Multi-scale Graph Principal Component Analysis for Connectomics

by

Steven Winter

Department of Statistical Science
Duke University

Date: ____________________________

Approved:

______________________________

David Dunson, Advisor

______________________________

Sayan Mukherjee

______________________________

Peter Hoff

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Statistical Science in the Graduate School of Duke University

2021
ABSTRACT

Multi-scale Graph Principal Component Analysis for Connectomics

by

Steven Winter

Department of Statistical Science
Duke University

Date: __________________________

Approved:

__________________________________________________________________________

David Dunson, Advisor

__________________________________________________________________________

Sayan Mukherjee

__________________________________________________________________________

Peter Hoff

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Statistical Science in the Graduate School of Duke University

2021
Abstract

In brain connectomics, it is common to divide the cortical surface into discrete regions of interest (ROIs), and then to use these regions to induce a graph. Nodes correspond to regions of interest and edges encode summaries of the strength of connections between pairs of regions. These spatial weighted graphs are often reduced to adjacency matrices, which are then used as inputs to downstream statistical analysis. The structure of these adjacency matrices depends critically on the chosen parcellation, with finer resolutions producing unique spare patterns. Consequently, both the available methods of analysis and the conclusions from analysis depend heavily on the chosen parcellation. To solve this problem we develop a multi-scale graph factorization model, which links together scale-specific factorizations through a common set of individual-specific latent factors. These scores combine information across from different parcellations to produce a single interpretable summary of an individuals brain structure. We develop a simple, efficient algorithm, and illustrate substantial advantages over comparable single-scale methods in both simulations and analyses of the Human Connectome Project dataset.
Acknowledgements

This work is done in collaboration with David Dunson (Duke University) and Zhengwu Zhang (University of North Carolina, Chapel Hill). I would like to express my sincere gratitude to David and Zhengwu for their enduring support. I thank the other members of my thesis committee, Sayan Mukherjee (Duke University) and Peter Hoff (Duke University), for many helpful conversations and exceptional class instructions. Finally, I thank my parents and my Grandmother for always encouraging my curiosity and giving me every opportunity to succeed.

This research is partially supported by grant 1R01MH118927-01 of the United States National Institute of Health (NIH). Data collection and sharing for this project was provided by the MGH-USC Human Connectome Project (HCP; Principal Investigators: Bruce Rosen, M.D., Ph.D., Arthur W. Toga, Ph.D., Van J. Weeden, MD). HCP funding was provided by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH), and the National Institute of Neurological Disorders and Stroke (NINDS). HCP data are disseminated by the Laboratory of Neuro Imaging at the University of Southern California.
## Contents

**Abstract**                                                                 iv  
**Acknowledgements**                                                        v  
**List of Figures**                                                          vii  
**1 Introduction**                                                           1  
**2 Data**                                                                   4  
  2.1 The Human Connectome Project                                             4  
  2.1.1 Connectome Extraction                                                 4  
  2.1.2 Trait Identification                                                  7  
**3 Methods**                                                                8  
  3.1 Modelling                                                               8  
    3.1.1 Motivation                                                          8  
    3.1.2 Implementation                                                      9  
    3.1.3 Connectome Specific Interpretations                                  14  
  3.2 Results                                                                15  
    3.2.1 Simulations                                                         15  
    3.2.2 Discovering Group Connectome Differences                            17  
    3.2.3 Improved Trait Predictions                                          20  
    3.2.4 Connectome Hypothesis Testing via Latent Factors                    22  
**4 Conclusions**                                                            24  
**Bibliography**                                                             26
List of Figures

2.1 Visualizing surface-enhanced tractography. .................... 6

2.2 Examples of brain networks under different parcellations. .... 7

3.1 Estimating simulated latent factors. ............................ 16

3.2 Changes in connectivity with increased lifetime binge drinking. 18

3.3 Out-of-sample trait predictions. ............................... 21

3.4 Detecting trait differences. .................................... 22
Chapter 1

Introduction

Recent improvements in neuroimaging have allowed scientists to obtain extremely high-resolution images of the brain, which has facilitated many novel methods of statistical analysis. In particular, there has been rapidly increasing interest in understanding how the brain connectome varies across individuals, both randomly and in relation to individual-specific factors, ranging from genotype to traits and behaviours [FZB13], [MAAB+16], [GCR+16], [ZAZD19]. A mathematically rigorous and computationally tractable representation of the human connectome is a necessary precursor for all such analysis. In the literature, the most common approach is to encode the connectome as a weighted adjacency matrix where each entry measures the strength of connection between a pair of regions of interest (ROIs).

In functional connectomics, blood flow correlation is a standard measure of connection strength, and it is common to study sequences of networks indexed by time, e.g. as a patient watches a movie. In structural connectomics, connectivity strength is measured using summaries of white matter (WM) fiber tracks between regions. Endpoint-related summaries include raw streamline counts and measures of connected surface area; diffusion-related summaries include statistics of fractional anisotropy; geometry-related summaries include average streamline length and cluster configuration. Some authors have considered the functional and structural connectome jointly [CMSO+20]. Though our methods apply to any set of adjacency matrices, regardless of which summary of which connectome they encode, we will focus primarily on streamline counts.

Raw imaging data is many steps removed from streamline count adjacency ma-
trices. First, each subject’s white matter tracts must be reconstructed from diffusion imagery and all tracts must be aligned to a common coordinate system \([\text{FG13}, \text{GWDD14}, \text{TSW}^{+15}; \text{GCR}^{+16}]\); only then can a parcellation be applied. Each parcellation corresponds to a pre-defined set of regions of interest - for example, the Desikan or Destrieux atlas \([\text{DSF}^{+06}, \text{DFDH10}]\). Weighted adjacency matrices are produced by computing summaries between regions after applying the chosen atlas. Many statistical tools are available once the raw imaging data has been processed into networks, ranging from analyzing topological summaries of the network structure \([\text{BB11}, \text{KFWG14}]\), to comparing each connection separately with multiple testing adjustment \([\text{ZFB10}, \text{MFGG}^{+13}]\), or more sophisticated methods that act on the entire adjacency matrices. The latter class of methods include recently-developed approaches for replicated network analysis ranging from tensor PCA \([\text{ZAZD19}]\) to Bayesian hierarchical models \([\text{ABFX08}, \text{Rod12}, \text{MDMC}^{+14}, \text{DDV17}]\).

