{max}(.)$ represents the eigenvector of the maximum eigenvalue. The algorithm can be found in Figure 3.7.

3.2.3. Multi-Modal B-Tensor Factorization

In this section, B_{cm}, the multi-model 4D B-Tensor decomposed into different components to find low dimensional connectome representations and orthogonal basis sub-networks. Unlike Section 3.2.2 and Section 3.2.1, functional and structural networks can be combined. This optimization problem can be written as:

$$\begin{aligned} &\min_{d_q, v_q, u_q^c, w_q} \left\{ \sum_{c=1}^{C} \frac{1}{N_c} \left\| B_{cm}^c - \sum_{q=1}^{Q} d_q (v_q \circ v_q \circ u_q^c \circ w_q) \right\|_2^2 \right\} \\ &\text{s.t.} \ \sum_{c=1}^{C} ||u_q^c||_2^2 = 1; \ v_q^T v_j = \delta_{jq}; \ w_q^T w_q \leq 1 \end{aligned} \tag{3.35}$$

where C is the number of classes in the dataset, N_c is the number of patients in class c, Q represents the number of orthogonal basis sub-networks, $d_q \in \mathbb{R}$ is the q^{th} sub-network's scaling factor, $v_q \circ v_q$ indicates the q^{th} basis sub-network where $v_q \in \mathbb{R}^I$ (note that I is equal to J), $u_q^c \in \mathbb{R}^{N_c}$ is the corresponding projection coefficient, w_q is the modality scaling parameter which balances the functional-structural sparsity. We have set

$$w_q = [w_q^*; \alpha w_q^*] = \Theta w_q^*$$

where α is a fixed density factor, defined as $\alpha = density(\text{fNET})/density(\text{sNET})$ to employ parameter tying.

INPUT: $B_u^c \ \forall c \in \{1, \cdots, C\}$

OUTPUT: $U_u = [u_:^1 u_:^2 \cdots u_:^C]$

$$V_{q,u} = v_q \circ v_q, \ \forall q \in \{1, \cdots, Q\}$$

Let $\hat{B}_u^c = B_u^c \ \forall c \in \{1, \cdots, C\}$

for q=1, $\cdots$ Q **do**

Initialize : v_q and $u_q^c \ \forall c \in \{1, \cdots, C\}$, $iter = 1$

if q=1 **then**

$\Gamma_0 = I \in \mathbb{R}^{P \times P}$ identity matrix

else

$\Gamma_{q-1} = I - \sum_{s=1}^{q-1} v_s \circ v_s$

end if

$Obj(1) = \sum_{c=1}^{C} \frac{1}{N_c} \left(\hat{B}_u^c \times_1 (\Gamma_{q-1} v_q) \times_2 (\Gamma_{q-1} v_q) \times_3 u_q^c \right)$

while $err > 10^{-6}$ and $iter < 1000$ **do**

$$u_q^c = \frac{\hat{B}_u^c \times_1 v_q \times_2 v_q}{N_c \sqrt{\sum_{c=1}^{C} \left\| \frac{\hat{B}_u^c \times_1 v_q \times_2 v_q}{N_c} \right\|_2^2}}$$

$v_q = E_{max}(\Gamma_{q-1}(\sum_{c=1}^{C} \frac{1}{N_c} \hat{B}_u^c \times_3 u_q^c)\Gamma_{q-1})$

CP Scaling : $d_q = \sum_{c=1}^{C} \hat{B}_u^c \times_1 v_q \times_2 v_q \times_3 u_q^c$

$iter = iter + 1$

$Obj(iter) = \sum_{c=1}^{C} \frac{1}{N_c} \left(\hat{B}_u^c \times_1 (\Gamma_{q-1} v_q) \times_2 (\Gamma_{q-1} v_q) \times_3 u_q^c \right)$

$err = |\left(Obj(iter) - Obj(iter-1)\right) / Obj(1)|$

end while

Deflation : $\hat{B}_u^c = \hat{B}_u^c - d_q(v_q \circ v_q \circ u_q^c)$

end for

Figure 3.7. 3^{rd} order uni-modal B-Tensor factorization algorithm [74].

This model is a CP decomposition, and the Equation 3.35 can also be written as maximization [109]:

$$\begin{aligned} &\max_{v_q, u_q^c, w_q} \sum_{c=1}^{C} \frac{1}{N_c} \left(B_{cm}^c \times_1 (\Gamma_{q-1} v_q) \times_2 (\Gamma_{q-1} v_q) \times_3 u_q^c \times_4 w_q \right) \\ &\text{s.t. } \sum_{c=1}^{C} ||u_q^c||_2^2 = 1;\ v_q^\top v_j = \delta_{jq};\ w_q^\top w_q \leq 1 \end{aligned} \tag{3.36}$$

where $\Gamma_{q-1} = P_{q-1} P_{q-1}^\top$ is the projection matrix with $P_{q-1} = [v_1, \cdots, v_{q-1}]$ imposing the Gram-Schmidt orthogonality process. In order to keep the notation simple, subscripts of w_q, v_q, and u_q^c are dropped. The optimization problem stated in Equation 3.36 can be solved by using the Lagrange multipliers as:

$$\begin{aligned} L(v, u^c, w) &= \sum_{c=1}^{C} \frac{1}{N_c} B_{cm}^c \times_1 v \times_2 v \times_3 u^c \times_4 w \\ &- \lambda_1 (\sum_{c=1}^{C} (u^c)^\top (u^c) - 1) - \lambda_2 (v^\top v - 1) - \lambda_3 (w^\top w - 1) \end{aligned} \tag{3.37}$$

Taking the partial derivative of L w.r.t. u^c,

$$\frac{\partial L}{\partial u^c} = \frac{1}{N_c} \left((B_{cm}^c \times_1 v \times_2 v) w \right)^\top - 2\lambda_1 (u^c)^\top = 0 \tag{3.38}$$

$$\Rightarrow u^c = \frac{(B_{cm}^c \times_1 v \times_2 v) w}{N_c 2\lambda_1} \tag{3.39}$$

$$\lambda_1 \left(\sum_{c=1}^{C} (u^c)^\top (u^c) - 1 \right) = 0 \Rightarrow \sum_{c=1}^{C} ||u^c||_2^2 = 1$$

$$\Rightarrow \sum_{c=1}^{C} \left|\left| \frac{(B_{cm}^c \times_1 v \times_2 v) w}{N_c 2\lambda_1} \right|\right|_2^2 = 1$$

$$\Rightarrow \lambda_1 = \frac{1}{2} \sqrt{\sum_{c=1}^{C} \left|\left| \frac{(B_{cm}^c \times_1 v \times_2 v) w}{N_c} \right|\right|_2^2} \tag{3.40}$$

The Equations 3.39 and 3.40 give,

$$u^c = \frac{(B^c_{cm} \times_1 v \times_2 v)w}{N_c \sqrt{\sum_{c=1}^{C} \left|\left| \frac{B^c_{cm} \times_1 v \times_2 v)w}{N_c} \right|\right|_2^2}} \tag{3.41}$$

Taking the derivative of Equation 3.37 w.r.t. v;

$$\frac{\partial L}{\partial v} = 2\sum_{c=1}^{C} \frac{1}{N_c} v^\top (B^c_{cm} \times_3 u^c \times_4 w) - 2\lambda_2 v^\top = 0 \tag{3.42}$$

$$\Rightarrow \sum_{c=1}^{C} \frac{1}{N_c} (B^c_{cm} \times_3 u^c \times_4 w)^\top v = \lambda_2 v \tag{3.43}$$

eigenvector, corresponding with the maximum eigenvalue, maximizes the Equation 3.36 since it can be written as $\max_{v,u^c,w} \sum_{c=1}^{C} \frac{1}{N_c} v^\top (B^c_{cm} \times_3 u^c \times_4 w)^T v$ without imposing projection matrix. Finally, v can be found as:

$$v = E_{max} \left(\sum_{c=1}^{C} \frac{1}{N_c} (B^c_{cm} \times_3 u^c \times_4 w)^T \right) \tag{3.44}$$

Finally, letting $w = \begin{bmatrix} 1 \\ \\ \alpha \end{bmatrix} w^* = \Theta w^*$, where α is the density ratio of the two connectome modalities in B^c_{cm}, $w \in \mathbb{R}^{2\times 1}$, and taking the derivative of Equation 3.37 w.r.t. w^*,

$$\frac{\partial L}{\partial w^*} = \sum_{c=1}^{C} \frac{1}{N_c} (u^c)^\top (B^c_{cm} \times_1 v \times_2 v)\Theta - 2\lambda_3 (w^*)^\top \Theta^\top \Theta = 0 \tag{3.45}$$

$$w^* = \frac{\sum_{c=1}^{C} \frac{1}{N_c} \Theta^\top (B^c_{cm} \times_1 v \times_2 v)^\top u^c}{2\lambda_3 \Theta^\top \Theta} \tag{3.46}$$

and λ_3 can be found as :

$$\sum_{c=1}^{C} ||\Theta w^*||_2^2 = \left|\left| \frac{\Theta \sum_{c=1}^{C} \frac{1}{N_c} \Theta^\top (B_{cm}^c \times_1 v \times_2 v)^\top u^c}{2\lambda_3 \Theta^\top \Theta} \right|\right|_2^2 = 1$$
$$\Rightarrow \lambda_3 = \frac{1}{2} \left|\left| \frac{\Theta \sum_{c=1}^{C} \frac{1}{N_c} \Theta^\top (B_{cm}^c \times_1 v \times_2 v)^\top u^c}{\Theta^\top \Theta} \right|\right|_2 \tag{3.47}$$

Equations 3.46 and 3.47 give,

$$w^* = \frac{\sum_{c=1}^{C} \frac{1}{N_c} \Theta^\top (B_{cm}^c \times_1 v \times_2 v)^\top u^c}{\left|\left| \Theta \sum_{c=1}^{C} \frac{1}{N_c} \Theta^T (B_{cm}^c \times_1 v \times_2 v)^\top u^c \right|\right|_2} \tag{3.48}$$

from which we have $w = \begin{bmatrix} 1 \\ \alpha \end{bmatrix} w^*$.

Note that $w \in \mathbb{R}^{2\times 1}$ is two dimensional vector as the simplest case and if there are more than two modalities such as R, $w \in \mathbb{R}^{R\times 1}$, θ can be appended without any change in the formulation for parameter tying.

INPUT: $B^c_{cm}\ \forall c \in \{1, \cdots, C\}, \Theta$

OUTPUT: $U_{cm} = [u^1_{:} u^2_{:} \cdots u^C_{:}]$

$$V_{q,cm} = v_q \circ v_q, \forall q \in \{1, \cdots, Q\}$$

Let $\hat{B}^c_{cm} = B^c_{cm}\ \forall c \in \{1, \cdots, C\}$

for q=1, $\cdots$ Q **do**

Initialize : v_q, w_q and $u^c_q\ \forall c \in \{1, \cdots, C\}$, $iter = 1$

if q=1 **then**

$\mathbf{\Gamma}_0 = I \in \mathbb{R}^{P \times P}$ identity matrix

else

$\mathbf{\Gamma}_{q-1} = I - \sum_{s=1}^{q-1} v_s \circ v_s$

end if

$Obj(1) = \sum_{c=1}^{C} \frac{1}{N_c} \left(\hat{B}^c_{cm} \times_1 (\Gamma_{q-1} v_q) \times_2 (\Gamma_{q-1} v_q) \times_3 u^c_q \times_4 w_q \right)$

while $err > 10^{-6}$ and $iter < 1000$ **do**

$$u^c_q = \frac{(\hat{B}^c_{cm} \times_1 v_q \times_2 v_q) w_q}{N_c \sqrt{\sum_{c=1}^{C} \left\| \frac{(\hat{B}^c_{cm} \times_1 v_q \times_2 v_q) w_q}{N_c} \right\|_2^2}}$$

$$w^*_q = \frac{\left(\sum_{c=1}^{C} \frac{1}{N_c} \Theta^T (\hat{B}^c_{cm} \times_1 v_q \times_2 v_q)^T u^c_q \right)}{\left\| \Theta \sum_{c=1}^{C} \frac{1}{N_c} \Theta^T (\hat{B}^c_{cm} \times_1 v_q \times_2 v_q)^T u^c_q \right\|_2}$$

$w_q = \Theta w^*_q$

$v_q = E_{max}(\Gamma_{q-1} (\sum_{c=1}^{C} \frac{1}{N_c} \hat{B}^c_{cm} \times_3 u^c_q \times_4 w_q) \Gamma_{q-1})$

CP Scaling : $d_q = \sum_{c=1}^{C} \hat{B}^c_{cm} \times_1 v_q \times_2 v_q \times_3 u^c_q \times_4 w_q$

$iter = iter + 1$

$Obj(iter) = \sum_{c=1}^{C} \frac{1}{N_c} \left(\hat{B}^c_{cm} \times_1 (\Gamma_{q-1} v_q) \times_2 (\Gamma_{q-1} v_q) \times_3 u^c_q \times_4 w_q \right)$

$err = |\left(Obj(iter) - Obj(iter-1)\right) / Obj(1)|$

end while

Deflation : $\hat{B}^c_{cm} = \hat{B}^c_{cm} - d_q (v_q \circ v_q \circ u^c_q \circ w_q)$

end for

Figure 3.8. 4^{th} order multi-modal B-Tensor factorization algorithm [74].

4. EXPERIMENTS

4.1. Data

Clinical studies of the participants were conducted at Behavioral Neurology and Movement Disorders Unit, Department of Neurology, Istanbul Faculty of Medicine, Istanbul University. Raw data images are obtained in a single session by using the Philips Achieva 3T MRI system (Best, Netherlands) with a 32-channel head coil at Neuroimaging Unit of Hulusi Behçet Life Sciences Research Laboratory, Istanbul University. 3D Fast Field Echo (FFE) pulse sequence with multi-shot Turbo Field Echo (TFE) imaging mode has been utilized to obtain T1w MRI. Acquisition parameters of the MRI scans were flip-angle = 8°, SENSE reduction 2 (Foot-Head), FOV = $220(RL) \times 240(AP)$ mm^2, voxel size = $1.0 \times 1.0 \times 1.0$ mm^3, TE/TR = 3.8 ms./8.3 ms., and number of slices = 180. DWI data is obtained with a maximum gradient strength of 40 mT/m, and 200 mT/m/ms slew rate, using a single-shot, pulse-gradient spin echo (PGSE), echo planar imaging (EPI) sequence. The acquisition parameters were FOV= 200×236 mm^2, 2.27 mm isotropic voxel size, 112×112 reconstruction matrix, 71 slices, and $TE/TR = 92ms/9032ms$. 120 diffusion weighted gradient directions were used at various b-values between $0 - 3000$ s/mm^2.

National Institute on Aging/Alzheimer's Association (NIA-AA) criteria is used to diagnose participants as MCI [110]. Free and Cued Selective Reminding Test (FCSRT) consists of three sessions, and in each session, participants learn 16 words with corresponding semantic cues. Then, the Total Free Recall score (TFR) is obtained for each session based on how many of those words are retrieved by the participant with and without a cue. Hence, the maximum value of the TFR can be 48, if a participant retrieves all the words. Cue index (CI) is calculated by dividing the total recall with cues to the summation of number of cued recall and unknown words. An objective deficit in episodic memory is defined as participants with a CI value of 0.67 and lower and TFR value of 27 and lower [111]. Hence, the participants whose scores are above

the determined threshold are labeled as SCI. The choice of the cut-off value as 27 is more liberal compared to the 24 which is a highly sensitive score of predicting future dementia [112]. In addition, NIA-AA criteria for probable AD with an amnestic presentation is utilized to diagnose AD patients [113]. Cognitive Functions Instrument Subject form (CFI-S) or CFI Study Partner form (CFI-SP) which are used to detect cognitive decline are filled by the SCI patients [114]. The Clinical Dementia Rating (CDR) scale is used to measure cognitive and functional performance in six domains that are memory, orientation, judgment & problem solving, community affairs, home & hobbies, and personal care [115].The CDR scores of all AD participants are 0.5 or 1 which means that they have very mild, or mild dementia. All MCI and SCI patients have a score of 0.5 and 0, respectively. A panel of behavioral neurologists confirm the diagnosis of patients as SCI, MCI and AD after a comprehensive examination. Criteria of excluding the participants from the study are as follows: history of current/past neurological/psychiatric disorders, major head trauma with loss of consciousness, white matter hyperintensities on MRI with a Fazekas score of 2 and 3 and alcohol or substance abuse. The remaining data consists of 88 patients. This research is conducted with the written approval of all patients and the confirmation of the ethics committee (Istanbul University, Istanbul Faculty of Medicine, Ethics Committee Approval: 877/30.05.2014; Boğaziçi University, Ethics Committee Approval: 2014-1/17.02.201).

4.2. Projections onto Known Factors

In this study, all connectome definitions for sNETs from Section 3.1.3 are used for a total of 20 patients (7 AD, 13 SCI) which provides 4D multi-modal B-Tensor denoted by B_{sm}. In addition, the Destriuex Atlas parcellation provides 148 nodes to construct the connectomes. Hence, $B_{sm} \in \mathbb{R}^{I \times J \times K \times R}$ is acquired by $I = J = 148$ node indices, $K = 20$ patients and $R = 6$ modality numbers. G tensor represents the known canonical sub-networks which is obtained from 7 functional networks defined by Yeo et al. [98]. Each of these networks is constructed by assigning ones to the edges between the nodes that belong to same network, and note that some of the nodes belong more than one network. Therefore, $G \in \mathbb{R}^{I \times J \times Q}$ is defined with $I = J = 148$ node indices

and $Q = 7$ canonical sub-networks which are depicted in Figure using axial, coronal, and sagittal view.

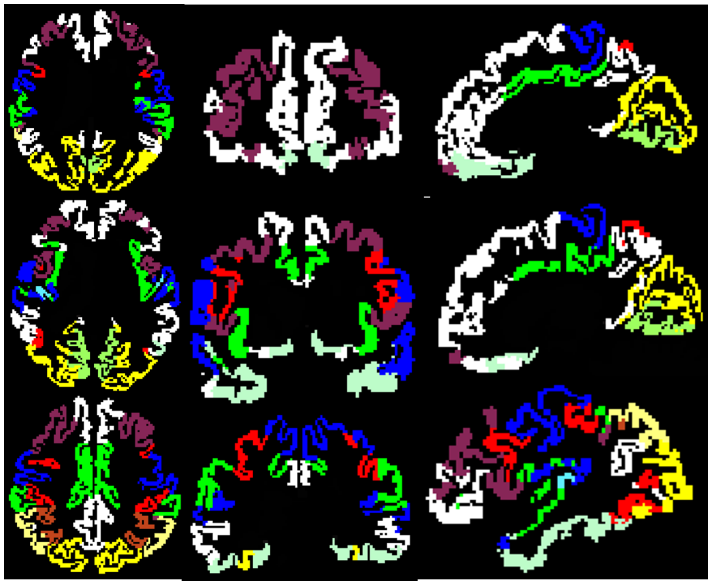

Figure 4.1. The 7 major functional sub-networks are depicted with color codes as: visual (yellow), somatomotor & auditory (blue), dorsal attention (red), salience (green), limbic (mint), frontoparietal (raspberry), and default mode (white) in axial, coronal and sagittal view [98].

