

Normal Modes of the Structural Connectome

Exploring the harmonic modes of the human brain using a graph
model of its geometry

Hoyt Patrick Taylor IV

Undergraduate Honors Thesis



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Physics

2019

Acknowledgements

This project would not have been possible without the generous support of the many wonderful people that helped me throughout the 2018-2019 academic year. I offer my deepest and most sincere gratitude to: Dr.

Bert Dempsey, a neighbor and long-time family friend, who took the time to read an early (and therefore confusing!) draft of this manuscript and offered helpful feedback; Ye Wu, who helped me learn how to use the MRtrix3 software package and generated tractograms of 20 million fibers for 10 subjects (which are not used in the results shown here, but will hopefully be used in continuation of this work); Dr. Zhengwang

Wu, who was incredibly kind and unfailingly generous with his time, and helped me with a myriad of different technical obstacles throughout the course of this project; my adviser Dr. Rosa Tamara Branca, who enabled me to conduct this research for the Physics department, kept me honest about both my time management and the physical interpretation of the model presented here, and who provided invaluable feedback on drafts of this manuscript; my advisers Dr. Han Zhang and Dr. Pew-Thian Yap, who met with

me for a minimum of 1 hour every week over the course of this project and responded to dozens (hundreds?) of rambling emails, providing technical guidance without which this work could not exist; my parents Katie and Hoyt Taylor, who provided much-needed moral support throughout this project, and to whom I owe both the countless opportunities that put me in the position to begin this work and my very

existence; and the many other individuals not explicitly named that helped me in various ways during the course of this project. I am humbled by the generosity of everyone that made this project possible.

Contents

1	Introduction	4
1.1	Modelling	4
1.2	Laplacian Eigenfunctions	4
1.3	Cortical Parcellation	5
2	Methods	6
2.1	Data and Preprocessing	6
2.2	Computation	7
2.2.1	MRtrix	7
2.2.2	MATLAB+Paraview	7
2.2.3	SciPy	7
2.2.4	SciKit-Learn	7
2.2.5	NumPy	8
2.3	Diffusion Tractography	8
2.4	White Matter Surfaces	9
3	Graph Formulation of Structural Connectome	10
3.1	Graphs	10
3.1.1	Surface Adjacency Matrix	11
3.1.2	Long-Range White Matter Fiber Adjacency Matrix	12
3.1.3	Total Adjacency Matrix	12
4	Normal Modes of the Structural Connectome	12
4.1	Individual	13
4.2	Group Average	15
4.3	Multi-Layer Graph	17
5	Clustering and Parcellation	17
5.1	Metrics of Variability	18
5.2	Color Mapping	18
5.3	Individual and Group-Average Parcellation	19
5.4	Multi-Layer Parcellation	21
5.5	Qualitative Comparison with Existing Parcellations	24

Abstract

Division of the human cortex into distinct regions is of high importance to neuroscientific inquiry. Fully-automated, multi-modal schemes of achieving such parcellation on an individual subject basis are particularly advantageous, however difficulties in inter-modal and inter-subject registration of brain images, as well as obstacles in preserving group-level correspondence of individual parcellation maps have slowed progress in this area. In parallel, there exists a relative dearth of data-driven parcellation schemes that incorporate high resolution structural connectivity metrics; the majority of widely-accepted parcellation maps in the literature have primarily used functional connectivity. Here, a fully data-driven, automated routine based on structural geometry and connectivity which achieves subject-specific cortical parcellation maps while maintaining group-level correspondence of maps is presented and optimized. Using high resolution white matter surface meshes and advanced fiber tracking techniques, a novel vertex-wise structural connectivity graph is constructed for each of 10 unrelated subjects, and the first k eigenvectors of the Laplacian Matrix of the graph's adjacency matrix are calculated. These eigenvectors represent the steady-state modes of vibration of the manifold described by this graph, and thus provide subject-specific maps of modes of connectivity in white matter. In order to obtain parcellations at varying levels of coarseness of the cortex from these eigenvectors, hierarchical agglomerative clustering is then performed on the surface mesh, where each vertex's feature vector is its profile in spectral space. Further, a multi-layer graph of all subjects is constructed, and individual parcellations with group level correspondence are obtained by agglomerative clustering of the eigenvectors of the laplacian matrix of the multi-layer graph.