All of the above approaches are critically dependent on the chosen atlas. An overwhelming number of atlases are available: many are based on different notions of “physiologically relevant” regions (e.g. the Deskian atlas), but ultimately these choices are arbitrary. The number of regions, hence the scale of the parcellation, can have a tremendous influence on the structure of the induced matrices. A large number of regions will produce massive, sparse adjacency matrices which demand different methods of analysis than coarse alternatives. From a clinical perspective, it is far from ideal to have inconsistent diagnoses of neurodegenerative diseases based on the size of a chosen atlas. Our key insight is that induced networks across different scales can all be viewed as containing incomplete information about a single underlying connectome. As such, it is reasonable to use multiple scales to estimate a single, scale-independent set of latent factors that summarizes an individual’s entire brain network. The goal of this document is to develop an easily interpretable and computationally
tractable model for producing such latent factors through a multi-scale graph PCA factorization.

Our approach builds upon Tensor-Network PCA [ZAZDI19], which proposed a Canonical Polyadic decomposition for single-scale semi-symmetric brain tensors. Their method returns a vector of individual-specific latent factors which can be used for visualization as well as inputs to downstream statistical models. They also analyze the Human Connectome Project (HCP) data and highlight several novel relationships between white matter structure and real world behaviours and traits. Unfortunately, their method produces different results for every parcellation. Our multi-scale generalization dramatically reduces sensitivity to the underlying parcellation through the aforementioned common latent factors.

Conceptually, our methodology has links to model averaging and ensemble learning methods [DMFV+13, HISVT15, KJKSI19]. These approaches typically focus on combining several models for highly accurate black-box prediction; instead we focus on sharing data between scales to produce a single set of latent factors that allow for highly interpretable analysis. For example, our method allows an analyst to map inference performed with the latent factors back on to the level of connectomes, as in Figure 3.2. We test our algorithm on simulated data as well as the HCP dataset and ultimately demonstrate that sharing data across scales highlights novel connectivity structures, improves error correction, and improves out-of-sample prediction. Code is freely available for public use at https://github.com/szwinter/MultiGraph_PCA.
Chapter 2

Data

2.1 The Human Connectome Project

All data analyzed in this document come from the Human Connectome Project (HCP), available via ConnectomeDB. The HCP contains 1221 high quality diffusion MRI (dMRI) scans from 1065 unique healthy subjects - each scan is composed of 6 runs summarizing three different gradient tables, with each table acquired once with right-to-left and then left-to-right phase encoding polarities. A typical gradient table contains 90 diffusion weighting directions plus 6 b0 acquisitions distributed throughout each run. Each run is composed of three shells (b = 1000, 2000 and 3000 s/mm²) with an equal number of acquisitions in each shell. All scans were performed with the same customized 3T Connectome Scanner. Basic processing yields final images with isotropic voxel size of 1.25 mm³. Please see GSW+13 and VSB+13 for precise details.

2.1.1 Connectome Extraction

We focus exclusively on surface-based structural connectivity (SC), wherein white matter streamlines are reconstructed using diffusion imagery and the projected to the cortical surface. Once a parcellation is defined on the surface, it is easy to compute the induced adjacency matrices by counting the number of streamlines starting in one region and ending in another. Traditional tractography methods fail to reliably project streamlines to the cortical surface for two primary reasons. First, diffusion in
gray matter (GM) obfuscates diffusion in WM, leading to unreliable reconstructions near the GM-WM boundary [GWDD14]. Second, the geometry of the cortical surface inherently leads to tract biases in areas of high curvature - for example, gyral blades [SODGDI18]. This latter point is exacerbated by lower relative spatial resolutions in these curved regions. A recent algorithm called surface-enhanced tractography (SET) [SODGDI18] solves both of these problems by using a surface flow technique to propagate WM streamlines through the GM-WM interface, entirely bypassing the need to use poor quality GM diffusion imagery.

SET was to build streamlines for 118 HCP subjects during the development of [CMSO+20]; these are the subjects under consideration here. For these data, flow is initiated from the white surface at $N_s = 3 \times 10^6$ discrete locations and propagated for a total of $t = 75$ time units; these numbers come from optimizing reproducibility. Figure 2.1 illustrates some results from the SET pipeline, where (a) shows the initial white matter segments constructed by SET and (b) shows the final tractography result. The final output of this algorithm is a brain network corresponding to a parcellation of many small triangles. Computing adjacency matrices under a new parcellation (e.g. the Desikan atlas) amounts to collapsing adjacent triangles.

In practice, we find our multi-scale graph method performs best when the parcellations contain mostly different information instead of mostly redundant information. It is thus important to obtain a simple, reproducible method for generating very different parcellations - we propose to do this by dividing regions of the Desikan atlas into equally sized subregions. For simplicity we focus on custom parcellations parameterized by two natural numbers, $(\ell, r)$, indicating the number of new regions created from each Desikan region in the left and right hemispheres, respectively. Parcellations with $\ell = r$ will be referred to as symmetric and parcellations with $\ell \neq r$ as asymmetric. For example, $(1, 1)$ corresponds exactly to the Desikan at-
Figure 2.1: Tractography results from SET. (a) shows the surface flow to one of the white surfaces, and (b) shows the final tractography with the reconstructed fanning structure near the white surfaces.

Las; (2, 2) splits every Desikan region into equally sized subregions; (2, 4) splits every region in the left hemisphere into two equally sized subregions and every region in the right hemisphere into four equally sized subregions, resulting in a graph with $34 \times 2 + 34 \times 4 = 204$ nodes for the cortical surfaces. Examples of these graphs are visualized in Figure 2.2.