$M \in \mathbb{R}^{K \times R \times Q}$ tensor is obtained by the matrix factorization approach described in Section 3.2.1 which is used for the classification of AD and SCI for each one of the 6 connectivity definitions. The classification of patients is assessed by using the Linear Discriminant Analysis (LDA) using 7D representation of every patient for each modality [116]. In addition, unit normal of separating hyperplanes is computed to understand how well each canonical axis is successful at classification of the two groups, see Figure 4.2. Moreover, the global network parameters such as the global clustering coefficient, the average shortest path length and the small worldness index are calculated to acquire the binary classification accuracy. The accuracy results obtained by the B-Tensor model ($Acc\ B_{sm}$) and the global parameters ($Acc\ Glb$) are presented in the Table 4.1 with the corresponding canonical axes for each sNETs definition.

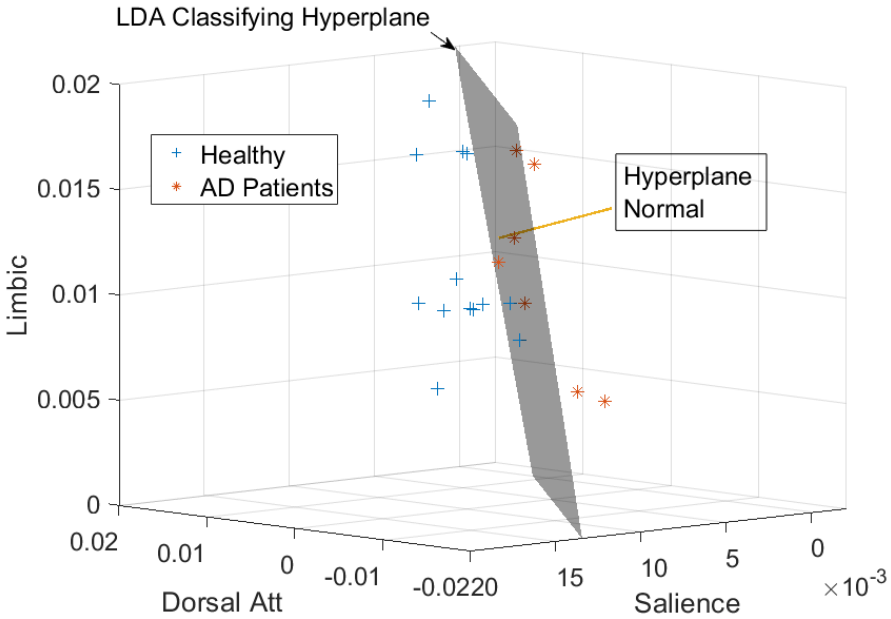

Figure 4.2. A depiction of the classification hyperplane in 3D for demonstration purposes.

Table 4.1. Classification accuracies of the LDA classifier for different modalities and the associated separating hyperplane normals, adopted by [72].

Conn.	Separating Hyperplane Normals							Acc	
	Vis.	Som.	Dors.	Sal.	Limb.	F-Par.	Def. M.	B_{sm}	*Glb*
e_{ij}^{w}	0.012	-0.01	-0.377	-0.887	0.211	-0.159	-0.022	0.95	0.8
e_{ij}^{n}	-0.326	-0.13	-0.53	-0.636	0.426	0.054	-0.073	0.95	0.85
e_{ij}^{l}	-0.224	-0.064	-0.351	-0.509	0.401	0.448	-0.446	0.9	0.75
e_{ij}^{h}	-0.352	0.231	-0.345	-0.703	0.299	0.056	0.34	0.9	0.9
e_{ij}^{a}	-0.352	-0.287	-0.298	-0.714	0.347	0.046	0.215	0.9	0.8
e_{ij}^{m}	-0.333	-0.296	-0.286	-0.741	0.347	0.046	0.215	0.9	0.75
$(\lvert\mu\rvert)$	0.2669	0.1702	0.3648	0.6987	0.3383	0.1367	0.228	-	-
(σ)	0.1336	0.1197	0.0878	0.1243	0.0767	0.1588	0.1605	-	-

All accuracies for B-Tensor model are above 90% and the corresponding classification accuracies with global features are between $75\% - 90\%$. Fiber count based sNET definitions (e^w_{ij}, e^n_{ij}) are the best performing modalities for the B-Tensor model. Each component of the canonical axes corresponds to one of the networks defined by Yeo, and absolute value of these components provides an insight for which network is more important in discriminating these two groups. Therefore, the standard deviations and the means of the magnitude of each component calculated over all connectome definitions are given in the Figure 4.3.

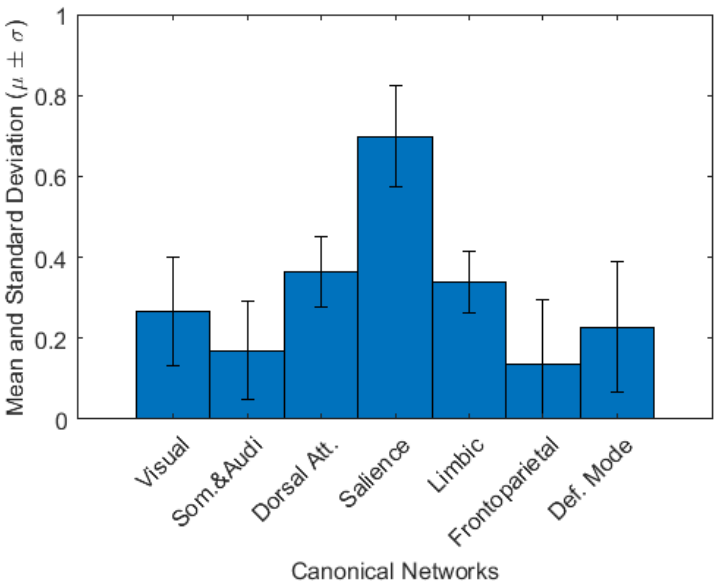

Figure 4.3. Means and standard deviations of the canonical axes' components computed over all modalities. The 4^{th} sub-network (Salience) is consistently observed to be the most important one [72].

The canonical sub-networks are ordered (from most to least important) based on their mean values provided in Figure 4.3, and used later to obtain discriminative power of each sNETs definition. LDA is employed using the top Q ($\in [1, 7]$) canonical sub-networks, separately for each modality. The results are shown in Figure 4.4. In general, adding more sub-networks' feature increases the discrimination score except for the Q $4-3$ case (i.e, using the dorsal attention and salience network).

Moreover, the weighted connectivity definition performs better in nearly all situations compared to FA-based network definitions.

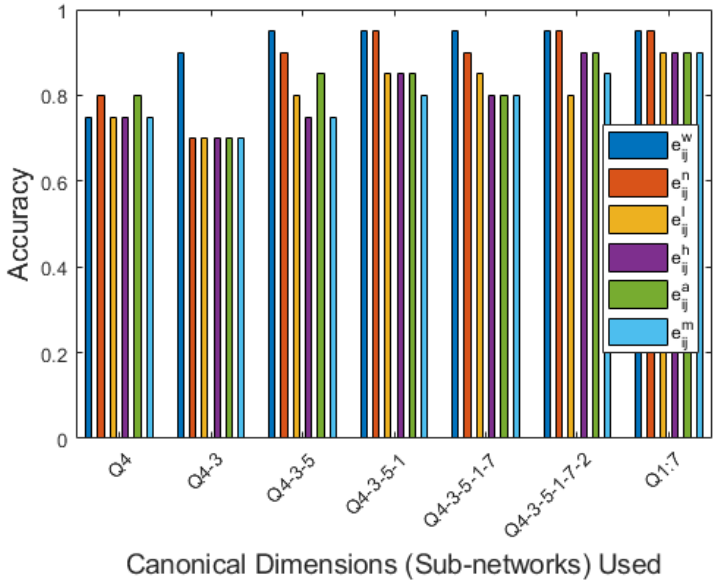

Figure 4.4. The accuracy of LDA classification on all 6 modalities using different subsets of 7 functional networks, Q1 - Visual, Q2 - Somatomotor & Auditory, Q3 - Dorsal attention, Q4 - Salience, Q5 - Limbic, Q6 - Frontoparietal, and Q7 - Default Mode [72].

4.3. Uni-Modal B-Tensor Factorization

This study was conducted on a total of 47 patients (an age and gender matched subset of all data), including 8 AD, 25 MCI and 14 SCI. Information about the subjects are given in Table 4.2. Experiments are done by using two types of fNETs and three types of sNETs: e_{ij}^{Pe} and e_{ij}^{Pc}, pearson correlation coefficient and partial correlation coefficient for fNETs, and e_{ij}^{n}, e_{ij}^{l}, e_{ij}^{a}; namely the normalized weighted fiber count, minimum fractional anisotropy, and average fractional anisotropy. In addition, the mentioned modalities are used separately to construct either B_f or B_s. Hence, with the usage of Destriuex Atlas parcellation dimensions for each modality becomes $B_u \in \mathbb{R}^{148\times148\times47}$.

Table 4.2. Summary of age and gender matched dataset [74].

Type	**AD**	**MCI**	**SCI**
# of Female/Male	5/3 (62.5%F)	16/9 (64%F)	9/5 (64%F)
Age$\in [51, 70]$	62.25 ± 6.61	61.28 ± 5.31	61.00 ± 5.99
CI-FCSRT	0.31 ± 0.21	0.66 ± 0.17	0.87 ± 0.12
TFR-FCSRT	7.25 ± 8.31	21.92 ± 4.54	34.5 ± 3.94

B_u decomposed into factors for all modality options with $Q \in \{3, 5, 7, 9, 11, 13, 15\}$ by using the algorithm shown in Figure 3.7. Initialization of the algorithm is randomly made 20 times, and the one with the minimum reconstruction error is chosen as an optimal solution. Reconstruction error is calculated by using l2-norm of the difference between the original tensor and its estimate, and dividing it to the number of elements. The average of all reconstruction coefficients for all modalities are less than $1e-4$.

Each individual connectome can be represented by the output of the decomposition in a Q-D space with columns of U_f and U_s, corresponding to fNETs and sNETs, respectively. Moreover, $V_{q,f}$, and $V_{q,s}$ represent the q^{th} sub-network of functional and structural connectomes that spans the Q-D spaces. U_f and U_s are used as feature vectors and employed to the classification of AD, MCI, and SCI. Moreover, multivariate analysis of variance (MANOVA) is utilized to obtain the significance of group separation in Q-D spaces [117]. All p-values for all B_f and B_s with different Q options are less than $1e^{-9}$, for details see Appendix Table D.1.

3-class linear support vector machine classifier is used to obtain separation performance based on leave-one-out cross-validation(LOOCV). The classification accuracies and the confusion matrices are presented in Tables 4.3 and 4.4, and confusion matrices for all Q values are represented in Appendix Table D.3. The Fisher-Pitman permutation test is employed to obtain the statistical significance of the observed classification accuracies [118, 119]. In order to implement Fisher-Pitman permutation test, the labels of Q-D connectome representations are randomly shuffled 1000 times at first, and

the same SVM approach is followed. The output accuracies are used to generate the null distribution of randomly acquired performances, and the p-values of the observed accuracies (less than $1e^{-}3$) show that the accuracy results are statistically significant for all modalities, see all p-values in Appendix Table D.2.

Table 4.3. Leave-one-out cross-validated classification accuracies for Q-dimensional connectome representations [74].

	Number of Basis Sub-networks (Q)						
	3	5	7	9	11	13	15
Functional Uni-modal (U_f)							
e_{ij}^{Pe}	0.96	1.00	1.00	1.00	0.96	0.94	0.94
e_{ij}^{Pc}	1.00	0.96	0.98	0.94	0.96	0.96	0.94
Structural Uni-modal (U_s)							
e_{ij}^{n}	0.85	0.91	0.87	0.85	0.85	0.83	0.83
e_{ij}^{l}	0.87	0.81	0.79	0.77	0.79	0.85	0.85
e_{ij}^{a}	0.81	0.79	0.81	0.79	0.81	0.87	0.87

Table 4.3 shows that the factorization of B_f generally performs better than B_s with respect to the classification accuracy. Table 4.4 shows the confusion matrices for both B_f and B_s factorizations when $Q = [5, 15]$. In structural connectomes, most of the misclassifications are observed between the AD-SCI groups. Moreover, fiber count based sNETs are relatively perform better than the FA based connectomes.

The distribution of the projection coefficients are displayed for $Q = 5$ sub-networks in Figure 4.5. CP scaling factors (d_q's) are given as percentages and represent each sub-networks role in reconstruction. As it can be seen from the Figure 4.5, fNETs' variation can be described mostly based on the first sub-network. On the other hand, sNETs' variation is almost evenly distributed across all sub-networks.

$V_{q,f}$ and $V_{q,s}$ sub-networks are displayed in the Figure 4.6 for $Q = 5$. The most apparent regions are the precuneus (L/R30), the angular gyrus (L/R25), the supra-marginal gyrus (L/R26), the superior parietal lobule (L/R27), the precentral gyrus

Table 4.4. Confusion matrices for 3-class classification when $Q = 5$ and $Q = 15$ [74].

		Q=5			Q=15		
		Predicted					
		SCI	MCI	AD	SCI	MCI	AD
e_{ij}^{Pe}	SCI	14	0	0	14	0	0
	MCI	0	25	0	0	25	0
	AD	0	0	8	2	1	5
e_{ij}^{Pc}	SCI	14	0	0	14	0	0
	MCI	0	25	0	0	25	0
	AD	2	0	6	3	0	5
e_{ij}^{n}	SCI	12	1	1	11	2	1
	MCI	1	24	0	3	22	0
	AD	1	0	7	2	0	6
e_{ij}^{l}	SCI	11	2	1	12	0	2
	MCI	2	23	0	1	24	0
	AD	4	0	4	4	0	4
e_{ij}^{a}	SCI	9	4	1	11	2	1
	MCI	3	22	0	2	23	0
	AD	2	0	6	1	0	7

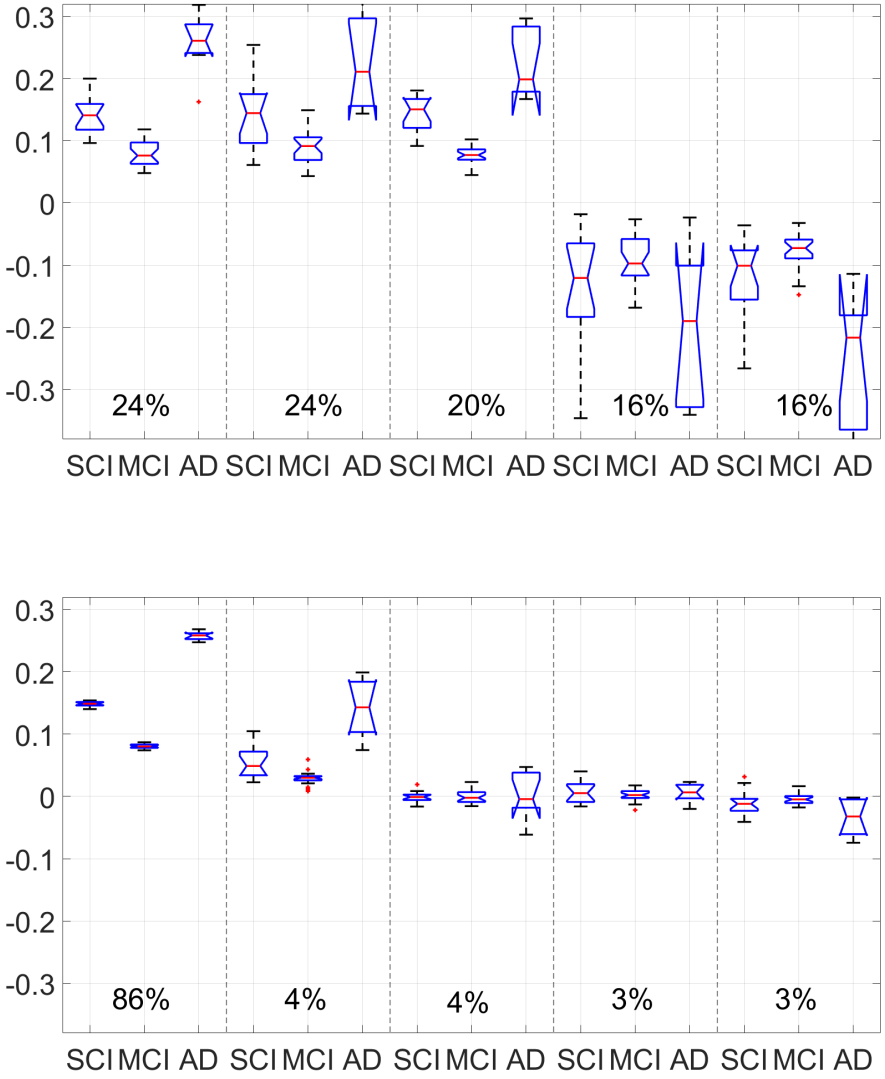

Figure 4.5. The distribution of projection coefficients of U_s (upper) and U_f (lower) per sub-network (left-to-right: $1^{st} - 5^{th}$) and per group (SCI/MCI/AD). The CP scaling factors (d_q's),given as percentages, represent the importance of corresponding sub-network in reconstruction of the B-Tensor [74].

(L/R29), the postcentral gyrus (L/R28), and the dorsal & ventral parts of posterior cingulate cortex (L/R 9-10). Left and right hemispheres are described with letter L (lower half) and letter R (upper half), respectively in the circular graph representation. In these sub-networks, only the strong connections are displayed which can be both positive or negative.

4.4. Multi-Modal B-Tensor Factorization

B_{cm} factorization has been made by using the same data described in Section 4.3, however, edge definitions are combined, one from sNETs and fNETs. Therefore, a total of 6 B_{cm} have been researched, and each of them has the same dimensions ($\in \mathbb{R}^{148\times148\times47\times2}$). The employed decomposition technique is based on the algorithm described in Figure 3.8 for all modality options with $Q \in \{3, 5, 7, 9, 11, 13, 15\}$.