1 Introduction

1.1 Modelling

When attempting to enumerate a theoretical framework governing the dynamics of complex systems, it is advantageous to simplify the system of interest. Although the potential experienced by an electron orbiting a proton is not solely described by the Coulomb interaction between the opposing charges, the wavefunctions and associated energies obtained by solving the Schrodinger equation with this simplified Hamiltonian of the Hydrogen atom are quite close to those obtained using degenerate perturbation theory and the full Hamiltonian. The human brain contains over 100 billion neurons, with on the order of 7^3 connections per neuron [1]. Aside from the reality that current empirical methods are not capable of accurately measuring the geometry of these connections at all scales of detail, the sheer geometric and cytoarchitectural complexity of this structure represent massive obstacles to bottom-up inference of the guiding principles of brain dynamics.

While individual neurons are capable of complex information processing, and their properties still constitute an active and open area of research, it has been proposed that on the population level, their behavior resembles that of tiny springs [2]. The model presented here draws on this simplification. We postulate that this simplification is akin to the simplification of the Hamiltonian of an electron in a Hydrogen atom to only its Coulomb interaction with the nucleus.

A proverbial tenant of biology states that the structure of biological systems underlies their function; function is solely explained by structure. In simple small-scale biological systems, this has been increasingly validated [3, 4, 5, 6]. There is more difficulty, however, in establishing this link in the human brain. While there has been moderate success of studies seeking to identify correlation between the structure of the brain and its temporal-spatial patterns of activation as measured by functional magnetic resonance imaging (fMRI), there are relatively few successful attempts to predict function from structure [7, 8, 9, 10].

This work describes a simple, fully data-driven model and processing pipeline that, from structural brain images, yields a complete, orthonormal basis of spatial patterns of functional activity on the human cortex. These patterns are referred to as structural connectome normal modes, harmonics, and eigenvectors. By virtue of its completeness and orthonormality, the basis of structural connectome normal modes is capable of describing any pattern of functional activation on the surface of the human brain quantified at or below the spatial resolution of the eigenfunctions. Further, we demonstrate that the information contained in a very small fraction of this basis can be used to calculate neuroanatomically meaningful parcellations of the cortex on the subject and group level at varying resolutions, and that calculated parcellations agree well with those in the literature derived from a wide array of brain features, including functional connectivity.

1.2 Laplacian Eigenfunctions

The Laplacian operator, the divergence of the gradient, is a fundamental feature of a broad array of physical theories including the Schrodinger equation, Maxwell's equations, the heat equation, Einstein's equations,

the general wave equation, and many other successful descriptions of physical phenomena. Many currently accepted explanations of complex physical systems owe their existence to the suitable formulation of a differential equation involving the Laplacian operator. Perhaps the most glaring example of this is the Schrodinger equation, which accurately predicts phenomena that were mystifying to the scientific community prior to its existence, and which is capable of describing dynamics of systems with infinite dimension.

The eigenfunctions of the Laplacian operator imposed on a flat, one-dimensional domain form the extraordinarily powerful Fourier basis, through which many of the most physically relevant dynamical paradigms can be described. The Fourier basis is universally used in analysis of countless physical systems that can be parameterized by one variable. Similarly, the orthonormal bases yielded by eigendecomposition of the Laplacian imposed on more complex domains are in many instances particularly convenient for representing physical dynamics parameterized by that domain [11, 12]. For instance: the steady-state modes of vibration of an elastic membrane held still at its edge by a rigid circle are the Bessel functions- the eigenfunctions of the Laplacian imposed on a circular, bounded domain. This basis can be used to efficiently represent all of the possible dynamics of the head of a drum. Here, a novel discrete model of a graph that parameterizes the structural geometry of the human cortex is presented, the eigenvectors of the Laplacian imposed on that domain are graph are calculated, and the efficacy of these normal modes as meaningful biological patterns is tested and verified. It is important to note that this model is largely based on that used by Atasoy et. al [13], however several important differences in the present model including the density of sampling of structural connectivity, the method of tractography used, the weighting of elements of the graph, and the implementation of multi-layer graph methods all render this model distinct and more robust. Additionally, the use of eigenvectors of the laplacian matrix of this graph model to parcellate the cortex into distinct functional regions is entirely novel.