Notice parcellations (1, 1) and (2, 2) show visually similar structures, hence they contain mostly redundant information about the connectome. Conversely, parcellations (1, 1) and (2, 4) look quite different - in particular, parcellation (2, 4) appears to be more sensitive to inter-hemisphere connections. In our experiments we find the greatest gains from multi-scale methods occur when both symmetric and asymmetric parcellations are used. This is not surprising given the high degree of cross-hemisphere correlation at lower resolutions. Some existing single-scale methods exploit this symmetry for more efficient dimensionality reduction and insightful exploratory analysis [vKK20]. In our case, we hypothesize the synthesis of disjoint views of the connectome allows us to inject more information about an individual’s brain into the latent
Figure 2.2: Multiple adjacency matrix representations of a random individual’s connectome. Notice the structural similarities between (1, 1) and (2, 2), contrasted with (2, 4).

factors than, say, several symmetric parcellations.

2.1.2 Trait Identification

The HCP contains rich trait measurement data for every subject in the study, quantifying cognitive, emotional, motor, sensory function and other domains. Most of the measurements were obtained in accordance with the NIH Toolbox for Assessment of Neurological and Behavioral function [GWH+13]. Five additional traits were measured in HCP, including visual processing; personality and adaptive function; delay discounting; fluid intelligence; and behavioural measures of emotional processing. In this document, we consider 175 traits for each subject in our analysis, spanning eight categories: cognitive ability; motor skills; substance use; psychiatric and life function; senses; emotional ability; personality; and general health. Each trait is measured as a binary, ordinal, or continuous variable. A detailed description of these 175 traits can be found in [BBH+13].
Chapter 3

Methods

3.1 Modelling

3.1.1 Motivation

Multi-scale graph data for individual $i$ comes as adjacency matrices $\{X_i^{(1)}, \ldots, X_i^{(R)}\}$, where $j = 1, \ldots, R$ indexes different parcellations of the cortical surface. Our goal is to replace these matrices with a real-valued vector of latent factors that summarizes the individual’s brain, and then use this vector to study relationships between brain structure and traits. Of course there are many ways this can be done without multi-scale machinery: for example, one could vectorize the lower triangular part of each matrix within each scale and use the resulting representations as inputs to linear or logistic regression with traits as responses. This would result in $R$ single-scale models which could be averaged or combined with e.g. Super Learner [VdLPH07]. This approach can lead to highly accurate predictions but suffers from a lack of interpretability. Specifically there is no principled method for visualizing ensemble-based inference on the level of connectomes, thus making such approaches sub-optimal for learning neural mechanisms underlying trait differences or partitioning subjects into groups in a clinical setting.

Our multi-scale model facilities interpretability by rephrasing the problem of computing latent factors as a problem of simultaneously approximating multiple graphs.
Our particular $K$-dimensional approximation for individual $i$’s data takes the form

$$X_i^{(j)} = \sum_{h=1}^{K} W_h^{(j)} u_{ih}, \quad j = 1, \ldots, R,$$

where $W_1^{(j)}, \ldots, W_K^{(j)}$ are independent rank 1 adjacency matrices and $u_i = (u_{i1}, \ldots, u_{iK})$ is the row vector we use to summarize an individual’s connectome. Notice the $W$ matrices depend on the scale $j$ but not the individual $i$; conversely the latent factor vector $u_i$ depends on the individual but not the scale. This dichotomy is critical: the graphs $W_1^{(j)}, \ldots, W_K^{(j)}$ encapsulate scale-specific, population level graph structures - in essence, they provide the best $K$ dimensional approximation to an average connectome under parcellation $j$. The latent factors $u_i$ encode scale-independent individual deviations from the population average, hence providing a convenient summary of the unique features of an individual’s connectome. Any inference performed with the latent factors can be mapped back on to the connectome level, allowing for clear, intuitive, and interpretable visualizations such as Figure 3.2.

### 3.1.2 Implementation

The multi-scale graph model presented above is over-parametrized. Each $W_h^{(j)}$ is a rank 1 adjacency matrix and can hence be decomposed as $d_h^{(j)} v_h^{(j)} \odot v_h^{(j)}$, where $d_h^{(j)}$ is a positive constant, $v_h^{(j)}$ is a vector of unit length, and $\odot$ denotes the outer product. We also impose within-scale orthogonality of the $v_h^{(j)}$ to reliably estimate these separate contributors. With these constraints, our multi-scale graph model can be expressed as

$$X_i^{(j)} = \sum_{h=1}^{K} d_h^{(j)} v_h^{(j)} \odot v_h^{(j)} u_{ih}, \quad j = 1, \ldots, R,$$

which has considerably fewer free parameters compared with the original data.
Recall that imposing orthogonality constraints for tensors is much more severe than in the matrix setting \cite{Rob16}; it is reasonable to worry that brain networks are not in the small subspace of semi-symmetric semi-orthogonal tensors. In our experiments on real data we did not observe any material gains from removing these constraints, suggesting that the semi-symmetric semi-orthogonal model is flexible enough for brain networks with mild noise. Nonetheless, the publicly available implementation of our algorithm has an option to turn off orthogonality constraints. Regardless of which constraints are enforced, finding optimal decompositions is non-trivial and necessitates expressing the problem in tensor notation. We adopt the conventions of \cite{KB09}: henceforth scalars will be denoted by $x$; vectors by $x$; matrices by $X$. We will call $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M}$ an $M$-mode tensor; the inner product of two $M$-mode tensors $A, B \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M}$ is

$$\langle A, B \rangle = \sum_{i_1} \sum_{i_2} \ldots \sum_{i_M} a_{i_1,i_2,\ldots,i_M} b_{i_1,i_2,\ldots,i_M}.$$ 

This induces the Frobenius norm via $||\mathcal{X}||_2 = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$. The $n$-mode multiplication of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M}$ with a matrix $A \in \mathbb{R}^{J_n \times I_n}$ (analogously, a vector) is denoted by $\mathcal{X} \times_n A$ and is defined as

$$(\mathcal{X} \times_n A)_{i_1,\ldots,i_{n-1},j,i_{n+1},\ldots,i_M} = \sum_{i_n} x_{i_1,i_2,\ldots,i_M} a_{j,i_n},$$

which is an element in $\mathbb{R}^{I_1 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_M}$.