Initialization of the factor matrices is randomly made 20 times, and the one with the minimum reconstruction error has been chosen as the optimal solution. Reconstruction errors of randomly initialized decompositions are averaged and found to be less than $1e^{-4}$. Columns of U_{cm} are used as feature vectors which are employed to 3-class linear SVM classifier to separate AD, MCI, and SCI based on LOOCV. In addition, MANOVA analysis is used to acquire the significance of classification in Q-D spaces. P-values of the MANOVA analyses are added to the Appendix Table D.4, and to summarize, p-values for all options with Q and modalities are less than $1e^{-19}$. Moreover, the accuracy results are further analyzed in terms of statistical significance by using the Fisher-Pitman permutation test as described in Section 4.3. All p-values are less than $1e^{-3}$, and the details can be found in Appendix Table D.5.

$V_{q,cm}$ sub-networks are displayed in Figure 4.8 when $Q = 5$. The most apparent regions are the precuneus (L/R30), the angular gyrus (L/R25), the supramarginal gyrus (L/R26), the superior parietal lobule (L/R27), the precentral gyrus (L/R29), and the postcentral gyrus (L/R28).

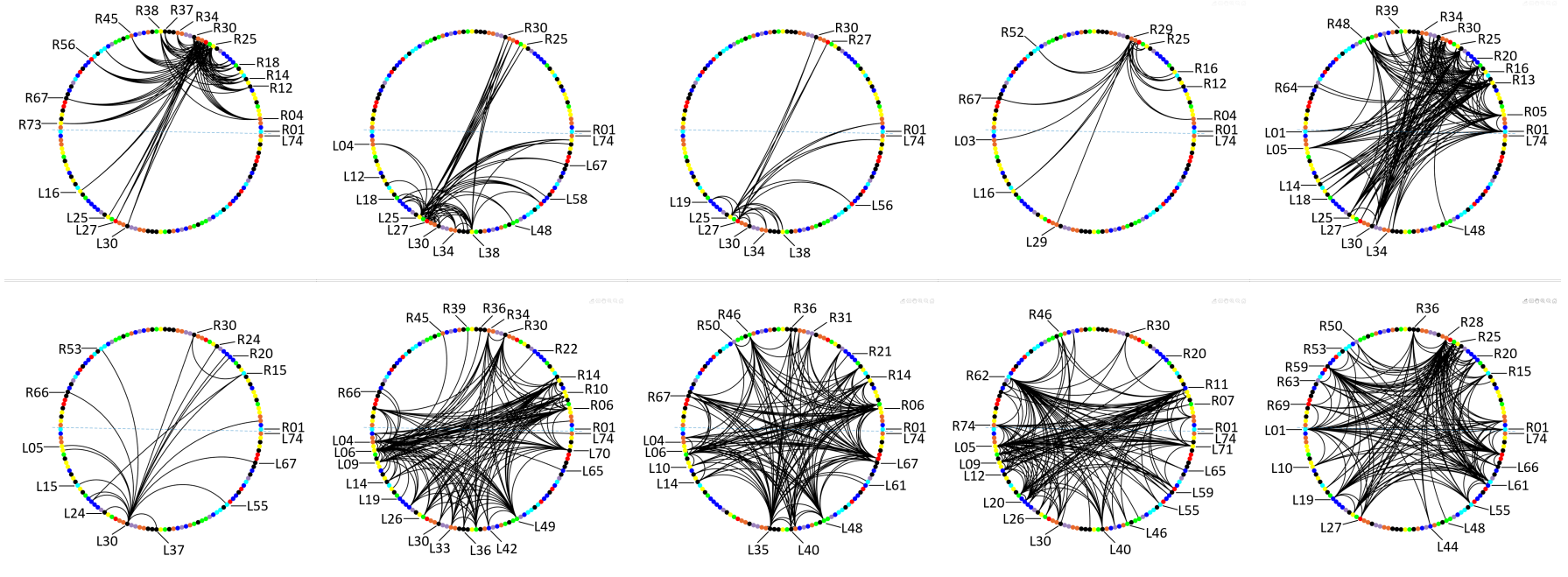

Figure 4.6. $V_{q,s}$ structural (upper row) and $V_{q,f}$ functional (lower row) basis sub-networks from left to right for $q = (1, \cdots, 5)$. The basis sub-networks are ordered with respect to $d_q s$ (left-to-right). The colors in the circular graphs represent seven functional networks (Red: Dorsal Attention, Green: Salience, Blue: Visual, Cyan: Frontoparietal, Purple: Limbic, Orange: Somatomotor & Auditory, Yellow: Default Mode, Black: Multiple Overlapping Networks) described in the literature [98]. Only the connections bigger than $\mu + 3.5\sigma$ are shown in the graphs [74].

The classification accuracies given in Table 4.5 show that multi-modal approach outperforms the uni-modal case even if the difference is small with respect to uni-modal fNETs' accuracy results (B_f). Projection coefficients' distribution of B_{cm} (e_{ij}^{n}+e_{ij}^{Pc}) which shows an evenly distributed sub-networks, is also displayed in Figure 4.7 when $Q = 5$. Confusion matrices for all pairwise modalities are given in Appendix Table D.6.

Table 4.5. Leave-one-out cross-validated classification accuracies for Q-dimensional connectome representations of B_{cm} [74].

	Number of Basis Sub-networks (Q)						
	3	5	7	9	11	13	15
Multi-modal (U_{cm})							
e_{ij}^{n}+e_{ij}^{Pc}	1.00	1.00	1.00	1.00	1.00	0.98	0.98
e_{ij}^{l}+e_{ij}^{Pc}	1.00	1.00	0.98	0.98	0.94	1.00	0.96
e_{ij}^{a}+e_{ij}^{Pc}	1.00	1.00	0.98	0.98	0.96	0.96	0.96
e_{ij}^{n}+e_{ij}^{Pe}	1.00	1.00	1.00	1.00	1.00	1.00	0.98
e_{ij}^{l}+e_{ij}^{Pe}	0.96	1.00	1.00	1.00	0.96	0.94	0.96
e_{ij}^{a}+e_{ij}^{Pe}	0.96	0.98	1.00	1.00	0.98	1.00	0.98

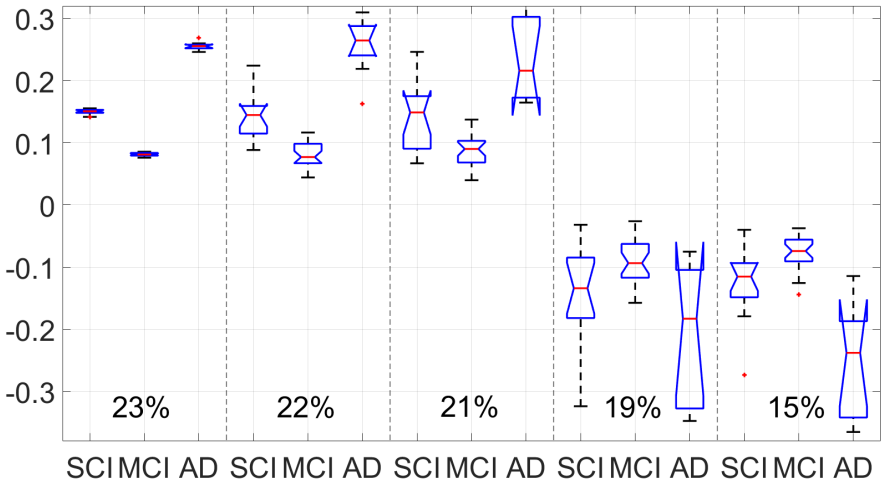

Figure 4.7. The distribution of the projection coefficients of U_{cm} per sub-network (left-to-right: $1^{st} - 5^{th}$) and per group (SCI/MCI/AD) for (e_{ij}^{n}+e_{ij}^{Pc}). The CP scaling factors (d_q's), given as a percentages, represent the importance of corresponding sub-network in reconstruction of the B-Tensor [74].

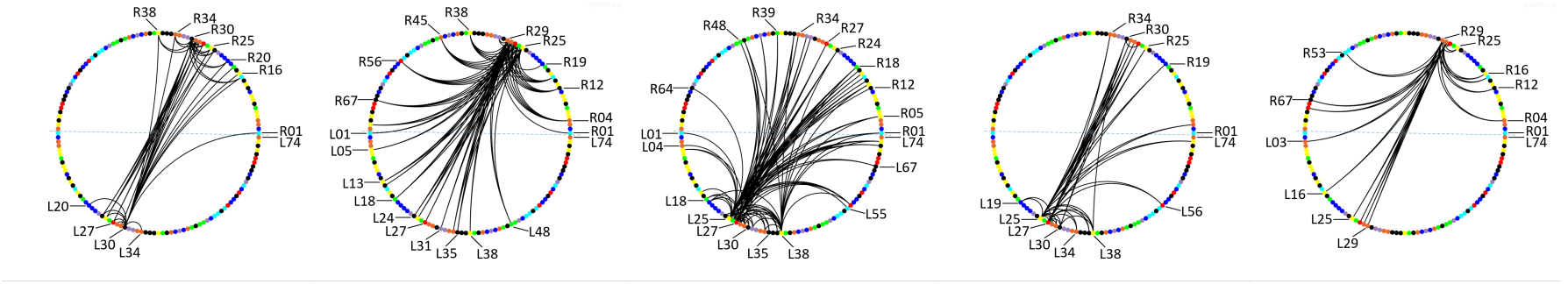

Figure 4.8. $V_{q,cm}$ basis sub-networks from left to right for $q = (1, \cdots, 5)$ with modalities $(e^{n}_{ij}+e^{Pc}_{ij})$ are depicted with respect to corresponding scaling factors $d_q s$ in descending order (left-to-right). The colors in the circular graphs represent seven functional networks (Red: Dorsal Attention, Green: Salience, Blue: Visual, Cyan: Frontoparietal, Purple: Limbic, Orange: Somatomotor & Auditory, Yellow: Default Mode, Black: Multiple Overlapping Networks) described in the literature [98]. Only the connections bigger than $\mu + 3.5\sigma$ are shown in the graphs [74].

5. DISCUSSION

Pathology of AD is defined by amyloid beta (Aβ) accumulation (senile plaques) and neurofibrillary tangles (tau concentration) in CSF [5, 120], and their relationships with different levels of AD are observed [121, 122]. Hence; Aβ, phosphorylated tau protein and tau protein in CSF are used as a biomarker for AD [123]. Increased levels of p-tau, and decreased levels of Aβ can be seen in AD patients with respect to controls or MCI [124]. Unlike Aβ aggregation, tau pathology correlates better with disease severity and cognitive decline [125]. Aβ and p-tau measurements of a subset of 35 patients in our data are observed to be related to the literature above, and displayed in Figure 5.1. As it can be seen, Aβ distribution decreases while p-tau concentration increases with respect to dementia stages of AD. However, both of them can not separate the groups well enough. In addition, Braak et al. observed stages in AD with respect to neurofibrillary tangles (NFTs) which is mostly related with the progression of atrophy found by MRI [126, 127]. Therefore, network analysis obtained by MRI is useful to obtain the underlying progress of AD. Hence, in this thesis, the methodologies based on pure network discrepancy highlight the differences of stages in AD by using sNETs and fNETs of patients.

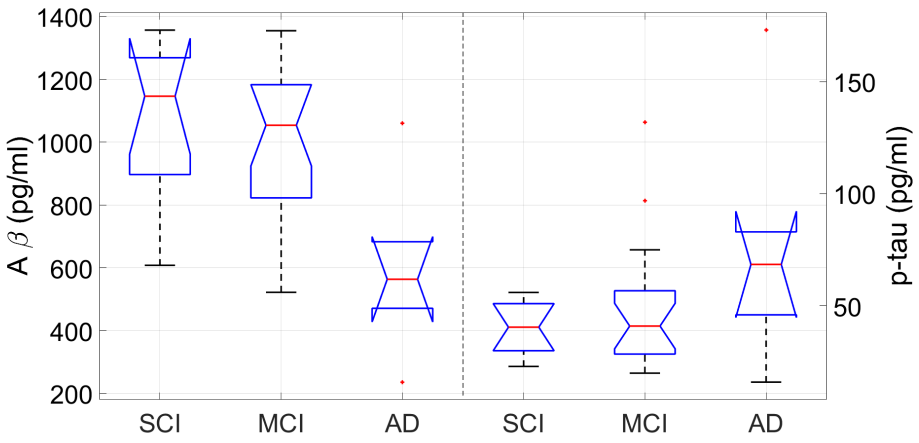

Figure 5.1. The Aβ and p-tau concentrations in CSF of 35 subjects [74].

The canonical sub-networks (G) are employed into the factorization of B_{sm} to obtain low dimensional representation (M) which allows us to acquire the importance of the sub-networks in discriminating the groups. AD patients and healthy controls are observed to be clearly classified with respect to these sub-networks in the study described in Section 4.2. The most prominent one among the 7 canonical sub-networks is the Salience sub-network. The dorsal attention and the limbic sub-networks were the second most important ones in differentiating AD from healthy control. Although, memory loss, one of the main symptoms of AD, is associated with the limbic sub-network, there are studies showing that the salience sub-network also has an effect on the progression of the disease [128]. Moreover, the diagnoses of AD patients are defined as early-AD by neurologists which may be effective in explaining the observed importance of salience sub-network. The limitation of projection onto known functional sub-networks (factors) stems from the fixed definitions of canonical sub-networks. However, this methodology can be developed further to obtain discriminative sub-networks rather than fixed sub-networks. Thus, more information can be acquired about the different levels of progression in neurodegenerative diseases.

Considering the performances of different sNETs' definitions, it can be seen that fiber-based definitions are more distinctive than FA-based definitions. FA measurements, seen as an indirect calculation of information flow between cortical regions, cannot be interpreted in this way with respect to the results. This could potentially be due to the shortcomings of diffusion tensor models which is employed to obtain FA values. Models using the microstructural integrity of fiber paths such as neurite orientation dispersion and density imaging (NODDI) will likely offer a higher discriminating performance [129].

The M tensor used in this study is a low-dimensional representation that exhibits how much canonical sub-networks are effective to reconstruct sNETs for each patient and modality. It may be possible to stage the progression of the disease with a regression analysis between the clinical test results and the M tensor. Although this study is limited to sNET definitions, it can be used to include fNETs due to its 4-dimensional

structure. With such an analysis, the relationships between functional and structural changes, and the progression of the disease can be observed. Small dataset size is the main limitation of this study. For this reason, the LDA analysis may be overfitted to represent all data, resulting in extreme precision. Studies with a larger data set will likely provide stronger results.

The study in Section 4.3, B_s and B_f are factorized, and the accuracy results that are shown in Table 4.3 clearly indicate the fNETs superiority over sNETs. The fNETs' performances are in line with the current literature which states that resting state functional networks are considered as a biomarker for neurodegenerative diseases [130]. In addition, performances of sNETs are actually high enough to state that the AD also propagates on the structure of the brain connectomes [131]. The resulting sub-networks, besides the achieved accuracy performances, give opportunity to gain insight into the spread of the disease with a functional and structural perspectives. Studies suggest that the sub-networks of the brain can be considered as a biomarker of AD [132]. Figure 4.6 illustrates the sub-networks found by uni-modal factorization. The precuneus, the angular gyrus, the supramarginal gyrus, the superior parietal lobule, the precentral gyrus and the postcentral gyrus are found to be the most prominent regions for sNETs factorization.

The precuneus is known to be the central node of the default mode network (DMN) and it contains a major way-station of the Papez circuit which is related to the episodic memory [133,134]. In addition, it is the earliest structure which shows an aggregated amyloid during the pre-clinical stage of AD, and this amyloid load disrupts both the intra-network and inter-network connectivity [135,136]. Moreover, disruptions in the precuneus are observed in individuals who are susceptible to AD even if amyloid accumulation does not exist [137]. During the MCI stage, the precuneus is also invaded by NFTs, that is to say, Braak & Braak stage of III-IV corresponds pathologically the involvement of the precuneus [126]. Hence, it is the most prominent structure in the neuropathalogical process of AD.

The angular gyrus and the supramarginal gyrus, (known as the two gyri of inferior parietal lobule), and superior parietal lobule (SPL) correspond to Braak & Braak stage of V which indicates the earliest phase of dementia [138]. The angular gyrus and the supramarginal gyrus contains a hub of the DMN and the ventral attention network (VAN), respectively. Moreover, SPL includes a hub of the dorsal attention network (DAN). Note that DAN and VAN (also known as salience network) are found to be the most prominent sub-networks in the earlier study [72]. By the disruption of these regions, linguistic and navigational functions are affected, and this fulfills the requirements of the description of dementia in which multiple cognitive deficits are observed [138].

NFTs first affect the entorhinal cortex, then manifest on hippocampus, and finally propagates to the neocortex which contains part of the cerebral cortex and cortical parts of the limbic system. In addition to these, association cortices exhibit a similar pattern. This progress has been used to define different stages of AD by Braak et al. [126, 139]. Even at the final stage (VI), the precentral and the postcentral gyri are not supposed to be affected as associated areas. Buckner et al. state that the inferior/-superior parietal lobules and the precuneus are cortical hub regions which are found to be more vulnerable to amyloid depositions that reflect the pathology of AD. However, there are not any cortical hubs found in the primary sensory and motor areas which corresponds to precentral and postcentral gyri [140]. Cortical hub regions are further investigated to assess the reason for their vulnerability, and this investigation leads to Activity Dependent Degeneration (ADD) theory that states the networks are damaged according to their neuronal activities that are mostly observed in hub regions [141]. The precuneus and the precentral gyrus are found as cortical hubs in ADD study. Moreover, postcentral gyrus, left precentral gyrus, precuneus, and superior parietal lobule are found to be statistically significant with respect to cortical thinning [142]. Furthermore, in another work, all prominent regions but angular gyrus are also found to be discriminative between AD and healthy controls [143].

Precentral and postcentral gyri are found to be discriminative in uni-modal sNETs factorization, unexpectedly since these regions are thought to be almost preserved in AD for most of the previous studies [126,140,144–146]. Precentral gyrus is in Brodmann area 4 which is related with the primary motor area, and postcentral gyrus is part of the somatosensory cortex which is responsible to process sensory information. Dysfunction in precentral gyrus can cause paralysis. In addition, lesion in postcentral gyrus can lead to agraphesthesia and/or astereognosis which are found to be highly related with AD [147]. However, these regions occasionally affected only in stage VI of Braak & Braak [126]. Cortical hubs are found to be vulnerable to Aβ depositions by Buckner et al., however, these regions are not presented as cortical hubs [140]. On the other hand, the precentral gyrus is presented as a cortical hub which shows a degeneration in this region because of its neuronal activities [141]. Moreover, cortical thinning is observed at postcentral and left precentral gyri in AD patients [142]. In another study, it has been found that primary motor and primary sensory cortex are involved in disease progression [148]. Stephen et al. observed that somatosensory cortex is affected early by the AD progress [149]. The effects of these regions on the progression of AD should be further investigated, while the results reveal conflicting findings, yet our study indicates that these regions are involved in the disease progression.