1.3 Cortical Parcellation

Cortical parcellation refers to the division of the human cortex into distinct functional regions. This is important as neuroscience seeks to associate human behaviors and psychiatric pathologies with specific brain regions. The first and most extensively referenced parcellation scheme was devised by Korbinian Brodmann, and was based on a post-mortem examination of the cytoarchitecture of the brain. Research since then has revealed that in many instances, the layout of cell types in the brain does not always directly correspond to the layout of functional regions. Further, it is becoming increasingly clear that the organization of the brain is hierarchical; for any given region, there exist multiple sub-regions with distinct function. Based on this, it is obvious that there might be no ground-truth division of the cortex, as there are a functionally unlimited number of resolutions at which the brain can be parcellated. For this reason, it is difficult to establish a consensus in the community on the best method of parcellation.

It has previously been shown that the eigenvectors of the discrete Laplacian, or Laplacian matrix imposed on a polygonal mesh representing the surface of the cortex can be used to segment the cortex into its major

distinct lobes [14]. This result shows that the vibrational modes of a membrane shaped like the surface of the cortex are predictive of its large scale functional organization. Many previous studies have used measures of long-range white matter connections to parcellate the cortex [15, 16, 17]. Although several of these have achieved agreement with existing functional and histological measurements, to the author’s knowledge, all previous work has relied on coarsely discretizing the cortex using prior anatomical information. In contrast, this work does not employ any neuroanatomical priors, and discretizes the cortex and its long-range structural connectivity at an unprecedented level of detail, resulting in unbiased, potentially more accurate, and more detailed parcellations.

2 Methods

2.1 Data and Preprocessing

Anatomical or T1-weighted MRI scans provide high contrast between the brain’s white matter- the regions of the brain comprised of long-range neuronal axon fibers- and the brain’s gray matter- the tissue comprising the cortex (the outer layer involved in all high-level cognitive processes) containing a high density of neuronal cell bodies. The white matter/ gray matter interface refers to the region separating these two types of tissue.

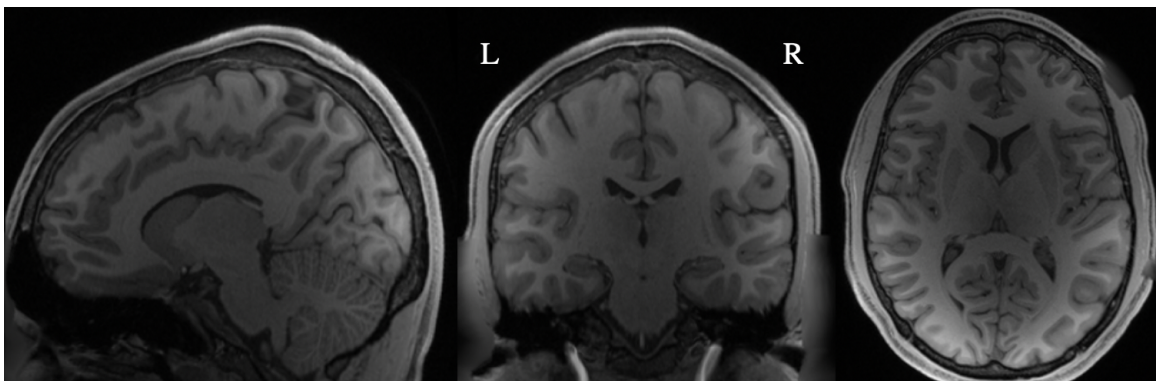


Figure 1: Example T1-weighted anatomical MR image for subject 1 from different view angles.