Concatenating the $N$ brain network matrices within each scale of data produces a semi-symmetric 3-mode tensor $\mathcal{X}(j) \in \mathbb{R}^{P_j \times P_j \times N}$, where $P_j$ is the number of ROIs in the parcellation scheme $j$. For example, the complete collection of streamline count adjacency matrices from the HCP data under the Desikan parcellation would produce
one such tensor of dimension $68 \times 68 \times 1065$, and splitting each Desikan ROI into two would produce another tensor of dimension $136 \times 136 \times 1065$. We can represent our multi-scale graph model at each scale as

$$\mathcal{X}^{(j)} = \sum_{h=1}^{K} d_h^{(j)} v_h^{(j)} \circ v_h^{(j)} \circ u_h, \quad j = 1, \ldots, R.$$  

Constraining the $u_h$ to be unit length allows one to solve each single scale approximation with a greedy, one-at-a-time rank-one algorithm referred to as the tensor power method [All12]. The greedy algorithm results in the basis $\{d_h^{(j)} v_h^{(j)} \circ v_h^{(j)}, h = 1, \ldots, K\}$ that explains the most variation in the observed networks. Multilinear algebra is used to rephrase the greedy rank-one minimization problem as a maximization problem:

$$\max_{\mathcal{X} \times_1 P_{h-1} v_h \times_2 P_{h-1} v_h \times_3 u_h} \quad \text{subject to } ||u_h||_2 = 1, ||v_h||_2 = 1$$

where $P_{h-1} = I - V_{h-1} V_{h-1}^T$ with $V_{h-1} = [v_1, \ldots, v_{h-1}]$. Here the orthogonality constraint in the $v_h$ manifests through the orthogonal projection matrices $P_h$; turning the constraint off is equivalent to replacing each of these with an identity matrix. Obtaining a joint approximation demands that we collapse these $R$ maximization problems into a single objective. We propose that one maximize the induced sum of squares. Specifically:

$$\max_{u_h, v_h^{(1)}, \ldots, v_h^{(R)}} \sum_{j=1}^{R} \left( \mathcal{X}^{(j)} \times_1 P_h^{(j)} v_h^{(j)} \times_2 P_{h-1} v_h^{(j)} \times_3 u_h \right)^2$$

$$\text{subject to } ||u_h||_2 = 1, ||v_h^{(1)}||_2 = 1, \ldots, ||v_h^{(R)}||_2 = 1.$$ 

To solve this maximization problem, blockwise coordinate ascent is sufficient:
iteratively optimizing with respect to the $u$ and $v$ components yields closed form updates:

$$\hat{u}_h | v_h^{(1)}, \ldots, v_h^{(R)} = E_{\text{max}} \left( \sum_{j=1}^{R} \left( A^{(j)} \times_1 P_{h-1}^{(j)} v_{h}^{(j)} \times_2 P_{h-1}^{(j)} v_{h}^{(j)} \right) \right. \left. \left( A^{(j)} \times_1 P_{h-1}^{(j)} v_{h}^{(j)} \times_2 P_{h-1}^{(j)} v_{h}^{(j)} \right)^T \right)$$

$$\hat{v}_{h}^{(j)} | u_h = E_{\text{max}} \left( P_{h-1}^{(j)} (A^{(j)} \times_3 u_h) P_{h-1}^{(j)} \right),$$

where $E_{\text{max}}(A)$ is the eigenvector corresponding to the maximum eigenvalue of the matrix $A$. As the objective is not globally convex, there is the possibility of converging to a local optima. Warm starts can reduce this risk. In particular, we recommend computing the Higher Order Singular Value Decomposition (HOSVD) at each scale to obtain initial estimates of $v^{(j)}$ and using these to find the first $u$ update. We also advocate re-running the optimization routine several times for each rank 1 problem; taking the best result across repeat runs further reduces the risk of selecting a local optima.

The form of $\hat{u}$ is appealing: using a sum of squares penalty results in an update rule that organically combines spectral information from all scales into a single vector. A linear penalty also yields a closed form update, but fails to incorporate any such spectral information. These developments are summarized in Algorithm 1.
Algorithm 1: Multi-scale graph principal components analysis.

Data: Tensors $\mathcal{X}^{(j)} \in \mathbb{R}^{P_j \times P_j \times N}$, $j = 1, \ldots, R$ and a positive integer $K$.

Result: Vectors $d^{(j)} = (d_1^{(j)}, \ldots, d_K^{(j)}) \in \mathbb{R}^K$ and matrices

$$
V^{(j)} = [v_1^{(j)}, \ldots, v_K^{(j)}] \in \mathbb{R}^{P_j \times K},
U = [u_1, \ldots, u_K] \in \mathbb{R}^{N \times K}.
$$

for $j \leftarrow 1$ to $R$ do
  $\hat{X}^{(j)} \leftarrow \mathcal{X}^{(j)}$

for $h \leftarrow 1$ to $K$ do
  initialize $u_h$;
  for $j \leftarrow 1$ to $R$ do
    initialize $v_h^{(j)}$;
    $P_0^{(j)} \leftarrow I_{P_j}$
  repeat
    $\hat{u}_h | v_h^{(1)}, \ldots, v_h^{(R)} \leftarrow E_{\text{max}} \left( \sum_{j=1}^R \left( \mathcal{X}^{(j)} \times_1 P_{h-1}^{(j)} v_h^{(j)} \times_2 P_{h-1}^{(j)} v_h^{(j)} \right) \left( \mathcal{X}^{(j)} \times_1 P_{h-1}^{(j)} v_h^{(j)} \times_2 P_{h-1}^{(j)} v_h^{(j)} \right)^T \right)$;
    for $j \leftarrow 1$ to $R$ do
      $\hat{v}_h^{(j)} | u_h \leftarrow E_{\text{max}} \left( P_{h-1}^{(j)} (\mathcal{X}^{(j)} \times_3 u_h) P_{h-1}^{(j)} \right)$;
  until objective converged;

for $j \leftarrow 1$ to $R$ do
  $V_h^{(j)} \leftarrow [v_1^{(j)}, \ldots, v_h^{(j)}]$;
  $P_h^{(j)} \leftarrow I - V_h^{(j)} V_h^{(j)T}$;
  $d_h^{(j)} \leftarrow \hat{X}^{(j)} \times_1 v_h^{(j)} \times_2 v_h^{(j)} \times_3 u_h$;
  $\hat{X}^{(j)} \leftarrow \hat{X}^{(j)} - d_h^{(j)} \circ v_h^{(j)} \circ u_h$;