The importance of sub-networks found by factorizing B_f, unlike sNETs, indicates that the first sub-network is the most prominent, and the others can be neglected. In uni-modal fNETs factorization, the precuneus and dorsal & ventral parts of the posterior cingulate cortex (PCC) are observed to be the most conspicuous cortical regions, see Figure 4.6. The precuneus is already mentioned as the key region for dementia. Braak & Braak stage III-IV corresponds pathologically the involvement of dorsal & ventral parts of the PCC that are also the components of Papez circuit. These regions found in uni-modal sNETs and fNETs are also prominent in multi-modal approach, see Figure 4.8. Hence, the observed sub-networks are in line with the neurological literature and offer insight into disease mechanism, yet postcentral and precentral gyri should be studied further. Seven functional sub-networks and observed discriminative sub-networks do not show any direct similarity [98].

Multi-modal B_{cm} factorization results indicate an improvement in the accuracies with respect to the uni-modal approaches. It allows us to consider both sNETs and fNETs simultaneously with a novel fashion. The Results specify that sNETs and fNETs can be seen as independent observations, which is also suggested by the current literature that different MRI modalities show their strength to distinguish stages of AD when employed in parallel [150, 151]. High classification accuracies between different levels of AD are obtained with the proposed B-Tensor factorizations, which supports the notion of being a non-invasive clinical biomarker. As it can be seen from the Figure 5.1, Aβ and p-tau concentrations in CSF may not reflect any significant changes while alterations in brain connectomes play a key role in identifying different stages of AD.

Recent developments in understanding the disease forces researchers to redefine AD with biological symptoms which is called AT(N) criteria. AT(N) represents amyloid beta, p-tau and neuronal injury biomarker, respectively [152]. Those criteria suggest continuum for AD. In addition, connectome patterns observed in this thesis may have the ability to show the discrepancy between patients who have the same diagnose when the continuum of the disease is considered. For now, it is not possible to claim that these patterns have the ability to propose disease continuum, however, in future work, representations found by tensor factorization with a blind study may show its power in this direction.

Clinical applicability of the proposed approaches can be assessed to a greater degree with a larger dataset to derive a conclusion. If the observed results are further studied with a multi-center dataset, it probably leads to improvement in statistical significance, and also highlight the effect of diverse data acquisitions. In order to observe the continuum in the progress of dementia, longitudinal study may reflect the identification of individuals who are more susceptible to AD. Finally, there are many different tensor decomposition techniques in the literature, and a comparative analysis may provide a valuable contribution [153–155].

6. CONCLUSION

B-Tensor is a novel representation of brain connectomes that offers an intuitive understanding of the underlying mechanisms in neurodegenerative diseases. Several B-tensor factorization approaches are presented to obtain the discriminative sub-networks and the corresponding low dimensional reconstruction coefficients.

When canonical networks are known, preliminary results indicate that the salience sub-network is the most prominent among 7 sub-networks with a high accuracy power. Fiber based structural networks perform better than FA-based connectivity definitions. The methodology of this study is suitable to be extended for a joint analysis of sNETs and fNETs.

When sub-networks are not known, uni-modal and multi-modal approaches allow us to explore discriminative connectivity patterns (sub-networks). Uni-modal B-Tensor factorization results for sNETs offer evenly distributed low dimensional sub-networks as well as high discriminative power. In addition, FA-based structural definitions did not distinguish different levels of AD better than simple fiber counts which is in line with the literature. Neurological interpretations of the observed structural basis sub-networks are mostly coherent with the literature, yet further studies are needed to discover unexpected regions such as pre/post central gyri. Representation of discriminative sub-networks for uni-modal functional approach can mostly be achieved by only one sub-network with a high accuracy power. Structural and functional connectomes are combined efficiently with the multi-modal methodology since the results show that the classification performances are increased with respect to uni-modal approaches.

Low dimensional representations should be studied further to understand its relationship with the clinical observations, and the discriminative power of B-Tensor factorizations are needed to be confirmed with a larger dataset by projecting connectomes onto the found sub-networks.

REFERENCES

1. "World population ageing", `https://www.un.org/en/development/desa/population/publications/pdf/ageing/ WorldPopulationAgeing2019-Highlights.pdf`, 2019, accessed at November 2020.

2. Reitz, C. and R. Mayeux, "Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers", *Biochemical Pharmacology*, Vol. 88, No. 4, pp. 640–651, 2014.

3. Wortmann, M., "Importance of national plans for Alzheimer's disease and dementia", *Alzheimer's Research & Therapy*, Vol. 5, No. 5, p. 40, 2013.

4. Reitz, C., C. Brayne and R. Mayeux, "Epidemiology of Alzheimer disease", *Nature Reviews Neurology*, Vol. 5, No. 6, pp. 137–152, 2011.

5. Buckner, R. L., A. Z. Snyder, B. J. Shannon, G. LaRossa, R. Sachs, A. F. Fotenos, Y. I. Sheline, W. E. Klunk, C. A. Mathis, J. C. Morris *et al.*, "Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory", *Journal of Neuroscience*, Vol. 25, No. 34, pp. 7709–7717, 2005.

6. Xia, M., J. Wang and Y. He, "BrainNet Viewer: a network visualization tool for human brain connectomics", *PloS ONE*, Vol. 8, No. 7, p. e68910, 2013.

7. Delbeuck, X., M. Van der Linden and F. Collette, "Alzheimer's disease as a disconnection syndrome?", *Neuropsychology Review*, Vol. 13, No. 2, pp. 79–92, 2003.

8. Palop, J. J., J. Chin and L. Mucke, "A network dysfunction perspective on neurodegenerative diseases", *Nature*, Vol. 443, No. 7113, p. 768, 2006.

9. Bassett, D. S. and E. T. Bullmore, "Human brain networks in health and disease", *Current Opinion in Neurology*, Vol. 22, No. 4, p. 340, 2009.

10. Tijms, B. M., A. M. Wink, W. de Haan, W. M. van der Flier, C. J. Stam, P. Scheltens and F. Barkhof, "Alzheimer's disease: connecting findings from graph theoretical studies of brain networks", *Neurobiology of Aging*, Vol. 34, No. 8, pp. 2023–2036, 2013.

11. Kaiser, M., "A tutorial in connectome analysis: topological and spatial features of brain networks", *NeuroImage*, Vol. 57, No. 3, pp. 892–907, 2011.

12. Onnela, J.-P., J. Saramäki, J. Kertész and K. Kaski, "Intensity and coherence of motifs in weighted complex networks", *Physical Review E*, Vol. 71, No. 6, p. 065103, 2005.

13. He, Y., Z. Chen and A. Evans, "Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's disease", *Journal of Neuroscience*, Vol. 28, No. 18, pp. 4756–4766, 2008.

14. Stam, C., W. De Haan, A. Daffertshofer, B. Jones, I. Manshanden, A.-M. van Cappellen van Walsum, T. Montez, J. Verbunt, J. De Munck, B. Van Dijk *et al.*, "Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease", *Brain*, Vol. 132, No. 1, pp. 213–224, 2009.

15. Yao, Z., Y. Zhang, L. Lin, Y. Zhou, C. Xu, T. Jiang, A. D. N. Initiative *et al.*, "Abnormal cortical networks in mild cognitive impairment and Alzheimer's disease", *PLoS Computational Biology*, Vol. 6, No. 11, p. e1001006, 2010.

16. Supekar, K., V. Menon, D. Rubin, M. Musen and M. D. Greicius, "Network analysis of intrinsic functional brain connectivity in Alzheimer's disease", *PLoS Computational Biology*, Vol. 4, No. 6, p. e1000100, 2008.

17. Sanz-Arigita, E. J., M. M. Schoonheim, J. S. Damoiseaux, S. A. Rombouts,

E. Maris, F. Barkhof, P. Scheltens and C. J. Stam, "Loss of 'small-world'networks in Alzheimer's disease: graph analysis of FMRI resting-state functional connectivity", *PLoS ONE*, Vol. 5, No. 11, p. e13788, 2010.

18. Lo, C.-Y., P.-N. Wang, K.-H. Chou, J. Wang, Y. He and C.-P. Lin, "Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in Alzheimer's disease", *Journal of Neuroscience*, Vol. 30, No. 50, pp. 16876–16885, 2010.

19. Stam, C. J., B. Jones, G. Nolte, M. Breakspear and P. Scheltens, "Small-world networks and functional connectivity in Alzheimer's disease", *Cerebral Cortex*, Vol. 17, No. 1, pp. 92–99, 2007.

20. Khazaee, A., A. Ebrahimzadeh and A. Babajani-Feremi, "Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory", *Clinical Neurophysiology*, Vol. 126, No. 11, pp. 2132–2141, 2015.

21. Watts, D. J., *Small worlds: The dynamics of networks between order and randomness*, Princeton University Press Princeton, 2000.

22. Watts, D. J. and S. H. Strogatz, "Collective dynamics of 'small-world' networks", *Nature*, Vol. 393, No. 6684, pp. 440–442, 1998.

23. Newman, M. E., "The structure and function of complex networks", *SIAM Review*, Vol. 45, No. 2, pp. 167–256, 2003.

24. Zhao, X., Y. Liu, X. Wang, B. Liu, Q. Xi, Q. Guo, H. Jiang, T. Jiang and P. Wang, "Disrupted small-world brain networks in moderate Alzheimer's disease: a resting-state FMRI study", *PloS ONE*, Vol. 7, No. 3, p. e33540, 2012.

25. Tijms, B. M., C. Möller, H. Vrenken, A. M. Wink, W. de Haan, W. M. van der Flier, C. J. Stam, P. Scheltens and F. Barkhof, "Single-subject grey matter graphs in Alzheimer's disease", *PloS ONE*, Vol. 8, No. 3, p. e58921, 2013.

26. Sporns, O. and C. J. Honey, "Small worlds inside big brains", *Proceedings of the National Academy of Sciences*, Vol. 103, No. 51, pp. 19219–19220, 2006.

27. Humphries, M. D. and K. Gurney, "Network 'small-world-ness': a quantitative method for determining canonical network equivalence", *PloS ONE*, Vol. 3, No. 4, p. e0002051, 2008.

28. de Haan, W., Y. A. Pijnenburg, R. L. Strijers, Y. van der Made, W. M. van der Flier, P. Scheltens and C. J. Stam, "Functional neural network analysis in frontotemporal dementia and Alzheimer's disease using EEG and graph theory", *BMC Neuroscience*, Vol. 10, No. 1, pp. 1–12, 2009.

29. Freeman, L. C., "A set of measures of centrality based on betweenness", *Sociometry*, pp. 35–41, 1977.

30. Zhu, W., W. Wen, Y. He, A. Xia, K. J. Anstey and P. Sachdev, "Changing topological patterns in normal aging using large-scale structural networks", *Neurobiology of Aging*, Vol. 33, No. 5, pp. 899–913, 2012.

31. Wang, L., Y. Li, P. Metzak, Y. He and T. S. Woodward, "Age-related changes in topological patterns of large-scale brain functional networks during memory encoding and recognition", *NeuroImage*, Vol. 50, No. 3, pp. 862–872, 2010.

32. Contreras, J. A., J. Goñi, S. L. Risacher, O. Sporns and A. J. Saykin, "The structural and functional connectome and prediction of risk for cognitive impairment in older adults", *Current Behavioral Neuroscience Reports*, Vol. 2, No. 4, pp. 234–245, 2015.

33. Alvarez-Hamelin, J. I., L. Dall'Asta, A. Barrat and A. Vespignani, "Large scale networks fingerprinting and visualization using the k-core decomposition", *Advances in neural information processing systems*, pp. 41–50, 2006.

34. Daianu, M., N. Jahanshad, T. M. Nir, A. W. Toga, C. R. Jack Jr, M. W.

Weiner and P. M. Thompson, for the Alzheimer's Disease Neuroimaging Initiative, "Breakdown of brain connectivity between normal aging and Alzheimer's disease: a structural k-core network analysis", *Brain Connectivity*, Vol. 3, No. 4, pp. 407–422, 2013.

35. Daianu, M., N. Jahanshad, T. M. Nir, C. R. Jack Jr, M. W. Weiner, M. A. Bernstein, P. M. Thompson and A. D. N. Initiative, "Rich club analysis in the Alzheimer's disease connectome reveals a relatively undisturbed structural core network", *Human Brain Mapping*, Vol. 36, No. 8, pp. 3087–3103, 2015.

36. Boccaletti, S., V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang, "Complex networks: Structure and dynamics", *Physics Reports*, Vol. 424, No. 4-5, pp. 175–308, 2006.

37. Xie, T. and Y. He, "Mapping the Alzheimer's brain with connectomics", *Frontiers in Psychiatry*, Vol. 2, p. 77, 2012.

38. He, Y., Z. Chen, G. Gong and A. Evans, "Neuronal networks in Alzheimer's disease", *The Neuroscientist*, Vol. 15, No. 4, pp. 333–350, 2009.

39. Stam, C. J. and J. C. Reijneveld, "Graph theoretical analysis of complex networks in the brain", *Nonlinear Biomedical Physics*, Vol. 1, No. 1, p. 3, 2007.

40. delEtoile, J. and H. Adeli, "Graph theory and brain connectivity in Alzheimer's disease", *The Neuroscientist*, Vol. 23, No. 6, pp. 616–626, 2017.

41. Fornito, A., A. Zalesky and E. Bullmore, *Fundamentals of brain network analysis*, Academic Press, 2016.

42. Kelly, C., B. B. Biswal, R. C. Craddock, F. X. Castellanos and M. P. Milham, "Characterizing variation in the functional connectome: promise and pitfalls", *Trends in Cognitive Sciences*, Vol. 16, No. 3, pp. 181–188, 2012.

43. Varoquaux, G. and R. C. Craddock, "Learning and comparing functional connectomes across subjects", *NeuroImage*, Vol. 80, pp. 405–415, 2013.

44. Dosenbach, N. U., B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-Schlaggar *et al.*, "Prediction of individual brain maturity using fMRI", *Science*, Vol. 329, No. 5997, pp. 1358–1361, 2010.

45. Jbabdi, S. and H. Johansen-Berg, "Tractography: where do we go from here?", *Brain Connectivity*, Vol. 1, No. 3, pp. 169–183, 2011.

46. Rubinov, M. and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations", *NeuroImage*, Vol. 52, No. 3, pp. 1059–1069, 2010.

47. Craddock, R. C., S. Jbabdi, C.-G. Yan, J. T. Vogelstein, F. X. Castellanos, A. Di Martino, C. Kelly, K. Heberlein, S. Colcombe and M. P. Milham, "Imaging human connectomes at the macroscale", *Nature Methods*, Vol. 10, No. 6, p. 524, 2013.

48. Efron, B. *et al.*, "Simultaneous inference: When should hypothesis testing problems be combined?", *The Annals of Applied Statistics*, Vol. 2, No. 1, pp. 197–223, 2008.

49. Zalesky, A., A. Fornito and E. T. Bullmore, "Network-based statistic: identifying differences in brain networks", *NeuroImage*, Vol. 53, No. 4, pp. 1197–1207, 2010.

50. Ahuja, R. K., J. B. Orlin and T. L. Magnanti, *Network flows: theory, algorithms, and applications*, Prentice-Hall, 1993.

51. Zalesky, A., L. Cocchi, A. Fornito, M. M. Murray and E. Bullmore, "Connectivity differences in brain networks", *NeuroImage*, Vol. 60, No. 2, pp. 1055–1062, 2012.

52. Meskaldji, D. E., M.-C. Ottet, L. Cammoun, P. Hagmann, R. Meuli, S. Eliez,

J. P. Thiran and S. Morgenthaler, "Adaptive strategy for the statistical analysis of connectomes", *PloS ONE*, Vol. 6, No. 8, p. e23009, 2011.

53. Meskaldji, D.-E., S. Morgenthaler and D. Van De Ville, "Statistical methods for comparing brain connectomes at different scales", *Wavelets and Sparsity XVI*, Vol. 9597, p. 95971L, International Society for Optics and Photonics, 2015.

54. Kim, J., W. Pan, A. D. N. Initiative *et al.*, "Highly adaptive tests for group differences in brain functional connectivity", *NeuroImage: Clinical*, Vol. 9, pp. 625–639, 2015.

55. Baumgartner, R., L. Ryner, W. Richter, R. Summers, M. Jarmasz and R. Somorjai, "Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis", *Magnetic Resonance Imaging*, Vol. 18, No. 1, pp. 89–94, 2000.

56. Ding, C. and X. He, "K-means clustering via principal component analysis", *Proceedings of the twenty-first international conference on Machine learning*, p. 29, 2004.

57. Leonardi, N., J. Richiardi, M. Gschwind, S. Simioni, J.-M. Annoni, M. Schluep, P. Vuilleumier and D. Van De Ville, "Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest", *NeuroImage*, Vol. 83, pp. 937–950, 2013.

58. Ding, C., X. He and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering", *Proceedings of the 2005 SIAM international conference on data mining*, pp. 606–610, SIAM, 2005.

59. Beckmann, C. F., M. DeLuca, J. T. Devlin and S. M. Smith, "Investigations into resting-state connectivity using independent component analysis", *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 360, No. 1457, pp.

1001–1013, 2005.

60. Calhoun, V. D., J. Liu and T. Adalı, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data", *NeuroImage*, Vol. 45, No. 1, pp. S163–S172, 2009.

61. Anderson, A., P. K. Douglas, W. T. Kerr, V. S. Haynes, A. L. Yuille, J. Xie, Y. N. Wu, J. A. Brown and M. S. Cohen, "Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD", *NeuroImage*, Vol. 102, pp. 207–219, 2014.

62. Lee, D. D. and H. S. Seung, "Algorithms for non-negative matrix factorization", *Advances in Neural Information Processing Systems*, pp. 556–562, 2001.

63. Khosla, M., K. Jamison, G. H. Ngo, A. Kuceyeski and M. R. Sabuncu, "Machine learning in resting-state fMRI analysis", *Magnetic Resonance Imaging*, Vol. 64, pp. 101–121, 2019.

64. Ou, J., L. Xie, X. Li, D. Zhu, D. P. Terry, A. N. Puente, R. Jiang, Y. Chen, L. Wang, D. Shen *et al.*, "Atomic connectomics signatures for characterization and differentiation of mild cognitive impairment", *Brain Imaging and Behavior*, Vol. 9, No. 4, pp. 663–677, 2015.

65. Yang, Z. and E. Oja, "Linear and nonlinear projective nonnegative matrix factorization", *IEEE Transactions on Neural Networks*, Vol. 21, No. 5, pp. 734–749, 2010.

66. Eavani, H., T. D. Satterthwaite, R. E. Gur, R. C. Gur and C. Davatzikos, "Discriminative sparse connectivity patterns for classification of fMRI data", *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 193–200, Springer, 2014.