The high resolution T1 weighted and diffusion weighted images, and associated files from the human connectome project (HCP) preprocessing pipeline from 10 unrelated subjects aged 22-25 years old were downloaded from the HCP 100 unrelated subjects release of the human connectome project [18, 19]. No further preprocessing was performed. Of the additional files from the HCP preprocessing pipeline, only the white matter surfaces, the volumetric segmentation and the spherically inflated white matter registration files (used as described in **2.2.2** and **2.4**) were used in subsequent analyses.

2.2 Computation

All computations were performed locally on a laptop computer with a 6 core Intel i9 3.5 GHz processor and 32GB of RAM. As the computational scope of this work is quite broad, in instances in which existing open-source software packages were capable of achieving the computational processes or set of processes efficiently, these packages were downloaded and used.

2.2.1 MRtrix

All algorithms used in the processing of diffusion weighted MRI images to produce tractograms of 2.8 million fibers per subject (described in **2.3**) are included in the open-source MRtrix3 software package, which is extensively documented at <https://mrtrix.readthedocs.io/en/latest/index.html>. [20, 21, 22].

2.2.2 MATLAB+Paraview

All visualization was performed using ParaView [23]. Interfacing between Python and ParaView was achieved by first writing relevant arrays from Python to text files, and then writing ParaView-compatible files using the text files and an open-source MATLAB-wrapped package called `mvtk`.

2.2.3 SciPy

The matrices constructed in this work are too large to be stored and manipulated using standard numerical data structures on the local machine with 32GB of RAM. The matrices are in general sparse, and therefore special data structures designed for dealing with sparse matrices allowed for efficient manipulation. Specifically, methods within the `scipy.sparse` sub-library of SciPy- an open-source scientific programming library included in most downloads of Python3- was used to construct, manipulate, and solve for the eigenvectors and eigenvalues of the matrices [24].

Additionally, the calculation of dice score detailed in **5.1** was performed using the `dice` method of the `scipy.distance` sub-library within SciPy.

2.2.4 SciKit-Learn

For the matching of the 2.8 million endpoint coordinates to their nearest neighboring of the 81,924 white matter surface vertices (described in **3.1.2**) and for the matching of white matter surface vertices lying on the mid-brain structures to their nearest neighboring vertex on the opposite hemisphere (described in **3.1.1**), an efficient nearest-neighbors algorithm was necessary. Of the several different open-source algorithms tested, the most efficient was determined to be the `NearestNeighbors` method from the `neighbors` sub-library of SciKit-Learn open-source Python3 library [25].

The agglomerative hierarchical clustering approach described in **5** was also implemented using tools available in SciKit-Learn, specifically using the `AgglomerativeClustering` method of the `cluster` sub-

library. Additionally, the normalized mutual information score (described in 5.1) was calculated using the `normalized_mutual_info_score` method of the `metrics` sub-library of SciKit-Learn.

2.2.5 NumPy

All other data manipulations not explicitly mentioned in 2.2.1-2.2.4 were either performed directly using methods available within the NumPy python library, or performed using Python functions heavily relying on NumPy methods written by the author as detailed below.

Functions used in the construction of matrices and eigendecomposition thereof are contained in a script called `spectral_graph.py`. Functions used in the analysis of the vector spaces described by individual sets of eigenvectors (described in 4.1) are contained in a script called `SubjectAnalysis.py`. Functions used to perform spectral clustering are contained in a script called `Clustering.py`. Functions used to construct the super-adjacency matrices described in 4.3 and to perform numerical analysis of the consistency of parcellation results described in 5.1 are contained in a script called `ClusterProcesses.py`. Any and all of these scripts are available upon request to the author. Any data manipulations not performed with functions included in these scripts were performed in the iPython console and therefore are not explicitly documented, but attempted descriptions by the author are available upon request.