Interpretations of algorithm inputs/outputs and comments on connectome specific applications follow.
3.1.3 Connectome Specific Interpretations

Our multi-scale algorithm takes in tensors $\mathcal{X}^{(j)} \in \mathbb{R}^{P_j \times P_j \times N}$ for $j = 1, \ldots, R$, each corresponding to a stack of $N$ single-scale brain network observations. It returns vectors $d^{(j)} = (d^{(j)}_1, \ldots, d^{(j)}_K) \in \mathbb{R}^K$ and matrices $V^{(j)} = [v^{(j)}_1, \ldots, v^{(j)}_K] \in \mathbb{R}^{P_j \times K}$, $U = [u_1, \ldots, u_K] \in \mathbb{R}^{N \times K}$. These are compactly stored as the sets $\{d^{(j)}, V^{(j)}, U\}$, $j = 1, \ldots, R$, which we informally refer to as Kruskal decompositions. The rows of the latent factor matrix $U$ summarize the connectomes of each individual - these are the outputs that can be used as features for statistical analysis. For the $j$th parcellation, the brain network modes $\{v^{(j)}_1 \circ v^{(j)}_1, \ldots, v^{(j)}_K \circ v^{(j)}_K\}$ explain more variation across subjects as $K$ increases.

Our approach has a single tuning parameter $K$, which can be chosen using strategies for choosing the number of components in PCA. If the primary goal is to extract brain principal components predictive of traits of the individual, then choosing $K$ to maximize predictive accuracy out-of-sample is a reasonable strategy. Alternatively, if the primary goal is to obtain a parsimonious representation of the brain connectome that discards limited information, then $K$ can be chosen so that at least some prespecified proportion of the variance is explained \cite{All12}. Formally, the multi-scale cumulative proportion of variance explained (CPVE) by the first $K$ components is given by

$$\text{CPVE} = \min_j \frac{||\mathcal{X}^{(j)} \times_1 P_{V^{(j)}_K} \times_2 P_{U^{(j)}_K}||}{||\mathcal{X}^{(j)}||},$$

where $P_{V^{(j)}_K}$ (resp. $P_{U^{(j)}_K}$) is the projection onto the first $K$ columns of $V^{(j)}$ (resp. $U^{(j)}$).
3.2 Results

Each subsection of the results highlights a different advantage of multi-scale modelling over single-scale alternatives. First, simulations are used to demonstrate improved estimation of latent factors in the presence of serious noise. Following our simulations is an analysis of 118 healthy individuals from the HCP whom were all processed during the early stages of [CMSO+20]. A method for detecting group connectome differences with varying trait values is showcased; in particular, we use multi-scale modelling to detect a relationship between reduced connectivity and increased binge drinking which appears to be unavailable to single-scale methods [Figure 3.2]. We close by demonstrating improved trait predictions and greater sensitivity for detecting connectome differences.

3.2.1 Simulations

We conduct a simulation study demonstrating how multi-scale modelling yields more accurate estimation of latent factors in the presence of noise. The majority of noise in connectomics applications occurs during imaging. Here one encounters radio frequency emissions due to thermal motion inside the brain, general measurement error, and significant head movements [Nol01, GAS18].

Our simulations span two types of graph structures (random and sparse) and two types of noise (random and Rademacher). To begin, adjacency matrices were simulated at scales \{25, 50, 75\}, each representing \(N = 100\) subjects with a true rank of 10. The common latent factors were drawn entrywise from a \(N(0, 1)\) distribution and the network components were drawn entrywise from a Gamma(1, 1) distribution. Sparse tensors were created by thresholding 75\% of the network modes. All matrices were normalized at this stage, and then noise was added. In half the experiments this
was centered normal noise with the standard deviation taken to be one third of the signal range; in the other half of the experiments it was Rademacher noise created by randomly flipping 25% of the graph edges for each subject.

Our metric for the performance of an estimate \( \hat{U} \) of \( U \) is the variance explained, which we define to be \( \| P_{\hat{U}} U \|_2 / \| U \|_2 \). If \( \hat{U} \) is an excellent reconstruction of \( U \) then \( P_{\hat{U}} U \approx U \), so the variance explained should be close to 1. For each \( K = 1, \ldots, 10 \) we recorded the multi-scale variance explained along with the mean variances explained for TN-PCA, HOSVD, and HOOI. This test was repeated 10 times; the average results are reported in Figure 3.1.

Multi-scale modelling results in favorable performance in all of these cases, with the largest gains visible under severe normal noise. This makes intuitive sense: the
more data we have, the more likely we are to be able to differentiate signal from noise. Joint modelling across scales simply leverages more data for each estimate of the latent factors. We also point out that common single-scale methods suffer from identifiability issues in estimating latent factors - that is, repeated runs can produce U estimates with an arbitrary sign, making it difficult to combine estimates across scales. This issue is avoided with multi-scale modelling.

### 3.2.2 Discovering Group Connectome Differences

One of our primary goals is to improve methodology for relating brain structure to other variables measured on the individual, ranging from neuropsychiatric conditions and cognitive traits to exposures. We would like to understand differences in the brain structure of individuals having low and high values of a trait of interest, while allowing for variability in brain structure across individuals in each group.

In this subsection we showcase a method tailored for such applications and then present an example with a specific trait: worst lifetime binge drinking. We use a symmetric and an asymmetric parcellation; combining these two distinct views to the brain leads to discovery of novel edge changes across groups not readily apparent from single-scale methods. Our symmetric parcellation of choice is the Desikan atlas - (1,1) in our notation - and our asymmetric parcellation is (2,4), corresponding to splitting every region in the left hemisphere into two equally sized subregions and every region in the right hemisphere into four equally sized subregions.