67. Eavani, H., T. D. Satterthwaite, R. Filipovych, R. E. Gur, R. C. Gur and C. Da-

vatzikos, "Identifying sparse connectivity patterns in the brain using resting-state fMRI", *NeuroImage*, Vol. 105, pp. 286–299, 2015.

68. Batmanghelich, N. K., B. Taskar and C. Davatzikos, "Generative-discriminative basis learning for medical imaging", *IEEE Transactions on Medical Imaging*, Vol. 31, No. 1, pp. 51–69, 2011.

69. Zille, P., V. D. Calhoun, J. M. Stephen, T. W. Wilson and Y.-P. Wang, "Fused estimation of sparse connectivity patterns from rest fMRI—Application to comparison of children and adult brains", *IEEE Transactions on Medical Imaging*, Vol. 37, No. 10, pp. 2165–2175, 2017.

70. Allen, E. A., E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state", *Cerebral Cortex*, Vol. 24, No. 3, pp. 663–676, 2014.

71. Cai, B., P. Zille, J. M. Stephen, T. W. Wilson, V. D. Calhoun and Y. P. Wang, "Estimation of dynamic sparse connectivity patterns from resting state fMRI", *IEEE Transactions on Medical Imaging*, Vol. 37, No. 5, pp. 1224–1234, 2017.

72. Durusoy, G., A. Karaaslanlı, D. Y. Dal, Z. Yıldırım and B. Acar, "Multi-modal Brain Tensor Factorization: Preliminary Results with AD Patients", *International Workshop on Connectomics in Neuroimaging*, pp. 29–37, Springer, 2018.

73. Zhang, Z., G. I. Allen, H. Zhu and D. Dunson, "Tensor network factorizations: Relationships between brain structural connectomes and traits", *NeuroImage*, Vol. 197, pp. 330–343, 2019.

74. Durusoy, G., Z. Yildirim, D. Y. Dal, C. Ulasoglu-Yildiz, E. Kurt, G. Bayir, E. Ozacar, E. Ozarslan, A. Demirtas-Tatlidede, B. Bilgic *et al.*, "B-Tensor: Brain Connectome Tensor Factorization for Alzheimer's Disease", *IEEE Journal of Biomedical and Health Informatics*, 2020, in press.

75. Kolda, T. and B. Bader, "Tensor decompositions and applications", *SIAM Review*, Vol. 51, No. 3, pp. 455–500, 2009.

76. Yang, W. and Y. Cui, "Tensor factorization-based classification of Alzheimer's disease vs healthy controls", *2012 5th International Conference on BioMedical Engineering and Informatics*, pp. 371–374, IEEE, 2012.

77. Karahan, E., P. A. Rojas-Lopez, M. L. Bringas-Vega, P. A. Valdés-Hernández and P. A. Valdes-Sosa, "Tensor analysis and fusion of multimodal brain images", *Proceedings of the IEEE*, Vol. 103, No. 9, pp. 1531–1559, 2015.

78. De Vos, M., D. Nion, S. Van Huffel and L. De Lathauwer, "A combination of parallel factor and independent component analysis", *Signal Processing*, Vol. 92, No. 12, pp. 2990–2999, 2012.

79. Cao, B., L. He, X. Wei, M. Xing, P. S. Yu, H. Klumpp and A. D. Leow, "t-bne: Tensor-based brain network embedding", *Proceedings of the 2017 SIAM International Conference on Data Mining*, pp. 189–197, SIAM, 2017.

80. Caiafa, C. F., O. Sporns, A. Saykin and F. Pestilli, "Unified representation of tractography and diffusion-weighted MRI data using sparse multidimensional arrays", *Advances in Neural Information Processing Systems*, pp. 4340–4351, 2017.

81. Aminmansour, F., A. Patterson, L. Le, Y. Peng, D. Mitchell, F. Pestilli, C. F. Caiafa, R. Greiner and M. White, "Learning macroscopic brain connectomes via group-sparse factorization", *Advances in Neural Information Processing Systems*, pp. 8849–8859, 2019.

82. Hamdi, S. M., Y. Wu, R. Angryk, L. C. Krishnamurthy and R. Morris, "Identification of discriminative subnetwork from fmri-based complete functional connectivity networks", *International Journal of Semantic Computing*, Vol. 13, No. 01, pp. 25–44, 2019.

83. Zhang, Y.-P., L. Xiao, G. Zhang, B. Cai, J. M. Stephen, T. W. Wilson, V. D. Calhoun and Y.-P. Wang, "Multi-paradigm fMRI fusion via sparse tensor decomposition in brain functional connectivity study", *IEEE Journal of Biomedical and Health Informatics*, 2020.

84. Li, N., W. Zhang, S. Cheng and Y. Liu, "Structured Sparse Representation in Resting-State fMRI Image Processing for Mild Traumatic Brain Injury Analysis", *2019 International Conference on Communications, Information System and Computer Engineering (CISCE)*, pp. 445–450, IEEE, 2019.

85. "FSL Toolbox", `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL`, 2021, accessed at January 2021.

86. "FreeSurfer Toolbox", `https://surfer.nmr.mgh.harvard.edu/`, 2021, accessed at January 2021.

87. "Tortoise Toolbox", `https://tortoise.nibib.nih.gov/`, 2021, accessed at January 2021.

88. "Preprocessing Pipeline", `https://vavlab.boun.edu.tr/brainet-structural-and-functional-brain-network-analysis`, 2019, accessed at November 2020.

89. Glasser, M. F., S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni *et al.*, "The minimal preprocessing pipelines for the Human Connectome Project", *NeuroImage*, Vol. 80, pp. 105–124, 2013.

90. "Human Connectome Project Pipeline Scripts", `https://github.com/Washington-University/HCPpipelines`, 2021, accessed at January 2021.

91. Greve, D. N. and B. Fischl, "Accurate and robust brain image alignment using boundary-based registration", *NeuroImage*, Vol. 48, No. 1, pp. 63–72, 2009.

92. "Recon-all Steps", `https://surfer.nmr.mgh.harvard.edu/fswiki/recon-all`, 2021, accessed at January 2021.

93. Sporns, O., "The human connectome: a complex network", *Annals of the New York Academy of Sciences*, Vol. 1224, No. 1, pp. 109–125, 2011.

94. Thirion, B. and O. Faugeras, "Nonlinear dimension reduction of fMRI data: the Laplacian embedding approach", *2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*, pp. 372–375, IEEE, 2004.

95. Arslan, S., S. I. Ktena, A. Makropoulos, E. C. Robinson, D. Rueckert and S. Parisot, "Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex", *NeuroImage*, Vol. 170, pp. 5–30, 2018.

96. Fischl, B., A. Van Der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy *et al.*, "Automatically parcellating the human cerebral cortex", *Cerebral Cortex*, Vol. 14, No. 1, pp. 11–22, 2004.

97. Destrieux, C., B. Fischl, A. Dale and E. Halgren, "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature", *NeuroImage*, Vol. 53, No. 1, pp. 1–15, 2010.

98. Yeo, B. T., F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni *et al.*, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity", *Journal of Neurophysiology*, 2011.

99. O'Donnell, L. J. and C.-F. Westin, "An introduction to diffusion tensor image analysis", *Neurosurgery Clinics*, Vol. 22, No. 2, pp. 185–196, 2011.

100. Basser, P. J., "Inferring microstructural features and the physiological state of

tissues from diffusion-weighted images", *NMR in Biomedicine*, Vol. 8, No. 7, pp. 333–344, 1995.

101. Le Bihan, D., "Molecular diffusion nuclear magnetic resonance imaging.", *Magnetic Resonance Quarterly*, Vol. 7, No. 1, p. 1, 1991.

102. Kingsley, P. B., "Introduction to diffusion tensor imaging mathematics: Part III. Tensor calculation, noise, simulations, and optimization", *Concepts in Magnetic Resonance Part A*, Vol. 28, No. 2, pp. 155–179, 2006.

103. Le Bihan, D., J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko and H. Chabriat, "Diffusion tensor imaging: concepts and applications", *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, Vol. 13, No. 4, pp. 534–546, 2001.

104. Tench, C., P. Morgan, M. Wilson and L. Blumhardt, "White matter mapping using diffusion tensor MRI", *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, Vol. 47, No. 5, pp. 967–972, 2002.

105. Moyer, D., B. A. Gutman, J. Faskowitz, N. Jahanshad and P. M. Thompson, "Continuous representations of brain connectivity using spatial point processes", *Medical Image Analysis*, Vol. 41, pp. 32–39, 2017.

106. Bowman, F. D., L. Zhang, G. Derado and S. Chen, "Determining functional connectivity using fMRI data with diffusion-based anatomical weighting", *NeuroImage*, Vol. 62, No. 3, pp. 1769–1779, 2012.

107. Sanchez, E. and B. R. Kowalski, "Tensorial resolution: a direct trilinear decomposition", *Journal of Chemometrics*, Vol. 4, No. 1, pp. 29–45, 1990.

108. Harshman, R. A. and M. E. Lundy, "Uniqueness proof for a family of models sharing features of Tucker's three-mode factor analysis and PARAFAC/CANDE-

COMP", *Psychometrika*, Vol. 61, No. 1, pp. 133–154, 1996.

109. Allen, G., "Sparse higher-order principal components analysis", *Artificial Intelligence and Statistics*, pp. 27–36, 2012.

110. Albert, M. S., S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, N. C. Fox, A. Gamst, D. M. Holtzman, W. J. Jagust, R. C. Petersen *et al.*, "The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease", *Alzheimer's & Dementia*, Vol. 7, No. 3, pp. 270–279, 2011.

111. Buschke, H., "Selective reminding for analysis of memory and learning", *Journal of Verbal Learning and Verbal Behavior*, Vol. 12, No. 5, pp. 543–550, 1973.

112. Grober, E., A. E. Sanders, C. Hall and R. B. Lipton, "Free and cued selective reminding identifies very mild dementia in primary care", *Alzheimer Disease and Associated Disorders*, Vol. 24, No. 3, p. 284, 2010.

113. McKhann, G. M., D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack Jr, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux *et al.*, "The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease", *Alzheimer's & Dementia*, Vol. 7, No. 3, pp. 263–269, 2011.

114. Li, C., J. Neugroschl, X. Luo, C. Zhu, P. Aisen, S. Ferris and M. Sano, "The utility of the cognitive function instrument (CFI) to detect cognitive decline in non-demented older adults", *Journal of Alzheimer's Disease*, Vol. 60, No. 2, pp. 427–437, 2017.

115. Hughes, C. P., L. Berg, W. Danziger, L. A. Coben and R. L. Martin, "A new

clinical scale for the staging of dementia", *The British Journal of Psychiatry*, Vol. 140, No. 6, pp. 566–572, 1982.

116. Fisher, R. A., "The use of multiple measurements in taxonomic problems", *Annals of Eugenics*, Vol. 7, No. 2, pp. 179–188, 1936.

117. Johnson, R. and D. Wichern, *Applied Multivariate Statistical Analysis*, Pearson Modern Classics for Advanced Statistics Series, Pearson, 2018.

118. Good, P., *Permutation tests: a practical guide to resampling methods for testing hypotheses*, Springer Science & Business Media, 2013.

119. Boik, R. J., "The Fisher-Pitman permutation test: A non-robust alternative to the normal theory F test when variances are heterogeneous", *British Journal of Mathematical and Statistical Psychology*, Vol. 40, No. 1, pp. 26–42, 1987.

120. Mirra, S. S., A. Heyman, D. McKeel, S. Sumi, B. J. Crain, L. Brownlee, F. Vogel, J. Hughes, G. Van Belle, L. Berg *et al.*, "The Consortium to Establish a Registry for Alzheimer's Disease (CERAD): Part II. Standardization of the neuropathologic assessment of Alzheimer's disease", *Neurology*, Vol. 41, No. 4, pp. 479–479, 1991.

121. Jack, C., D. S. Knopman, W. J. Jagust, L. Shaw, P. Aisen, M. Weiner, R. Petersen and J. Trojanowski, "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade", *The Lancet Neurology*, Vol. 9, No. 1, pp. 119–128, 2010.

122. Jack, C., D. Bennett, K. Blennow, M. Carrillo, B. Dunn, H. S.B., D. Holtzman, W. Jagust, J. Jessen, F. amd Karlawish, E. Liu, J. Molinuevo, T. Montine, C. Phelps, K. Rankin, C. Rowe, P. Scheltens, E. Siemers, H. Snyder and R. Sperling, "NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease.", *Alzheimer's & Dementia*, Vol. 14, No. 4, pp. 535–562, 2018.

123. Blennow, K. and H. Hampel, "CSF markers for incipient Alzheimer's disease",

The Lancet Neurology, Vol. 2, No. 10, pp. 605–613, 2003.

124. Höglund, K., O. Hansson, P. Buchhave, H. Zetterberg, P. Lewczuk, E. Londos, K. Blennow, L. Minthon and J. Wiltfang, "Prediction of Alzheimer's disease using a cerebrospinal fluid pattern of C-terminally truncated β-amyloid peptides", *Neurodegenerative Diseases*, Vol. 5, No. 5, pp. 268–276, 2008.

125. Pais, M., L. Martinez, O. Ribeiro, J. Loureiro, R. Fernandez, L. Valiengo, P. Canineu, F. Stella, L. Talib, M. Radanovic *et al.*, "Early diagnosis and treatment of Alzheimer's disease: new definitions and challenges", *Brazilian Journal of Psychiatry*, Vol. 42, No. 4, 2020.

126. Braak, H. and E. Braak, "Staging of Alzheimer's disease-related neurofibrillary changes", *Neurobiology of Aging*, Vol. 16, No. 3, pp. 271–278, 1995.

127. Johnson, K. A., N. C. Fox, R. A. Sperling and W. E. Klunk, "Brain imaging in Alzheimer disease", *Cold Spring Harbor Perspectives in Medicine*, Vol. 2, No. 4, p. a006213, 2012.

128. Seeley, W. W., R. K. Crawford, J. Zhou, B. L. Miller and M. D. Greicius, "Neurodegenerative diseases target large-scale human brain networks", *Neuron*, Vol. 62, No. 1, pp. 42–52, 2009.

129. Zhang, H., T. Schneider, C. A. Wheeler-Kingshott and D. C. Alexander, "NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain", *NeuroImage*, Vol. 61, No. 4, pp. 1000–1016, 2012.

130. Hohenfeld, C., C. J. Werner and K. Reetz, "Resting-state connectivity in neurodegenerative disorders: Is there potential for an imaging biomarker?", *NeuroImage: Clinical*, Vol. 18, pp. 849–870, 2018.

131. Peraza, L. R., A. Díaz-Parra, O. Kennion, D. Moratal, J.-P. Taylor, M. Kaiser, R. Bauer, A. D. N. Initiative *et al.*, "Structural connectivity centrality changes

mark the path toward Alzheimer's disease", *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, Vol. 11, pp. 98–107, 2019.

132. Teipel, S., M. J. Grothe, J. Zhou, J. Sepulcre, M. Dyrba, C. Sorg and C. Babiloni, "Measuring cortical connectivity in Alzheimer's disease as a brain neural network pathology: toward clinical applications", *Journal of the International Neuropsychological Society*, Vol. 22, No. 2, pp. 138–163, 2016.

133. Mesulam, M.-M., "Large-scale neurocognitive networks and distributed processing for attention, language, and memory", *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, Vol. 28, No. 5, pp. 597–613, 1990.

134. Catani, M., F. Dell'Acqua and M. T. De Schotten, "A revised limbic system model for memory, emotion and behaviour", *Neuroscience & Biobehavioral Reviews*, Vol. 37, No. 8, pp. 1724–1737, 2013.

135. Villemagne, V. L., K. E. Pike, G. Chetelat, K. A. Ellis, R. S. Mulligan, P. Bourgeat, U. Ackermann, G. Jones, C. Szoeke, O. Salvado, R. Martins, G. O'Keefe, C. Mathis, W. Klunk, D. Ames, C. Masters and C. Rowe, "Longitudinal assessment of Aβ and cognition in aging and Alzheimer disease", *Annals of Neurology*, Vol. 69, No. 1, pp. 181–192, 2011.

136. Sheline, Y. I., M. E. Raichle, A. Z. Snyder, J. C. Morris, D. Head, S. Wang and M. A. Mintun, "Amyloid plaques disrupt resting state default mode network connectivity in cognitively normal elderly", *Biological Psychiatry*, Vol. 67, No. 6, pp. 584–587, 2010.

137. Sheline, Y. I., J. C. Morris, A. Z. Snyder, J. L. Price, Z. Yan, G. D'Angelo, L. Collin, D. Sachin, B. Tammie, F. Anne, G. Alison and M. A. Mintun, "APOE4 allele disrupts resting state fMRI connectivity in the absence of amyloid plaques or decreased CSF Abeta42", *Journal of Neuroscience*, Vol. 30, No. 50, pp. 17035–

17040, 2010.

138. Mesulam, M., *Principles of behavioral and cognitive neurology*, Oxford University Press, 2000.

139. Pooler, A. M., M. Polydoro, S. Wegmann, S. B. Nicholls, T. L. Spires-Jones and B. T. Hyman, "Propagation of tau pathology in Alzheimer's disease: identification of novel therapeutic targets", *Alzheimer's Research & Therapy*, Vol. 5, No. 5, p. 49, 2013.

140. Buckner, R. L., J. Sepulcre, T. Talukdar, F. M. Krienen, H. Liu, T. Hedden, J. R. Andrews-Hanna, R. A. Sperling and K. A. Johnson, "Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease", *Journal of Neuroscience*, Vol. 29, No. 6, pp. 1860–1873, 2009.

141. de Haan, W., K. Mott, E. C. van Straaten, P. Scheltens and C. J. Stam, "Activity dependent degeneration explains hub vulnerability in Alzheimer's disease", *PLoS Computational Biology*, Vol. 8, No. 8, p. e1002582, 2012.

142. Kang, K., U. Yoon, J.-M. Lee and H.-W. Lee, "Idiopathic normal-pressure hydrocephalus, cortical thinning, and the cerebrospinal fluid tap test", *Journal of the Neurological Sciences*, Vol. 334, No. 1-2, pp. 55–62, 2013.

143. Zhang, Y., Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang and T.-F. Yuan, "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning", *Frontiers in Computational Neuroscience*, Vol. 9, p. 66, 2015.

144. Rapoport, S. I., B. Horwitz, J. V. Haxby and C. L. Grady, "Alzheimer's disease: metabolic uncoupling of associative brain regions", *Canadian Journal of Neurological Sciences*, Vol. 13, No. S4, pp. 540–545, 1986.

145. Kemper, T. L., "Neuroanatomical and neuropathological changes during aging and dementia.", *Clinical Neurology of Aging*, 1994.