2.3 Diffusion Tractography

The diffusion MRI scans used provide a measure of the diffusivity of water in 60 different directions at every voxel (cube of side length 1.25mm) in the brain. Water diffuses more easily in the direction of white matter axonal fiber bundles. Diffusion tractography seeks to estimate the orientation of white matter fibers based on the diffusion magnetic resonance signal.

From the multi-shell, high resolution, registered diffusion MRI scans of each subject, the $b=1000 \frac{mm}{s^2}$ shell was extracted for each subject and the response function of the diffusion signal was estimated using the Tournier algorithm with the default parameters. Next, the fiber orientation distribution was calculated from the diffusion signal and estimated response function using the constrained spherical deconvolution algorithm with the default parameters. 10 million fiber tracts were then generated for each subject from the fiber orientation distribution using the iFOD2 probabilistic tractography algorithm. The seeding and termination of tracks was constrained to the white matter-gray matter boundary using the ‘anatomically constrained tractography’ option in the track generation algorithm, where the FreeSurfer segmentation file included in the data release was used to determine the gray matter white matter interface. Each tractogram of 10 million fibers was sifted to 2.8 million fibers using the fiber orientation distribution and the SIFT algorithm, and the spatial coordinates and connections of the endpoints of each tractogram were stored.

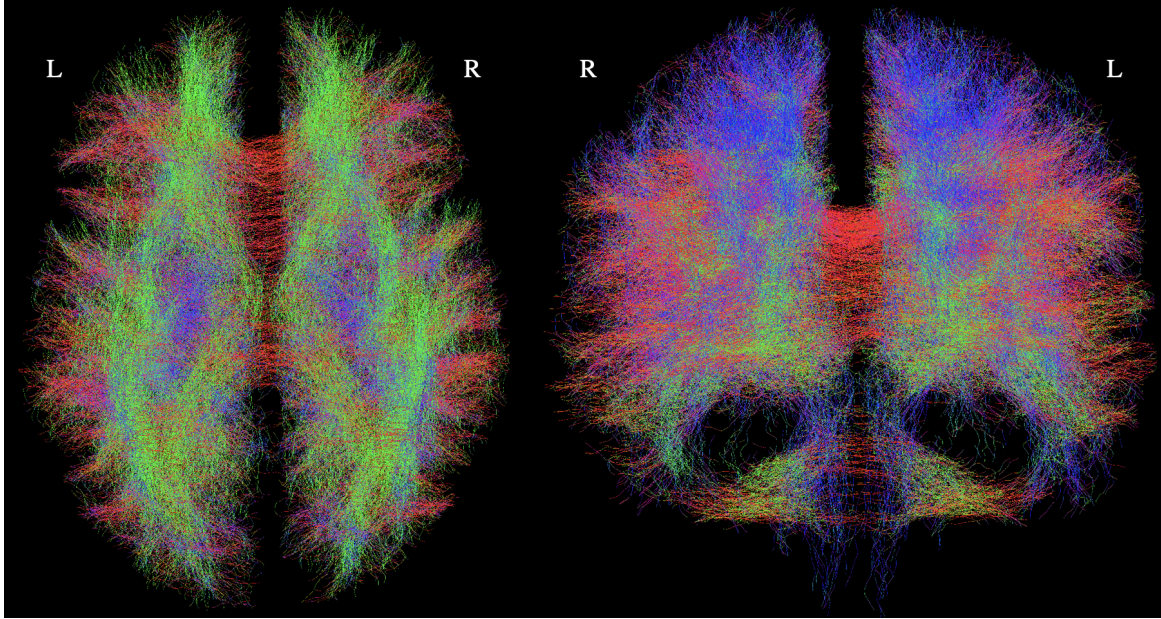


Figure 2: Example tractogram of subject 1 with 100,000 tracks. A lower number of tracks are shown for visualization purposes, as individual fibers are not visible in the tractograms of 2.8 million fibers. Fibers are colored according to direction.