Inferences on how connectomes change with traits is performed by finding a direction \( \mathbf{w} \in \mathbb{R}^K \) in the embedding space that is highly correlated with a trait, and then mapping that direction back on to the space of brain networks. If \( \mathbf{Y} \) denotes the centered trait scores and \( \mathbf{U} \) denotes the centered latent factors, then this is equivalent
Figure 3.2: Largest changes in connectivity with increased lifetime binge drinking. Purple nodes correspond to the left hemisphere; red nodes correspond to the right hemisphere. Networks were multiplied by a scale-dependent constant for a more readable legend.
to choosing

\[ w = \arg\max_{||\tilde{w}||_2=1} \tilde{w}^T U^T Y. \]

When the trait is continuous this may be easily done with canonical correlation analysis (CCA) [Hot92]; when the trait is categorical we instead use linear discriminant analysis (LDA) [Fis36]. The vector \( w \) captures how connectomes change with the trait score - we can then map this back onto the brain network via

\[ \Delta^{(j)} = s \sum_{h=1}^{K} d_h^{(j)} v_h^{(j)} \circ v_h^{(j)} w_h, \]

where \( s > 0 \) is a scaling parameter. Defining \( s = w^T U^T Y / (||Uw||_2^2 ||Y||_2^2) \) to be the correlation between the projected scores and the traits allows one to compare networks across traits. Note that by using different \( v^{(j)} \) we inspect the connectome change \( \Delta^{(j)} \) at different scales, however we advocate always presenting network plots on the finest scale possible so as not to inadvertently censor potentially insightful trait-connectome relationships.

Figure 3.2 illustrates this process for the ordinal “worst binge drinking” trait, which takes values from 1 to 7 for males and 1 to 6 for females. A value of 1 indicates no more than three drinks in a day; 2 indicates four to six drinks; 3 indicates seven to nine; 4 indicates ten to twelve; 5 indicates thirteen to fifteen; 6 indicates sixteen to twenty for males and over sixteen for females, and 7 indicates over twenty-one drinks in a single day for males. We treat this trait as continuous so that we may use CCA. Single- and multi-scale latent factors were computed using \( K = 10 \); the resulting \( \Delta \) networks were thresholded to retain the 100 entries with the largest absolute values at different scales. Column (a) and (c) shows the results using the single-scale TN-PCA.
method [ZAZD19] and column (b) and (d) shows the results from our method.

At the coarse scale, we see similar patterns in the results of our method and the single-scaled TN-PCA method: the Δ network within each hemisphere contains mixed positive and negative values; between-hemisphere connections are dominated by negative values, focusing on the frontal cortex band. The results indicate that increased alcohol intake is related to decreased connection strength between most of brain regions, which is consistent with the literature [ZAZD19], [MGK01]. However, we see multi-scale modelling reveals additional negative connectivity changes between the left rostral middle frontal gyrus, superior frontal gyri and the right frontal and right temporal lobes, which are missed by the single-scale alternatives. Importantly, at the higher resolution, our multi-scale (2, 4) plot in panel (d) presents a more detailed picture of the information contained in the connection between ROI pairs compared with the signal scale method. Careful inspection of the connection reduction between the left and right superior frontal ROIs indicates that the reduction in connectivity specifically happens between one half of the left superior frontal gyrus and all four subregions of the right superior frontal gyrus. These four new edges all have different weights, indicating that the connectivity changes are most relevant to specific subregions of the right superior frontal gyrus. One could continue to add finer scales to our multi-scale model to shine more light on this finding.

3.2.3 Improved Trait Predictions

The synthesis of both symmetric and asymmetric graph data yields significant improvements in predictive performance over comparably interpretable single-scale methods. We highlight this using the multi-scale \{(1, 1), (2, 4)\} data from 118 random HCP individuals. These individuals were randomly split into a 70% training set and a 30%
(a) Histograms of MSE improvements. The left plot has one omitted outlier around $-200$.

(b) Highlighting improvements in predicting traits specifically related to alcohol use.

**Figure 3.3:** Visualizing the improved predictions from multi-scale methods. All calculations use the median MSE.

test set; a ridge regression model was then trained for each trait using the first 70 latent factors. Predictions were made on the test set, and the mean squared error (MSE) was recorded for every trait. Overall, we repeated this test 100 times with different splits. Figure 3.3 shows distributions of the relative changes in median MSE, as well as highlights a few specific traits related to alcohol use. Our multi-scale model achieved a lower median MSE for 88.8% of traits when compared to the single-scale
Figure 3.4: Visualizing $p$-values from comparing $K = 20$ latent factors with the MMD test. Horizontal and vertical lines correspond to FDR < 0.05 for single- and multi-scale $p$-values respectively.

(1, 1) model and 73.1% when compared to the single-scale (2, 4) model.

The negative changes in MSE are largely confined to traits within the domains of education, income, dexterity/endurance, delay discounting, and negative affectivity. The positive changes include the majority of traits related to substance use and dependence (alcohol, tobacco, marijuana, and hard drugs), psychiatric function (attention, aggression, anxiety, depression, insomnia, etc), sensory ability (vision, touch, taste, hearing, smell), emotion recognition, measures of healthy relationships, and family medical history.

3.2.4 Connectome Hypothesis Testing via Latent Factors

We can indirectly measure the significance of connectome-trait relationships by performing a hypothesis test for differences in the distribution of latent factors between individuals with high trait values and low trait values. Single-scale models are at a significant theoretical disadvantage to multi-scale models in this context for two main reasons. Firstly, they greatly amplify the number of comparisons being made.
If we have $R$ parcellations and we analyze each trait separately for each parcellation, then we are performing $175 \times R$ hypothesis tests; a multi-scale model leveraging all $R$ parcellations only requires 175 tests. Secondly, $p$-values from single-scale methods cannot be interpreted on the connectome level. Hypothesis tests performed with a single discretization of the cortical surface can only be used for inference related to that discretization; multi-scale modelling frees one from these confines and allows for more general, connectome-level tests.

We used the maximum mean discrepancy (MMD) test to detect differences in the latent factors between individuals with trait values in the upper quartile 25% and those in the lower quartile (groups of $N \in [25, 30]$ for our 118 unique individuals based on data availability and tied values). Figure 3.4 visualizes $p$-values for the multi-scale $\{(2, 4), (4, 2)\}$ model against a single-scale alternative, with horizontal and vertical lines corresponding to significance thresholds after adjusting for a false discovery rate (FDR) [BH95] of no more than 5%.