146. Brun, A. and E. Englund, "Regional pattern of degeneration in Alzheimer's disease: neuronal loss and histopathological grading", *Histopathology*, Vol. 5, No. 5, pp. 549–564, 1981.

147. Davis, A. S., A. Mazur-Mosiewicz and R. S. Dean, "The presence and predictive value of astereognosis and agraphesthesia in patients with Alzheimer's disease", *Applied Neuropsychology*, Vol. 17, No. 4, pp. 262–266, 2010.

148. Suva, D., I. Favre, R. Kraftsik, M. Esteban, A. Lobrinus and J. Miklossy, "Primary motor cortex involvement in Alzheimer disease", *Journal of Neuropathology and Experimental Neurology*, Vol. 58, No. 11, pp. 1125–1134, 1999.

149. Stephen, J. M., R. Montaño, C. H. Donahue, J. C. Adair, J. Knoefel, C. Qualls, B. Hart, D. Ranken and C. J. Aine, "Somatosensory responses in normal aging, mild cognitive impairment, and Alzheimer's disease", *Journal of Neural Transmission*, Vol. 117, No. 2, pp. 217–225, 2010.

150. Schouten, T. M., M. Koini, F. de Vos, S. Seiler, J. van der Grond, A. Lechner, A. Hafkemeijer, C. Möller, R. Schmidt, M. de Rooij *et al.*, "Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease", *NeuroImage: Clinical*, Vol. 11, pp. 46–51, 2016.

151. Damoiseaux, J. S. and M. D. Greicius, "Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity", *Brain Structure and Function*, Vol. 213, No. 6, pp. 525–533, 2009.

152. Jack Jr, C. R., D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish *et al.*, "NIA-

AA research framework: toward a biological definition of Alzheimer's disease", *Alzheimer's & Dementia*, Vol. 14, No. 4, pp. 535–562, 2018.

153. Che, M., A. Cichocki and Y. Wei, "Neural networks for computing best rank-one approximations of tensors and its applications", *Neurocomputing*, Vol. 267, pp. 114–133, 2017.

154. Che, M. and Y. Wei, "Randomized algorithms for the approximations of Tucker and the tensor train decompositions", *Advances in Computational Mathematics*, Vol. 45, pp. 395–428, 2019.

155. Che, M., Y. Wei and H. Yan, "The computation of low multilinear rank approximations of tensors via power scheme and random projection", *SIAM Journal on Matrix Analysis and Applications*, Vol. 41, pp. 605–636, 2020.

156. Kolda, T. G., *Multilinear operators for higher-order decompositions*, Tech. rep., Citeseer, 2006.

APPENDIX A: Tensor Algebra

Tensor factorization techniques can be understood well if the algebraic knowledge is obtained. Therefore it is necessary to give detail about tensor multiplication and norm definitions since those information will be used extremely in the algorithms given in this thesis, details can be found in [156].

A.1. Matrix Products

Matrix products and their properties will be explained in this section since they will be useful in later to understand tensor decomposition algebra.

A.1.1. Kronecker Product

The Kronecker product of matrices $A \in \mathbb{R}^{I\times J}$ and $B \in \mathbb{R}^{K\times L}$ is denoted by $A \otimes B$ and the result is of size $(IK \times JL)$. In addition, it can be defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1j}B \\ a_{21}B & a_{22}B & \cdots & a_{2j}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}B & a_{I2}B & \cdots & a_{IJ}B \end{bmatrix}$$

Proposition A.1. *Let $A \in \mathbb{R}^{I\times J}$, $B \in \mathbb{R}^{K\times L}$, $C \in \mathbb{R}^{J\times M}$, and $D \in \mathbb{R}^{L\times T}$ Then,*

(i) $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$,
(ii) $(A \otimes B)^{\dagger} = A^{\dagger} \otimes B^{\dagger}$.

A.1.2. Khatri-Rao Product

The Khatri-Rao product of matrices $A \in \mathbb{R}^{I\times J}$ and $B \in \mathbb{R}^{K\times J}$ is denoted by $A \odot B$ and the result is of size $(IK \times J)$. It is a columnwise Kronecker product, and it can be defined as

$$A \odot B = \begin{bmatrix} a_{11}b_{:1} & a_{12}b_{:2} & \cdots & a_{1j}b_{:j} \\ a_{21}b_{:1} & a_{22}b_{:2} & \cdots & a_{2j}b_{:j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}b_{:1} & a_{I2}b_{:2} & \cdots & a_{IJ}b_{:j} \end{bmatrix}$$

Proposition A.2. *Let $A \in \mathbb{R}^{I\times L}$, $B \in \mathbb{R}^{J\times L}$,and $C \in \mathbb{R}^{K\times L}$. Then,*

(i) $A \odot B \odot C = (A \odot B) \odot C = A \odot (B \odot C)$,
(ii) $(A \odot B)^\top (A \odot B) = A^\top A * B^\top B$,
(iii) $(A \odot B)^\dagger = ((A^\top A * B^\top B))^\dagger (A \odot B)^\top$

A.1.3. Hadamard Product

The Hadamard product of matrices $A \in \mathbb{R}^{I\times J}$ and $B \in \mathbb{R}^{I\times J}$ is denoted by $A * B$ and the result is of size $(I\text{x}J)$. It is an element-wise matrix product, and it can be defined as

$$A * B = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1J}b_{1J} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2J}b_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}b_{I1} & a_{I2}b_{I2} & \cdots & a_{IJ}b_{IJ} \end{bmatrix}$$

A.2. Matricization of a Tensor

It is possible to arrange a tensor as a matrix. Assuming that $X \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, and the set of dimensions $\mathrm{R} = \{r_1, \cdots, r_L\}$, $\mathrm{C} = \{c_1, \cdots, c_M\}$ represent rows and columns. Finally, the matricized tensor can be denoted by

$$X_{\mathrm{R}\times\mathrm{C}} \in \mathbb{R}^{J \times K} \text{with } J = \prod_{n \in \mathrm{R}} I_n \text{ and } K = \prod_{n \in \mathrm{C}} I_n \tag{A.1}$$

Details of the formulation can be found in [156]. The special case of this matricization, called the *n-mode matricization*, is mostly used in this thesis. $X_{(n)}$ notation is used to describe matricized tensor with respect to n^{th} dimension. Elements of matricized tensor (i_n, j) can be found as:

$$j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^{N} (i_k - 1) J_k \quad with \quad J_k = \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m \tag{A.2}$$

Proposition A.3. *Let $Y \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ be an N-way tensor, and $\mathrm{N} = 1, \cdots, N$.*

(i) If $A \in \mathbb{R}^{I \times J_n}$. Then,
$X = Y \times_n A \Leftrightarrow X_{(n)} = A Y_{(n)}$

(ii) Let $A \in \mathbb{R}^{I_n \times J_n}$ for all $n \in \mathrm{N}$. If $\mathrm{R} = \{r_1, \cdots, r_L\}$, $\mathrm{C} = \{c_1, \cdots, c_M\}$, then,
$X = Y \times_1 A^{(1)} \times_2 A^{(2)} \cdots \times_N A^{(N)} \Leftrightarrow$
$X_{(\mathrm{R}\times\mathrm{C})} = \left(A^{(r_L)} \otimes \cdots \otimes A^{(r_1)}\right) Y_{(\mathrm{R}\times\mathrm{C})} \left(A^{(c_M)} \otimes \cdots \otimes A^{(c_1)}\right)^T$

(iii) As a result, if $A \in \mathbb{R}^{I_n x J_n}$ for all $n \in \mathrm{N}$, for any specific $n \in \mathrm{N}$, matricized version becomes:
$X = Y \times_1 A^{(1)} \times_2 A^{(1)} \cdots \times_N A^{(N)} \Leftrightarrow$
$X_{(n)} = A^{(n)} Y_{(n)} \left(A^{(N)} \otimes \cdots \otimes A^{(n+1)} \otimes A^{(n-1)} \otimes \cdots \otimes A^{(1)}\right)^T$

A.3. Tensor Multiplication

Assuming that $X \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $U \in \mathbb{R}^{J \times I_n}$. Then, *n-mode product* of X and U is denoted by $X \times_n U$ and the size becomes $I_1 \times \cdots I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$. It can be written element-wise as:

$$(X \times_n U)_{i_1 \ldots i_{n-1} j i_{n+1} \ldots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \ldots i_N} u_{j i_n}$$

Proposition A.4. *Let $Y \in \mathbb{R}^{J_1 \times J_2 \times \ldots \times J_N}$ be an N-way tensor.*

(i) Given matrices $A \in \mathbb{R}^{I_m \times J_m}$ and $B \in \mathbb{R}^{I_n \times J_n}$,

$$Y \times_m A \times_n B = (Y \times_m A) \times_n B = (Y \times_n B) \times_m A \quad (m \neq n)$$

(ii) Given matrices $A \in \mathbb{R}^{I \times J_n}$ and $B \in \mathbb{R}^{K \times I}$,

$$Y \times_n A \times_n B = Y \times_n (BA)$$

APPENDIX B: Preprocessing

B.1. Tortoise Settings

Table B.1. Tortoise settings [87].

Parameter Name	**Explanation**	**Chosen Option**
Effect	Specifies which undesirable effects will be corrected	motion_distortion
DWI_B0_registration	Perform motion and eddy distortion correction on DWI data	on
smoothing	Smoothing prior to registration.	off
automatic_noise_computation	Automatically compute noise	on
verbose	Verbose level	6
nbins	Number of bins. Used to compute normalized mutual information	48
img_resample_res	Final diffusion weighted image output resolution (value in mm)	1.5, 1.5, 1.5
structural_output_res	Final structural image output resolution (value in mm)	1.5, 1.5, 1.5
mask_fsl	Obtain bounding box based on FSL mask.	on
ITK_structural_registration	ITKs registration pipeline	on
initial_global_search	Initial rigid body global search	off
upsampling	Upsampling options (all: upsampling in all directions)	all
upsampling_factor	Upsampling factor	2

Table B.1. Tortoise settings [87]. (cont.)

Parameter Name	**Explanation**	**Chosen Option**
upsampling_method	Upsampling method	bicubic
scale_signal_factor	Scale diffusion intensity	1
cropped_FOV	Desired FOV. Applied to template image during cropping (value in mm)	200, 240, 190
bspline_EPI_correction	Perform EPI distortion correction with B-Spline registration.	off
eddy_correction_for_B0_to_structural	Activates the second order Eddy current transformations	on
image_format	The output structural and template images will be in the chosen format	NIFTI
output_endianism	Output Endianism.	big
write_interpolated_variance	Write the noise variance image	on
compute_corr_matrix	Use the correlation matrix	off
start_at_step	Start at step	0
stop_at_step	Stop at step.	5
keep_step4_output	Keep step 4 outputs	off
keep_intermediate_data	Keep ALL the intermediate, generated data.	on

B.2. Registration Script

```
#!/usr/bin/python
import sys
import gzip
import os
in_fold = sys.argv[1]
pathList = in_fold.split("/")
Pat_Name = pathList[len(pathList)-1]
reslice_ref=in_fold+"/T1w/"+"T1w_acpc_dc_restore.nii.gz"
import nipype.interfaces.freesurfer
from nipype.interfaces.freesurfer import ReconAll
from nipype.interfaces.freesurfer import MRIConvert
#-----------------------------------------------
# Convert .mgz to .nii format with reference image
main_folder = in_fold+"/Parcellation/"+Pat_Name
input_file = main_folder + "/mri/aparc.a2009s+aseg.mgz"
output_file = main_folder+"_PreRegParc.nii"
mc = MRIConvert()
mc.inputs.in_file=input_file
mc.inputs.out_type ="nii"
mc.inputs.out_file = output_file
mc.inputs.reslice_like=reslice_ref
mc.inputs.resample_type='nearest'
result = mc.run()
print "conversion completed"
#-----------------------------------------------
# HCP_Pre is registered to T1 which is output of Tortoise
processed_folder = in_fold+"/Processed/"+Pat_Name
input_file = main_folder+"/mri"+"/T1w_hires.nii.gz"
ref_image = processed_folder+"_dMRI_DMC_REORIENTATION.nii"
output_file = processed_folder+"_registeredT1.nii"
```

```
output_mat_file = processed_folder+"_regT1toTortT1.mat"
import nipype.interfaces.fsl as fsl
print "Applying transformation to High Res. T1: Started";
flt = fsl.FLIRT()
flt.inputs.in_file = input_file
flt.inputs.reference =ref_image
flt.inputs.out_file = output_file
flt.inputs.output_type = "NIFTI"
flt.inputs.out_matrix_file = output_mat_file
result = flt.run()
print "Applying transformation to High Res. T1: Completed";
#-----------------------------------------------
# Using Transformation Matrix, Parcellation of 1mm T1 is
# registered to T1 which is output of Tortoise
input_file = processed_folder+"_PreRegParc.nii"
output_file = processed_folder+".par.nii"
ref_image = processed_folder+"_dMRI_DMC_REORIENTATION.nii"
input_mat_file= processed_folder+"_regT1toTortT1.mat"
import nipype.interfaces.fsl as fsl
print "Applying Transformation Matrix : Started"
applyxfm_t2 = fsl.ApplyXFM()
applyxfm_t2.inputs.in_file = input_file
applyxfm_t2.inputs.in_matrix_file = input_mat_file
applyxfm_t2.inputs.out_file = output_file
applyxfm_t2.inputs.reference = ref_image
applyxfm_t2.inputs.interp='nearestneighbour'
applyxfm_t2.inputs.apply_xfm = True
applyxfm_t2.inputs.output_type='NIFTI'
applyxfm_t2 = applyxfm_t2.run()
print "Applying Transformation Matrix : Completed"
#-----------------------------------------------
```

```
#Using Transform Matrix, T1 Brain is downsampled to 1.5 mm
TXb_fold = in_fold+"/T1w/"
input_file = TXb_fold+"T1w_acpc_dc_restore_brain.nii.gz"
output_file= processed_folder+"_T1Brain.nii"
ref_image = processed_folder+"_dMRI_DMC_REORIENTATION.nii"
input_mat_file= processed_folder+"_regT1toTortT1.mat"
import nipype.interfaces.fsl as fsl
print "Applying Transform Matrix to T1 Brain: Started"
applyxfm_t2 = fsl.ApplyXFM()
applyxfm_t2.inputs.in_file = input_file
applyxfm_t2.inputs.in_matrix_file = input_mat_file
applyxfm_t2.inputs.out_file = output_file
applyxfm_t2.inputs.reference = ref_image
applyxfm_t2.inputs.interp='nearestneighbour'
applyxfm_t2.inputs.apply_xfm = True
applyxfm_t2.inputs.output_type='NIFTI'
applyxfm_t2 = applyxfm_t2.run()
print "Applying Transform Matrix to T1 Brain: Completed"
#-----------------------------------------------------
#Using Transform Matrix, T2 Brain is downsampled to 1.5 mm
input_file = TXb_fold+"T2w_acpc_dc_restore_brain.nii.gz"
output_file = processed_folder+"_T2Brain.nii"
ref_image = processed_folder+"_dMRI_DMCstructural.nii"
input_mat_file= processed_folder+"_regT1toTortT1.mat"
import nipype.interfaces.fsl as fsl
print input_file_path_T1Parc
print output_file_path_T1Parc
print "Applying Transform Matrix to T2 Brain: Started"
applyxfm_t2 = fsl.ApplyXFM()
applyxfm_t2.inputs.in_file = input_file
applyxfm_t2.inputs.in_matrix_file = input_mat_file
```

```
applyxfm_t2.inputs.out_file = output_file
applyxfm_t2.inputs.reference = ref_image
applyxfm_t2.inputs.interp='nearestneighbour'
applyxfm_t2.inputs.apply_xfm = True
applyxfm_t2.inputs.output_type='NIFTI'
applyxfm_t2 = applyxfm_t2.run()
print "Applying Transform Matrix to T2 Brain: Completed"
```

B.3. Segmentation Script

```
#!/usr/bin/python
import sys
import os
import gzip
import shutil
cwd = os.getcwd()
input_folder = sys.argv[1]
proc_folder = input_folder+"/Processed/"
#----------------------------------------
shutil.copy("b_seg.py", proc_folder+"b_seg.py");
os.chdir(proc_folder)
pathList = input_folder.split("/")
Patient_Name = pathList[len(pathList)-1]
input_file = proc_folder+Patient_Name+".t1.nii"
import nipype.interfaces.fsl as fsl
# Segmentation (FAST)
print "Segmentation: Started";
fastr = fsl.FAST()
fastr.inputs.in_files = input_file
fastr.inputs.out_basename = Patient_Name
fastr.inputs.output_type = 'NIFTI'
```

```
result = fastr.run()
print "Segmentation: Completed"
#----------------------------------------
os.remove("b_seg.py")
os.chdir(cwd)
output_path = proc_folder+Patient_Name+"_seg.nii"
base = os.path.splitext(output_path)[0]
os.rename(output_path, proc_folder+Patient_Name+".seg.nii")
```

B.4. Masking Script

```
#!/usr/bin/python

import sys
import gzip

in_fold = sys.argv[1]
pathList = in_fold.split("/")
Pat_Name = pathList[len(pathList)-1]
#----------------------------------------
main_folder = in_fold+"/Parcellation/"+Pat_Name
processed_folder = in_fold+"/Processed/"+Pat_Name
input_file = main_folder+"_dMRI_DMC_REORIENTATION.nii"
output_file = processed_folder+".t1.nii"
mask_file_brain = processed_folder+"_T1Brain.nii"
import nipype.interfaces.fsl as fsl
print "Brain Extraction for T1: Started";
applymask_dwi = fsl.ApplyMask()
applymask_dwi.inputs.in_file = input_file
applymask_dwi.inputs.mask_file = mask_file_brain
applymask_dwi.inputs.out_file = output_file
```

```
applymask_dwi.inputs.output_type = 'NIFTI'
applymask_dwi = applymask_dwi.run()
print "Brain Extraction for T1: Completed";
#-----------------------------------------
# T2, DWI Masking
print "Masking Part Begins"
import nipype.interfaces.fsl as fsl
import nipype.interfaces.fsl.maths
import os
pro_folder = in_fold+"/Processed"
mask_file_brain = pro_folder+"_T1Brain.nii"
#search for files
for files in os.listdir(pro_folder):
        if files.endswith("_dMRI_DMC_RAW_4D.nii"):
                dwi_name = files
                dwi_path = os.path.join(pro_folder, dwi_name)
                new_name = Pat_Name + ".dwi.nii.gz"
                masked_dwi = os.path.join(pro_folder, new_name)
        elif files.endswith("_DMCstructural.nii"):
                t2_name = files
                t2_path = os.path.join(pro_folder, t2_name)
                new_name = Pat_Name + ".t2.nii"
                masked_t2 = os.path.join(pro_folder, new_name)
if len(dwi_path) == 0:
        print "DWI IMAGE IS NOT FOUND"
        sys.exit()
if len(NONMASKED_T2_PATH) == 0:
        print "T2 IMAGE IS NOT FOUND"
        sys.exit()

print "Masking: Started"
```

```
applymask_dwi = fsl.ApplyMask()
applymask_dwi.inputs.in_file = dwi_path
applymask_dwi.inputs.mask_file = mask_file_brain
applymask_dwi.inputs.out_file = masked_dwi
applymask_dwi.inputs.output_type = 'NIFTI_GZ'
applymask_dwi = applymask_dwi.run()

applymask_t2 = fsl.ApplyMask()
applymask_t2.inputs.in_file = t2_path
applymask_t2.inputs.mask_file = mask_file_brain
applymask_t2.inputs.out_file = masked_t2
applymask_t2.inputs.output_type = 'NIFTI'
applymask_t2 = applymask_t2.run()
print "Masking: Completed"
```