In particular when $K = 20$ our multi-scale model detects significantly different latent factor distributions for two traits related to fluid intelligence (Penn matrix test overall score and number of skipped questions), two traits related to reading ability (oral reading test scores, both age adjusted and unadjusted), two traits related to working memory (list sorting test scores, both age adjusted and unadjusted), and two traits related to overall psychiatric function (normalized anxiety/depression scores and raw DSM hyperactivity scores). None of these traits were significant under the associated single-scale models.
Chapter 4

Conclusions

In this document we develop a new method for summarizing an individual’s structural connectome using multiple discretizations of the cortical surface. Our proposed approach is tailored for interpretable analysis of relationships between brain structure and individual specific factors. By using data from multiple parcellations simultaneously, we can potentially improve power to detect relationships between connectomes and traits; for example, we observed a significant association between the connectome and binge drinking that was not apparent in single scale analyses. In addition, we demonstrated significantly improved predictive performance over comparable single-scale models, as well as generally more accurate detection of connectome differences and greater robustness to different types of noise. All code is available, ready-to-use, on GitHub.

Our algorithm performed best when we combined both low and high-resolution data. This makes intuitive sense: high-resolution data contain a great deal of information, but analysis is generally complicated by problems such as individual ROI misalignment and sparsity. Low-resolution data contain less information, but individuals are extremely well aligned and the observed networks are not sparse. Anchoring high-resolution data to low-resolution data via common latent factors allows us to leverage information only available at fine scales without being overwhelmed by the aforementioned complications.

Another consistent theme in our experiments was disappointing results with symmetric parcellations; surprisingly, asymmetric parcellations almost always yield better
performance - even when comparing among single-scale models. We believe this indicates that future research on defining optimal brain parcellations is paramount. Perhaps ROIs can be considered as network communities which can be learned from the ultra-high resolution adjacency matrix representation of the connectome. Different parcellations could be found by running e.g. hierarchical community discovery \cite{RB06, LFK09} with different tuning parameters, similarity metrics, and loss functions. One could imagine creating similarity metrics or loss functions that factor in trait information; for example detecting a set of communities with high interconnectivity for individuals with the highest trait values. Such methods could produce a wide variety of “optimal” parcellations.

The above approach may generate many parcellations that are optimal in some sense, but it is unlikely that all of these parcellations are necessary for multi-scale inference. It is therefore also important to develop methods for selecting ideal subsets of multi-scale data. Hypothesis testing provides a simple solution: the MMD test can be applied to see if the latent factors change significantly after adding a new scale. Another relatively simple solution may involve adding weights to each term of the objective function in our model; these weights could then be estimated in some supervised way, with thresholding used to eliminate irrelevant parcellations. Finally, it may be possible to define an influence function in terms of the latent factors, which would provide a clear interpretation of the effect of new parcellations as well as opening the door to analyzing the effects of head movement on inference.
Bibliography

[ABFX08] Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9(Sep):1981–2014, 2008.

[ACM12] William K. Allard, Guangliang Chen, and Mauro Maggioni. Multiscale geometric methods for data sets ii: Geometric multi-resolution analysis. *Applied and Computational Harmonic Analysis*, 32(3):435 – 462, 2012.

[AKM18] Salim Arslan, Sofia Ira Ktena, Antonios Makropoulos, Emma C. Robinson, Daniel Rueckert, and Sarah Parisot. Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *NeuroImage*, 170:5 – 30, 2018.

[All12] Genevera Allen. Sparse higher-order principal components analysis. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pages 27–36, La Palma, Canary Islands, 21–23 Apr 2012. PMLR.

[BB11] Edward T. Bullmore and Danielle S. Bassett. Brain graphs: Graphical models of the human brain connectome. *Annual Review of Clinical Psychology*, 7(1):113–140, 2011.

[BBH+13] Deanna M. Barch, Gregory C. Burgess, Michael P. Harms, Steven E. Petersen, Bradley L. Schlaggar, Maurizio Corbetta, Matthew F. Glasser, Sandra Curtiss, Sachin Dixit, Cindy Feldt, Dan Nolan, Edward Bryant, Tucker Hartley, Owen Footer, James M. Bjork, Russ Poldrack, Steve Smith, Heidi Johansen-Berg, Abraham Z. Snyder, and David C. Van Essen. Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, 80:169 – 189, 2013. Mapping the Connectome.

[BH95] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995.

[BL10] Göran Bergqvist and Erik G Larsson. The higher-order singular value decomposition: Theory and an application [lecture notes]. *IEEE Signal Processing Magazine*, 27(3):151–154, 2010.
[CMSO+20] Martin Cole, Kyle Douglass Murray, Etienne St-Onge, Benjamin Risk, Jianhui Zhong, Giovanni Schifitto, Maxime Descoteaux, and Zhengwu Zhang. Surface-based connectivity integration. *bioRxiv*, 2020.

[DDV17] Daniele Durante, David B Dunson, and Joshua T Vogelstein. Non-parametric Bayes modeling of populations of networks. *Journal of the American Statistical Association*, 112(520):1516–1530, 2017.

[DFDH10] Christophe Destrieux, Bruce Fischl, Anders Dale, and Eric Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53(1):1–15, 2010.

[DMFV+13] Benoit Da Mota, Virgile Fritsch, Gaël Varoquaux, Vincent Frouin, Jean-Baptiste Poline, and Bertrand Thirion. Enhancing the reproducibility of group analysis with randomized brain parcellations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 591–598. Springer, 2013.

[dRvdH13] Marcel A. de Reus and Martijn P. van den Heuvel. The parcellation-based connectome: Limitations and extensions. *NeuroImage*, 80:397–404, 2013.

[DSF+06] Rahul S. Desikan, Florent Ségonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, Anders M. Dale, R. Paul Maguire, Bradley T. Hyman, Marilyn S. Albert, and Ronald J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968 – 980, 2006.

[FG13] Martin A. Frost and Rainer Goebel. Functionally informed cortex based alignment: An integrated approach for whole-cortex macro-anatomical and ROI-based functional alignment. *NeuroImage*, 83:1002 – 1010, 2013.

[Fis36] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.

[FSD99] Bruce Fischl, Martin I Sereno, and Anders M Dale. Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2):195–207, 1999.

[FZB13] Alex Fornito, Andrew Zalesky, and Michael Breakspear. Graph analysis of the human connectome: promise, progress, and pitfalls. *NeuroImage*, 80:426–444, 2013.
Bhawna Goyal, Sunil Agrawal, and BS Sohi. Noise issues prevailing in various types of medical images. *Biomedical & Pharmacology Journal*, 11(3):1227, 2018.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

Matthew F. Glasser, Timothy S. Coalsona, Emma C. Robinson, Carl D. Hacker, John Harwell, Essa Yacoub, et al. A multi-modal parcellation of human cerebral cortex. *NeuroImage*, 536:171 – 178, 2016.

Matthew F. Glasser, Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L. Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R. Polimeni, David C. Van Essen, and Mark Jenkinson. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80:105 – 124, 2013.

Gabriel Girard, Kevin Whittingstall, Rachid Deriche, and Maxime Descoteaux. Towards quantitative connectivity analysis: reducing tractography biases. *NeuroImage*, 98:266 – 278, 2014.

Richard C. Gershon, Molly V. Wagster, Hugh C. Hendrie, Nathan A. Fox, Karon F. Cook, and Cindy J. Nowinski. NIH toolbox for assessment of neurological and behavioral function. *Neurology*, 80(11 Supplement 3):S2–S6, 2013.

Andrés Hoyos-Idrobo, Yannick Schwartz, Gaël Varoquaux, and Bertrand Thirion. Improving sparse recovery on structured images with bagged clustering. In 2015 International Workshop on Pattern Recognition in NeuroImaging, pages 73–76. IEEE, 2015.

Richard A. Harshman and Margaret E. Lundy. Parafac: Parallel factor analysis. *Computational Statistics Data Analysis*, 18(1):39 – 72, 1994.

Harold Hotelling. Relations between two sets of variates. In *Breakthroughs in Statistics*, pages 162–190. Springer, 1992.

Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

Mayuresh S. Korgaonkar, Alex Fornito, Leanne M. Williams, and Stuart M. Grieve. Abnormal structural networks characterize major depressive disorder: a connectome analysis. *Biological Psychiatry*, 76(7):567 – 574, 2014.
[KJKS19] Meenakshi Khosla, Keith Jamison, Amy Kuceyeski, and Mert R. Sabuncu. Ensemble learning with 3D convolutional neural networks for functional connectome-based prediction. *NeuroImage*, 199:651 – 662, 2019.

[LFK09] Andrea Lancichinetti, Santo Fortunato, and János Kertész. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 11(3):033015, 2009.

[MAAB+16] Karla L Miller, Fidel Alfaro-Almagro, Neal K Bangerter, David L Thomas, Essa Yacoub, Junqian Xu, Andreas J Bartsch, Saad Jbabdi, Stamatios N Sotiropoulos, Jesper LR Andersson, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, 19(11):1523, 2016.

[MDMC+14] Oscar Miranda-Dominguez, Brian D Mills, Samuel D Carpenter, Kathleen A Grant, Christopher D Kroenke, Joel T Nigg, and Damien A Fair. Connectotyping: model based fingerprinting of the functional connectome. *PloS One*, 9(11), 2014.

[Mes19] Arnaud Messé. Parcellation influence on the connectivity-based structure–function relationship in the human brain. *Human Brain Mapping*, 2019.

[MFGG+13] Djalel Eddine Meskaldji, Elda Fischi-Gomez, Alessandra Griffa, Patric Hagmann, Stephan Morgenthaler, and Jean-Philippe Thiran. Comparing connectomes across subjects and populations at different scales. *NeuroImage*, 80:416 – 425, 2013.

[MGF+17] Daniel Moyer, Boris A. Gutman, Joshua Faskowitz, Neda Jahanshad, and Paul M. Thompson. Continuous representations of brain connectivity using spatial point processes. *Medical Image Analysis*, 41:32 – 39, 2017.

[MGK01] Hamdy F. Moselhy, George Georgiou, and Ashraf Kahn. Frontal lobe changes in alcoholism: a review of the literature. *Alcohol and Alcoholism*, 36(5):357–368, 09 2001.

[Nol01] Douglas C Noll. A primer on MRI and functional MRI. *Technical Report, University of Michigan*, 2001.

[RB06] Jörg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Physical Review E*, 74(1):016110, 2006.

[Rob16] Elina Robeva. Orthogonal decomposition of symmetric tensors. *SIAM Journal on Matrix Analysis and Applications*, 37(1):86–102, 2016.
Abel Rodríguez. Modeling the dynamics of social networks using Bayesian hierarchical blockmodels. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 5(3):218–234, 2012.

Etienne St-Onge, Alessandro Daducci, Gabriel Girard, and Maxime Descoteaux. Surface-enhanced tractography (SET). *NeuroImage*, 169(November 2017):524–539, 2018.

Bernard N Sheehan and Yousef Saad. Higher order orthogonal iteration of tensors (hooi) and its relation to pca and glram. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, pages 355–365. SIAM, 2007.

Peter Stange. On the efficient update of the singular value decomposition. In *PAMM: Proceedings in Applied Mathematics and Mechanics*, volume 8, pages 10827–10828. Wiley Online Library, 2008.

Christine Lucas Tardif, Andreas Schäfer, Miriam Waehnert, Juliane Dinse, Robert Turner, and Pierre-Louis Bazin. Multi-contrast multi-scale surface registration for improved alignment of cortical areas. *NeuroImage*, 111:107 – 122, 2015.

Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

Mark J Van der Laan, Eric C Polley, and Alan E Hubbard. Super learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1), 2007.

Erik-Jan van Kesteren and Rogier A Kievit. Exploratory factor analysis with structured residuals for brain network data. *Network Neuroscience*, pages 1–27, 2020.

David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, and Kamil Ugurbil. The WU-Minn human connectome project: An overview. *NeuroImage*, 80:62 – 79, 2013.

Zhengwu Zhang, Genevera I. Allen, Hongtu Zhu, and David Dunson. Tensor network factorizations: Relationships between brain structural connectomes and traits. *NeuroImage*, 197:330 – 343, 2019.

Andrew Zalesky, Alex Fornito, and Edward T. Bullmore. Network-based statistic: Identifying differences in brain networks. *NeuroImage*, 53(4):1197 – 1207, 2010.