B.5. Preprocessing fMRI Script

```
import os
import subprocess
import sys
import string
import random
import shutil
import nibabel as nib
import nipype.interfaces.fsl as fsl
from time import gmtime, strftime

def id_generator(size=6,
        chars=string.ascii_uppercase + string.digits):
return ''.join(random.choice(chars) for _ in range(size))
in_fold = sys.argv[1]
```

```
pathList = in_fold.split("/")
pat_name = pathList[len(pathList)-1]
FMRI_FILE = ""
T1_FILE = ""

#-----------------------------------------------
for files in os.listdir(in_fold):
    if files == "RAW":
        currentPath = os.path.join(in_fold, files)
        for f in os.listdir(currentPath):
            if f.find("_fMRI.nii.gz") != -1:
                FMRI_FILE = os.path.join(currentPath, f)
    elif files == "Processed":
        currentPath = os.path.join(in_fold, files)
        for f in os.listdir(currentPath):
            if f.find(".t1.nii") != -1:
                T1_FILE = os.path.join(currentPath, f)
                break
if FMRI_FILE == "" :
        print "No raw FMRI image found!"
        sys.exit()

if T1_FILE == "" :
        print "No T1 image found!"
        sys.exit()
#-----------------------------------------------
times = strftime("%d%m%Y_%H%M%S", gmtime())
out_p=os.path.join(in_fold, "preprocessing_output_"+times)
n= "$FSLDIR/bin/fslnvols " + FMRI_FILE
proc=subprocess.Popen(n,stdout=subprocess.PIPE,shell=True)
(output, err) = proc.communicate()
```

```
p_status = proc.wait()
numVols = int(output)
k="$FSLDIR/bin/fslval " + FMRI_FILE + " pixdim4"
proc=subprocess.Popen(k,stdout=subprocess.PIPE,shell=True)
(output, err) = proc.communicate()
p_status = proc.wait()
TR = float(output)
s= "$FSLDIR/bin/fslval " + FMRI_FILE + " dim1"
proc=subprocess.Popen(s,stdout=subprocess.PIPE,shell=True)
(output, err) = proc.communicate()
p_status = proc.wait()
size1 = int(output)
s= "$FSLDIR/bin/fslval " + FMRI_FILE + " dim2"
proc=subprocess.Popen(s,stdout=subprocess.PIPE,shell=True)
(output, err) = proc.communicate()
p_status = proc.wait()
size2 = int(output)
s= "$FSLDIR/bin/fslval " + FMRI_FILE + " dim3"
proc=subprocess.Popen(s,stdout=subprocess.PIPE,shell=True)
(output, err) = proc.communicate()
p_status = proc.wait()
size3 = int(output)
s = "$FSLDIR/bin/fslval " + FMRI_FILE + " dim4"
proc=subprocess.Popen(s,stdout=subprocess.PIPE,shell=True)
(output, err) = proc.communicate()
p_status = proc.wait()
size4 = int(output)
numVoxels = size1*size2*size3*size4
#-----------------------------------------------
# CHANGE T1 HEADER IF XYZT FIELDS ARE EMPTY
img = nib.load(T1_FILE)
```

```
hdr = img.header
hdr['xyzt_units'] = 10

FSF_FILE = "./brainet_preprocessing_reg.fsf"
randomName = id_generator(7)
TEMP_FSF_FILENAME = randomName + ".fsf"
TEMP_FSF_PATH = "/tmp/" + TEMP_FSF_FILENAME
os.system("cp brainet_preprocessing_reg.fsf /tmp/")
command1 = "mv /tmp/brainet_preprocessing_reg.fsf "
renameCommand = command1 + TEMP_FSF_PATH
os.system(renameCommand)

with open(FSF_FILE,'r') as old, open(TEMP_FSF_PATH, 'w') as new:
        for line in old:
                if line.startswith("set feat_files"):
                        newLine = "set feat_files(1) \"" +
                        FMRI_FILE.split(".")[0] + "\"\n"
                        new.write(newLine)
                elif line.startswith("set fmri(outputdir)"):
                        newLine = "set fmri(outputdir) \"" +
                        out_p + "\"\n"
                        new.write(newLine)
                elif line.startswith("set fmri(totalVoxels)"):
                        newLine = "set fmri(totalVoxels) " +
                        str(numVoxels) + "\n"
                        new.write(newLine)
                elif line.startswith("set fmri(tr)"):
                        newLine = "set fmri(tr) " +
                        str(TR) + "\n"
                        new.write(newLine)
                elif line.startswith("set fmri(npts)"):
```

```
                    newLine = "set fmri(npts) " +
                    str(numVols) + "\n"
                    new.write(newLine)
                elif line.startswith("set highres_files(1)"):
                    newLine = "set highres_files(1) \""  +
                    T1_FILE + "\"\n"
                    new.write(newLine)
                else:
                    new.write(line)
#-----------------------------------------
print TEMP_FSF_PATH
runFeatCommand = "feat " + TEMP_FSF_PATH
os.system(runFeatCommand)
f= ".feat"
out1 = "filtered_func_data.nii.gz"
out2 = "reg/example_func2highres.nii.gz"
out3 = "reg/example_func2highres.mat"
OUTPUT_FMRI_FILE = os.path.join(out_p + f,out1)
OUTPUT_REF_FILE  = os.path.join(out_p + f, out2)
OUTPUT_MATRIX_FILE = os.path.join(out_p + f, out3)
splittedName = T1_FILE.split("/")
new_na = "_preprocessed_fmri.fmri.nii.gz"
splittedName[len(splittedName)-1] = pat_name + new_na
OUTPUT_REGISTERED_FMRI_FILE = "/".join(splittedName)
flt = fsl.FLIRT()
flt.inputs.in_file = OUTPUT_FMRI_FILE
flt.inputs.reference = OUTPUT_REF_FILE
flt.inputs.out_file = OUTPUT_REGISTERED_FMRI_FILE
flt.inputs.in_matrix_file = OUTPUT_MATRIX_FILE
flt.inputs.apply_xfm = True
result = flt.run()
```

B.6. Functional MRI Settings

Table B.2. Settings value of fMRI preprocessing used in script.

Name	Value	Name	Value	Name	Value	Name	Value	Name	Value
conpic_real.1	1	con_real1.2	0	level	1	regstandard_dof	12	brain_thresh	10
conname_real.1	""	conpic_orig.1	1	analysis	1	regstandard_nonlinear_yn	0	featwatcher_yn	1
shape1	0	conname_orig.1	""	help_yn	1	reginitial_highres_yn	0	sscleanup_yn	0
deriv_yn1	1	con_orig1.1	1	stats_yn	0	reginitial_highres_search	90	inputtype	2
paradigm_hp	100	bgimage	1	npts	200	reginitial_highres_dof	3	regunwarp_yn	0
ncopeinputs	0	conmask1_1	0	ndelete	5	regstandard_search	90	signallossthresh	10
confoundevs	0	tsplot_yn	1	mc	1	reghighres_search	90	prewhiten_yn	1
evtitle1	""	con_mode_old	orig	te	35	reghighres_dof	BBR	motionevs	0
poststats_yn	0	init_highres	""	st	5	regstandard_nonlinear_warpres	10	motionevsbeta	""
tempfilt_yn1	1	init_standard	""	sh_yn	0	alternateReference_yn	0	scriptevsbeta	""
convolve_phase1	0	overwrite_yn	0	noise	0.66	conmask_zerothresh_yn	0	reghighres_yn	1
regstandard_yn	0	gammasigma1	3	dwell	0.7	init_initial_highres	""	rendertype	1
zmax	8	gammadelay1	6	zmin	2	alternative_mask	""	filtering_yn	1
inmelodic	0	con_real1.1	1	st_file	""	tr	3.000001	version	6
relative_yn	0	temphp_yn	1	phase1	0	evs_real	2	convolve1	2
tagfirst	1	templp_yn	0	stop1	-1	evs_vox	0	threshmask	""

Table B.2. Settings value of fMRI preprocessing used in script. (cont.)

Name	**Value**	**Name**	**Value**	**Name**	**Value**	**Name**	**Value**	**Name**	**Value**
multiple	1	melodic_yn	0	ortho1.0	0	ncon_orig	1	thresh	3
noisear	0.34	skip1	0	ortho1.1	0	ncon_real	1	prob_thresh	0.05
critical_z	5.3	off1	30	perfsub_yn	0	nftests_orig	0	z_thresh	2.3
bet_yn	1	on1	30	evs_orig	1	nftests_real	0	zdisplay	0
smooth	6	unwarp_dir	y-	mixed_yn	2	constcol	0	con_mode	orig
norm_yn	1	robust_yn	0	regstandard	/usr/share/fsl/5.0/data/standard/MNI152_T1_2mm_brain				

B.7. Parcellation Node Names

Table B.3. List of anatomical parcellations of Destrieux Atlas, and corresponding sub-networks defined by Yeo et al. [97, 98].

Id	Short name	Long name	Yeo et al. 7-Network
1	G_and_S_frontomargin	Fronto-marginal gyrus and sulcus	Fparietal
2	G_and_S_occipital_inf	Inferior occipital gyrus and sulcus	Visual
3	G_and_S_paracentral	Paracentral lobule and sulcus	Som&Aud
4	G_and_S_subcentral	Subcentral gyrus and sulci	Som&Aud
5	G_and_S_transv_frontopol	Transverse frontopolar gyri and sulci	Def. Mode
6	G_and_S_cingul-Ant	Anterior part of the cingulate gyrus and sulcus	Def. Mode
7	G_and_S_cingul-Mid-Ant	Middle-anterior part of the cingulate gyrus and sulcus	Salience
8	G_and_S_cingul-Mid-Post	Middle-posterior part of the cingulate gyrus and sulcus	Salience, Som&Aud
9	G_cingul-Post-dorsal	Posterior-dorsal part of the cingulate gyrus	Def. Mode
10	G_cingul-Post-ventral	Posterior-ventral part of the cingulate gyrus	Def. Mode
11	G_cuneus	Cuneus	Visual
12	G_front_inf-Opercular	Opercular part of the inferior frontal gyrus	Salience, Fparietal
13	G_front_inf-Orbital	Orbital part of the inferior frontal gyrus	Def. Mode
14	G_front_inf-Triangul	Triangular part of the inferior frontal gyrus	Fparietal, Def. Mode
15	G_front_middle	Middle frontal gyrus	Fparietal
16	G_front_sup	Superior frontal gyrus	Def. Mode

Table B.3. List of anatomical parcellations of Destrieux Atlas, and corresponding sub-networks defined by Yeo et al. [97, 98]. (cont.)

Id	**Short name**	**Long name**	**Yeo et al. 7-Network**
17	G_Ins_lg_and_S_cent_ins	Long insular gyrus and central sulcus of the insula	Som&Aud , Salience
18	G_insular_short	Short insular gyri	Salience
19	G_occipital_middle	Middle occipital gyrus	Visual
20	G_occipital_sup	Superior occipital gyrus	Visual
21	G_oc-temp_lat-fusifor	Lateral occipito-temporal gyrus	Visual
22	G_oc-temp_med-Lingual	Lingual gyrus, ligual part of the medial occipito-temporal gyrus	Visual
23	G_oc-temp_med-Parahip	Parahippocampal gyrus, of the medial occipito-temporal gyrus	Limbic
24	G_orbital	Orbital gyri	Limbic, Fparietal, Def. Mode
25	G_pariet_inf-Angular	Angular gyrus	Def. Mode
26	G_pariet_inf-Supramar	Supramarginal gyrus	Salience
27	G_parietal_sup	Superior parietal lobule	Dors. Att.
28	G_postcentral	Postcentral gyrus	Som&Aud
29	G_precentral	Precentral gyrus	Som&Aud
30	G_precuneus	Precuneus (medial part of P1)	Dors. Att., Def. Mode
31	G_rectus	Straight gyrus, Gyrus rectus	Limbic
32	G_subcallosal	Subcallosal area, subcallosal gyrus	Limbic
33	G_temp_sup-G_T_transv	Anterior transverse temporal gyrus (of Heschl)	Som&Aud

Table B.3. List of anatomical parcellations of Destrieux Atlas, and corresponding sub-networks defined by Yeo et al. [97, 98]. (cont.)

Id	**Short name**	**Long name**	**Yeo et al. 7-Network**
34	G_temp_sup-Lateral	Lateral aspect of the superior temporal gyrus	Som&Aud
35	G_temp_sup-Plan_polar	Planum polare of the superior temporal gyrus	Som&Aud, Limbic
36	G_temp_sup-Plan_tempo	Planum temporale of the superior temporal gyrus	Som&Aud, Salience
37	G_temporal_inf	Inferior temporal gyrus	Dors. Att. , Limbic
38	G_temporal_middle	Middle temporal gyrus	Def. Mode
39	Lat_Fis-ant-Horizont	Horizontal ramus of the anterior segment of the lateral sulcus	Salience
40	Lat_Fis-ant-Vertical	Vertical ramus of the anterior segment of the lateral sulcus	Salience, Def. Mode
41	Lat_Fis-post	Posterior ramus of the lateral sulcus	Som&Aud
42	Pole_occipital	Occipital pole	Visual
43	Pole_temporal	Temporal pole	Limbic
44	S_calcarine	Calcarine sulcus	Visual
45	S_central	Central sulcus	Som&Aud
46	S_cingul-Marginalis	Marginal branch of the cingulate sulcus	Salience
47	S_circular_insula_ant	Anterior segment of the circular sulcus of the insula	Fparietal, Def. Mode
48	S_circular_insula_inf	Inferior segment of the circular sulcus of the insula	Salience
49	S_circular_insula_sup	Superior segment of the circular sulcus of the insula	Salience
50	S_collat_transv_ant	Anterior transverse collateral sulcus	Limbic

Table B.3. List of anatomical parcellations of Destrieux Atlas, and corresponding sub-networks defined by Yeo et al. [97, 98]. (cont.)

Id	**Short name**	**Long name**	**Yeo et al. 7-Network**
51	S_collat_transv_post	Posterior transverse collateral sulcus	Visual
52	S_front_inf	Inferior frontal sulcus	Fparietal
53	S_front_middle	Middle frontal sulcus	Fparietal
54	S_front_sup	Superior frontal sulcus	Fparietal, Def. Mode
55	S_interm_prim-Jensen	Sulcus intermedius primus	Fparietal
56	S_intrapariet_and_P_trans	Intraparietal sulcus and transverse parietal sulci	Dors. Att.
57	S_oc_middle_and_Lunatus	Middle occipital sulcus and lunatus sulcus	Visual
58	S_oc_sup_and_transversal	Superior occipital sulcus and transverse occipital sulcus	Visual
59	S_occipital_ant	Anterior occipital sulcus and preoccipital notch	Visual, Dors. Att.
60	S_oc-temp_lat	Lateral occipito-temporal sulcus	Dors. Att.
61	S_oc-temp_med_and_Lingual	Medial occipito-temporal sulcus and lingual sulcus	Visual
62	S_orbital_lateral	Lateral orbital sulcus	Fparietal
63	S_orbital_med-olfact	Medial orbital sulcus	Limbic
64	S_orbital-H_Shaped	Orbital sulci	Limbic, Fparietal
65	S_parieto_occipital	Parieto-occipital sulcus	Visual
66	S_pericallosal	Pericallosal sulcus	Fparietal, Def. Mode
67	S_postcentral	Postcentral sulcus	Som&Aud , Dors. Att.

Table B.3. List of anatomical parcellations of Destrieux Atlas, and corresponding sub-networks defined by Yeo et al. [97, 98]. (cont.)

Id	**Short name**	**Long name**	**Yeo et al. 7-Network**
68	S_precentral-inf-part	Inferior part of the precentral sulcus	Dors. Att.
69	S_precentral-sup-part	Superior part of the precentral sulcus	Dors. Att.
70	S_suborbital	Suborbital sulcus	Limbic, Def. Mode
71	S_subparietal	Subparietal sulcus	Def. Mode
72	S_temporal_inf	Inferior temporal sulcus	Fparietal, Def. Mode
73	S_temporal_sup	Superior temporal sulcus	Def. Mode
74	S_temporal_transverse	Transverse temporal sulcus	Som&Aud

APPENDIX C: CP Decomposition

The CP decomposition factorizes the tensor into sum of rank-one tensors. For example given a third-order tensor $X \in \mathbb{R}^{I \times J \times K}$ we want to write it as

$$X \approx \sum_{r=1}^{R} a_r \circ b_r \circ c_r \tag{C.1}$$

where R is a positive integer and $a_r \in \mathbb{R}^I$, $b_r \in \mathbb{R}^J$, and $c_r \in \mathbb{R}^K$. This expression can also be written in element-wise as

$$x_{ijk} \approx \sum_{r=1}^{R} a_r(i) b_r(j) c_r(k) \tag{C.2}$$

If we write factor matrices as a combination of rank-one components ,i.e., $A = [a_1, a_2 \cdots a_R]$, $B = [b_1, b_2 \cdots b_R]$, and $C = [c_1, c_2 \cdots c_R]$ where $A \in \mathbb{R}^{I \times R}$, $B \in \mathbb{R}^{J \times R}$, and $C \in \mathbb{R}^{K \times R}$. CP decomposition of a third-order tensor is illustrated in Figure C.1. In order to find

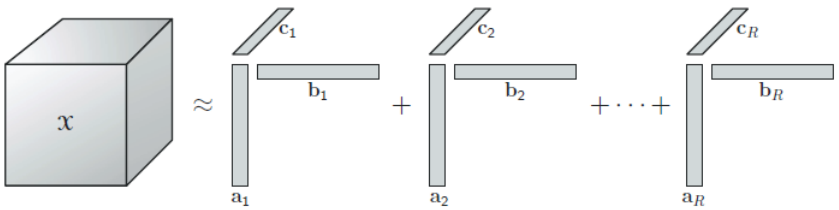

Figure C.1. CP decomposition of a three-way array [75]

factor matrices several algorithms can be used such that ALS, GD, MB-GD etc.. In all algorithms, factor matrices should be initialized at first which can be done randomly or using GEVD. In the next sub-section algorithms mentioned above are discussed in detail.

C.1. Alternating Least Square

Lets assume that X is an Nth-order tensor. Then CP decomposition of $X \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ will be

$$X_{i_1 i_2 \cdots i_N} \approx \sum_{r=1}^{R} \lambda_r a^{(1)}_{i_1 r} a^{(2)}_{i_2 r} \cdots \cdots a^{(N)}_{i_N r} \tag{C.3}$$

If we write factor matrices as a combination of rank-one components ,i.e., $A^{(1)} = [a^{(1)}_1, a^{(1)}_2 \cdots a^{(1)}_R]$, $A^{(2)} = [a^{(2)}_1, a^{(2)}_2 \cdots a^{(2)}_R]$, and all the way up to N as $A^{(N)} = [a^{(N)}_1, a^{(N)}_2 \cdots a^{(N)}_R]$ where $A^{(1)} \in \mathbb{R}^{I_1 \times R}$, $A^{(2)} \in \mathbb{R}^{I_2 \times R}$, and in the end $A^{(N)} \in \mathbb{R}^{I_N \times R}$. In addition, assume that $\lambda \in \mathbb{R}^{R \times R \times R}$ tensor with λ_r's along the superdiagonal. Then as written in section A.3 , this decomposition can be written as :

$$X \approx \lambda \times_1 A^{(1)} \times_2 A^{(2)} \cdots \times_N A^{(N)} \tag{C.4}$$

By using the knowledge in *n-mode matricization* as in section A.2 and propositon A.3(iii) it can be written as :

$$X_{(n)} = A^{(n)} \lambda_{(n)} \big(A^{(N)} \otimes \cdots \otimes A^{(n+1)} \otimes A^{(n-1)} \otimes \cdots \otimes A^{(1)}\big)^T \tag{C.5}$$

Because of the structure of λ ,e.i, being superdiagonal, Kronecker products of the factor matrices turn into Khatri-Rao product in the next formula as :

$$X_{(n)} = A^{(n)} \Lambda \big(A^{(N)} \odot \cdots \odot A^{(n+1)} \odot A^{(n-1)} \odot \cdots \odot A^{(1)}\big)^T \tag{C.6}$$

where $\Lambda = diag(\lambda_r)$. CP decomposition is an optimization problem which can be modeled as :

$$\min_{\hat{X}} \parallel X - \hat{X} \parallel \text{ with } \hat{X} = \lambda \times_1 A^{(1)} \times_2 A^{(2)} \cdots \times_N A^{(N)} \tag{C.7}$$

It is easier to compute CP decomposition by using the alternating least squares (ALS) algorithm when $X \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$, and a desired R are given. The idea behind the ALS is each factor matrix is solved in a turn, when the other factor matrices are fixed until some convergence criteria is satisfied. Then the optimization problem for one turn become :

$$\min_{A^{(n)}} \parallel X_{(n)} - A^{(n)} \Lambda \big(A^{(N)} \odot \cdots \odot A^{(n+1)} \odot A^{(n-1)} \odot \cdots \odot A^{(1)} \big)^T \parallel \tag{C.8}$$

which is a classic least squares problem. Then the optimal solution is computed as :

$$\hat{A}^{(n)} = X_{(n)} \Big[\big(A^{(N)} \odot \cdots \odot A^{(n+1)} \odot A^{(n-1)} \odot \cdots \odot A^{(1)} \big)^T \Big]^{\dagger} \tag{C.9}$$

where $\hat{A} = A\Lambda$. Since the Khatri-Rao product pseudoinverse has a special form, it can be rewritten as :

$$\hat{A}^{(n)} = X_{(n)} \big(A^{(N)} \odot \cdots \odot A^{(n+1)} \odot A^{(n-1)} \odot \cdots \odot A^{(1)} \big) V^{\dagger} \tag{C.10}$$

where

$$V = (A^{(N)\top} A^{(N)}) * \cdots (A^{(n+1)\top} A^{(n+1)}) * (A^{(n-1)\top} A^{(n-1)}) * \cdots * (A^{(1)\top} A^{(1)}) \tag{C.11}$$

C.2. Gradient Descent

Gradient descent is an iterative optimization algorithm which consists of finding gradient of the loss function and defining a proper step size for the update of parameters. Therefore, gradient of the loss function has to be calculated at first. Lets examine this example for a 3-way tensor $X \in \mathbb{R}^{I \times J \times K}$ and factor matrices with

INPUT: $X \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}, R > 0$

OUTPUT: $A^{(n)} \in \mathbb{R}^{I_n \times R}$ for $n = 1, \cdots, N$

initialize $A^{(n)} \in \mathbb{R}^{I_n \times R}$ for $n = 1, \cdots, N$

repeat

for $n = 1, \cdots, N$ **do**

$V \leftarrow (A^{(N)\top} A^{(N)}) * \cdots A^{(n+1)\top} A^{(n+1)} * A^{(n-1)\top} A^{(n-1)} * \cdots * (A^{(1)\top} A^{(1)})$

$A^{(n)} \leftarrow X_{(n)} \big(A^{(N)} \odot \cdots \odot A^{(n+1)} \odot A^{(n-1)} \odot \cdots \odot A^{(1)}\big) V^{\dagger}$

normalize columns of $A^{(n)}$ (storing norm as λ)

end for

until some convergence criteria is satisfied

Figure C.2. CP Decomposition with ALS

$A \in \mathbb{R}^{I \times R}, B \in \mathbb{R}^{J \times R}, C \in \mathbb{R}^{K \times R}$, gradient with respect to $A, B,$ and C will be

$$L(A, B, C) = ||X_{(1)} - A(C \odot B)^{\top}||_F^2 \tag{C.12}$$

$$\frac{\partial L(A, B, C)}{\partial A} = -2 \left(X_{(1)} - A(C \odot B)^{\top}\right) C \odot B \tag{C.13}$$

$$\frac{\partial L(A, B, C)}{\partial B} = -2 \left(X_{(2)} - B(C \odot A)^{\top}\right) C \odot A \tag{C.14}$$

$$\frac{\partial L(A, B, C)}{\partial C} = -2 \left(X_{(3)} - C(B \odot A)^{\top}\right) B \odot A \tag{C.15}$$

In this method factor matrix A is updated by using the following equation :

$$A = A - \eta \frac{\partial L(A, B, C)}{\partial A} \tag{C.16}$$

where η is the step size. Other factor matrices are updated in a similar way. Whole algorithm can be seen below.

C.3. Mini-Batch Gradient Descent

Let's assume that we have N way tensor $X \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, and we would like to decompose it as we have done in the upper section. Unlike the gradient descent

INPUT: $X \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}, R > 0, \eta$

OUTPUT: $A^{(n)} \in \mathbb{R}^{I_n \times R}$ for $n = 1, \cdots, N$

initialize $A^{(n)} \in \mathbb{R}^{I_n \times R}$ for $n = 1, \cdots, N$

repeat

for $n = 1, \cdots, N$ **do**

$V = \left(A^{(N)} \odot \cdots \odot A^{(n+1)} \odot A^{(n-1)} \odot \cdots \odot A^{(1)}\right)$

$Grad = (X_{(n)} - A^{(n)} V^\top) V$

$A^{(n)} = A^{(n)} + \eta Grad$

end for

until some convergence criteria is satisfied

Figure C.3. CP Decomposition with GD

method, stochastic gradient descent does not use all the data in order to calculate gradient, just randomly picked data. Lets examine this example for a 3-way tensor $X \in \mathbb{R}^{I \times J \times K}$ and factor matrices with $A \in \mathbb{R}^{I \times R}, B \in \mathbb{R}^{J \times R}, C \in \mathbb{R}^{K \times R}$, gradient with respect to $A, B,$ and C at one point will be

$$\frac{\partial}{\partial A_{i,r}} \left(X_{i,j,k} - \sum_{r=1}^{R} A_{i,r} B_{j,r} C_{k,r} \right)^2 = -2 \left(X_{i,j,k} - \sum_{\hat{r}=1}^{R} A_{i,\hat{r}} B_{j,\hat{r}} C_{k,\hat{r}} \right) C_{k,r} * B_{j,r} \quad \text{(C.17)}$$

$$\frac{\partial}{\partial C_{k,r}} \left(X_{i,j,k} - \sum_{r=1}^{R} A_{i,r} B_{j,r} C_{k,r} \right)^2 = -2 \left(X_{i,j,k} - \sum_{\hat{r}=1}^{R} A_{i,\hat{r}} B_{j,\hat{r}} C_{k,\hat{r}} \right) B_{j,r} * A_{i,r} \quad \text{(C.18)}$$

$$\frac{\partial}{\partial B_{j,r}} \left(X_{i,j,k} - \sum_{r=1}^{R} A_{i,r} B_{j,r} C_{k,r} \right)^2 = -2 \left(X_{i,j,k} - \sum_{\hat{r}=1}^{R} A_{i,\hat{r}} B_{j,\hat{r}} C_{k,\hat{r}} \right) C_{k,r} * A_{i,r} \quad \text{(C.19)}$$

In this method factor matrix A is updated by using the following equation :

$$A_{i,r} = A_{i,r} - \eta \frac{\partial \left(X_{i,j,k} - \sum_{r=1}^{R} A_{i,r} B_{j,r} C_{k,r} \right)^2}{\partial A_{i,r}} \quad \text{(C.20)}$$

where η is the step size. Other factor matrices are updated in a similar way. We have used mini-batch gradient descent algorithm which uses b examples in each iteration

to update factor matrices. Therefore, randomly picked b examples have been used to calculate gradient. Whole algorithm of mini-batch gradient descent can be seen below.

INPUT: $X \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, $R > 0$, $batch_{size} > 0$, M is the elements of X

OUTPUT: $A^{(n)} \in \mathbb{R}^{I_n \times R}$ for $n = 1, \cdots, N$

initialize $A^{(n)} \in \mathbb{R}^{I_n \times R}$ for $n = 1, \cdots, N$

X=random.shuffle(X)

$Batch = getbatches(X, batch_{size})$

repeat

Batch = random.shuffle(Batch)

for $i = 1, \cdots, M/batch_{size}$ **do**

for $n = 1, \cdots, N$ **do**

$Grad = evaluate\_grad(X, Batch(i), A^{(n)})$

$A^{(n)} = A^{(n)} + \eta Grad$

end for

end for

until some convergence criteria is satisfied

Figure C.4. CP Decomposition with Mini-Batch GD

APPENDIX D: Further Experiment Results

D.1. Uni-Modal B-Tensor

Table D.1. P-values of the MANOVA analysis for B_f and B_s with a set of Q values.

		Q=3	Q=5	Q=7	Q=9	Q=11	Q=13	Q=15
B_f	e_{ij}^{Pe}	9.59E-20	1.39E-20	4.02E-21	6.50E-22	4.41E-20	3.16E-20	1.12E-18
	e_{ij}^{Pc}	3.15E-41	1.24E-38	9.22E-35	4.13E-40	1.31E-37	2.60E-35	1.62E-35
B_s	e_{ij}^{n}	8.06E-14	1.43E-15	4.18E-17	2.33E-14	4.03E-15	7.83E-14	1.09E-12
	e_{ij}^{l}	1.84E-12	9.13E-11	2.15E-10	1.11E-10	2.90E-10	2.08E-10	1.35E-09
	e_{ij}^{a}	1.25E-11	4.46E-11	1.13E-12	7.87E-12	1.22E-09	1.67E-10	6.02E-11

Table D.2. P-values of the Fisher-Pitman permutation test for B_f and B_s with a set of Q values.

		Q=3	Q=5	Q=7	Q=9	Q=11	Q=13	Q=15
B_f	e_{ij}^{Pe}	2.51E-08	4.84E-10	2.03E-11	2.06E-11	1.80E-10	3.13E-09	3.36E-10
	e_{ij}^{Pc}	1.26E-09	1.82E-09	3.07E-11	6.40E-10	2.31E-11	1.99E-10	3.31E-10
B_s	e_{ij}^{n}	3.77E-07	6.10E-09	8.58E-08	1.05E-07	1.07E-07	2.64E-07	2.62E-07
	e_{ij}^{l}	4.79E-08	2.16E-06	2.37E-06	7.64E-06	3.32E-06	3.10E-07	1.25E-07
	e_{ij}^{a}	1.12E-06	8.29E-06	1.37E-06	4.85E-06	1.08E-06	7.28E-09	4.16E-08

Table D.3. Confusion matrices for B_f and B_s with a set of Q values. S, M and A stand for SCI, MCI, and AD.

		Q=3			Q=5			Q=7			Q=9			Q=11			Q=13			Q=15		
		Predicted																				
		S	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**
e_{ij}^{Pe}	**SCI**	13	0	1	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0
	MCI	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0
	AD	1	0	7	0	0	8	0	0	8	0	0	8	1	1	6	1	2	5	1	2	5
e_{ij}^{Pc}	**SCI**	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0
	MCI	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0
	AD	0	0	8	1	0	7	1	0	7	1	1	6	2	0	6	2	0	6	2	0	6
e_{ij}^{n}	**SCI**	10	3	1	12	1	1	11	2	1	11	2	1	11	2	1	11	2	1	11	2	1
	MCI	2	23	0	1	24	0	2	23	0	2	23	0	2	23	0	3	22	0	3	22	0
	AD	1	0	7	1	0	7	1	0	7	2	0	6	2	0	6	2	0	6	2	0	6
e_{ij}^{l}	**SCI**	12	1	1	11	2	1	11	1	2	10	2	2	10	2	2	11	2	1	12	0	2
	MCI	2	23	0	2	23	0	2	23	0	3	22	0	3	22	0	1	24	0	1	24	0
	AD	2	0	6	4	0	4	5	0	3	4	0	4	3	0	5	3	0	5	4	0	4
e_{ij}^{a}	**SCI**	10	3	1	9	4	1	11	2	1	10	3	1	10	3	1	11	2	1	11	2	1
	MCI	3	22	0	3	22	0	4	21	0	3	22	0	2	23	0	2	23	0	2	23	0
	AD	2	0	6	2	0	6	2	0	6	3	0	5	3	0	5	1	0	7	1	0	7

D.2. Multi-Modal B-Tensor

Table D.4. P-values of the MANOVA analysis for B_{cm} with a set of Q values.

B_{cm}	Q=3	Q=5	Q=7	Q=9	Q=qq	Q=13	Q=15
$e_{ij}^{Pe}+e_{ij}^{n}$	8.32E-28	5.27E-27	3.00E-25	1.03E-26	3.23E-25	4.99E-24	1.64E-22
$e_{ij}^{Pe}+e_{ij}^{l}$	7.60E-20	7.76E-22	1.06E-23	5.99E-25	6.30E-22	2.10E-21	2.39E-23
$e_{ij}^{Pe}+e_{ij}^{a}$	2.63E-20	6.63E-23	3.28E-24	2.09E-23	1.27E-22	2.67E-21	6.50E-20
$e_{ij}^{Pc}+e_{ij}^{n}$	3.73E-45	3.65E-43	2.21E-44	2.25E-41	1.72E-38	2.74E-35	4.76E-31
$e_{ij}^{Pc}+e_{ij}^{l}$	1.22E-45	6.02E-42	3.33E-44	1.05E-40	7.80E-37	9.17E-37	5.91E-33
$e_{ij}^{Pc}+e_{ij}^{a}$	1.41E-46	1.17E-41	3.90E-39	8.60E-40	3.31E-39	4.86E-36	4.98E-33

Table D.5. P-values of the Fisher-Pitman permutation test for B_{cm} with a set of Q values.

B_{cm}	Q=3	Q=5	Q=7	Q=9	Q=qq	Q=13	Q=15
$e_{ij}^{Pe}+e_{ij}^{n}$	2.85E-10	5.62E-12	1.15E-11	5.56E-12	1.59E-11	2.71E-11	8.71E-12
$e_{ij}^{Pe}+e_{ij}^{l}$	7.16E-09	8.05E-11	1.26E-11	1.07E-11	1.98E-10	2.28E-09	9.53E-12
$e_{ij}^{Pe}+e_{ij}^{a}$	3.69E-09	1.51E-10	4.00E-12	4.57E-12	3.71E-11	1.03E-11	5.42E-11
$e_{ij}^{Pc}+e_{ij}^{n}$	1.26E-10	3.30E-11	2.61E-11	1.55E-11	2.07E-12	2.80E-11	8.93E-12
$e_{ij}^{Pc}+e_{ij}^{l}$	2.32E-09	5.02E-11	4.96E-11	7.40E-11	2.13E-10	6.55E-12	8.30E-11
$e_{ij}^{Pc}+e_{ij}^{a}$	9.01E-10	5.86E-11	1.11E-10	1.91E-10	1.60E-10	1.48E-10	8.43E-11

Table D.6. Confusion matrices for B_{cm} with a set of Q values. S, M and A stand for SCI, MCI, and AD.

		Q=3			Q=5			Q=7			Q=9			Q=11			Q=13			Q=15		
		Predicted																				
		S	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**
$e_{ij}^{Pe}+e_{ij}^{n}$	**SCI**	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0
	MCI	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0
	AD	0	0	8	0	0	8	0	0	8	0	0	8	0	0	8	1	0	7	1	0	7
$e_{ij}^{Pe}+e_{ij}^{l}$	**SCI**	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0
	MCI	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0
	AD	0	0	8	0	0	8	0	1	7	1	0	7	2	1	5	0	0	8	2	0	6
$e_{ij}^{Pe}+e_{ij}^{a}$	**SCI**	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0
	MCI	0	25	0	0	25	0	0	25	0	0	25	0	1	24	0	1	24	0	1	24	0
	AD	0	0	8	0	0	8	1	0	7	1	0	7	1	0	7	1	0	7	1	0	7
$e_{ij}^{Pc}+e_{ij}^{n}$	**SCI**	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0
	MCI	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0
	AD	0	0	8	0	0	8	0	0	8	0	0	8	0	0	8	0	0	8	1	0	7
$e_{ij}^{Pc}+e_{ij}^{l}$	**SCI**	13	0	1	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0
	MCI	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0
	AD	1	0	7	0	0	8	0	0	8	0	0	8	2	0	6	2	1	5	1	1	6

Table D.6.Confusion matrices for B_{cm} with a set of Q values. S, M and A stand for SCI, MCI, and AD. (cont.)

		Q=3			Q=5			Q=7			Q=9			Q=11			Q=13			Q=15		
		Predicted																				
		S	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**	**S**	**M**	**A**
$e_{ij}^{Pc}+e_{ij}^{a}$	**SCI**	13	0	1	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0
	MCI	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0	0	25	0
	AD	1	0	7	1	0	7	0	0	8	0	0	8	1	0	7	0	0	8	1	0	7