2.4 White Matter Surfaces

The high resolution white matter surface mesh containing approximately 120,000 vertices, per hemisphere included in the data release were resampled to an equal number of vertices for each subject using barycentric interpolation, and were registered to a common space using the spherical demons method.

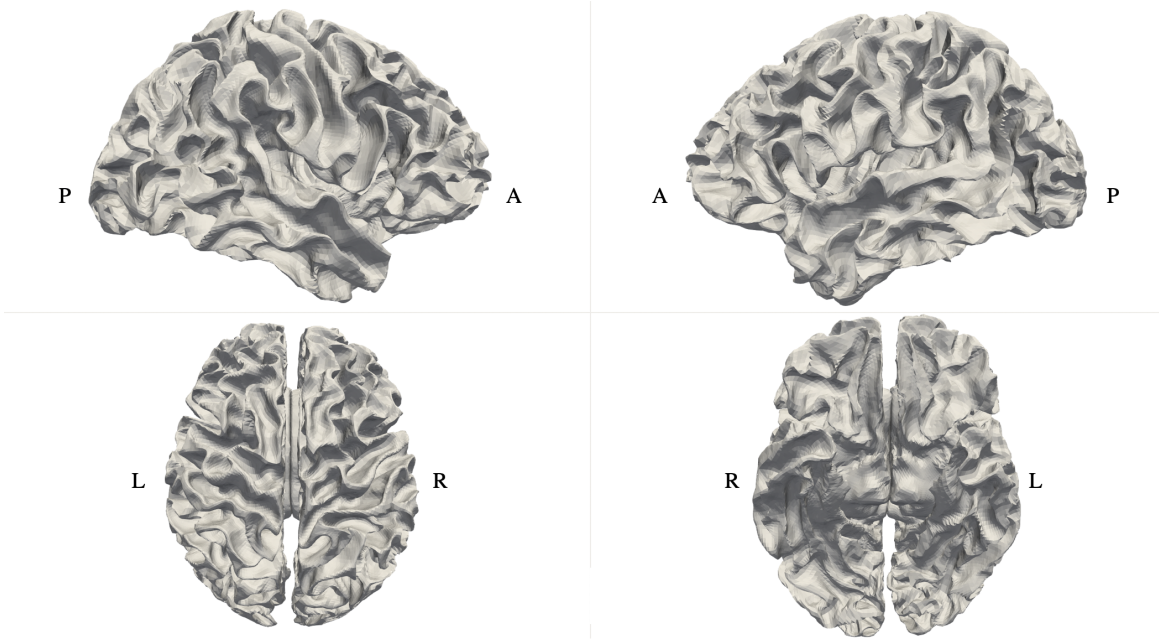


Figure 3: Resampled and registered white matter surface mesh of subject 1 from different view angles. "L" and "R" refer to left and right. "A" and "P" refer to anterior (front) and posterior (back).

The spatial coordinates of the vertices and the connections defining the polygonal mesh were extracted and stored for each subject. For a given subject, these vertices are given by

$$\begin{aligned}
 V &= \{v_i | i \in 0..1...N_{vert}\}, \\
 v_i &= (x_i, y_i, z_i)
 \end{aligned}
 \tag{1}$$

where $N_{vert} = 82,924$.

3 Graph Formulation of Structural Connectome

3.1 Graphs

A graph G is defined by a collection of vertices V and the edges E between those vertices, given by

$$E = \{e_{ij} | (v_i, v_j) \in (V \times V)\}.
 \tag{2}$$

Although the vertices V may be embedded space in a particular way, the connectivity of G and all of its topological properties can be described by its adjacency matrix \hat{A} , which is populated by

$$\hat{A}_{i,j} = \hat{A}_{j,i} = \begin{cases} w_{i,j} & \text{if } (i, j) \in E \\ 0 & \text{otherwise,} \end{cases}$$

(3)

where $w_{i,j}$ is the weight of the connection between vertices i and j .

Importantly, it has been shown that for a graph whose vertices are a sampling of a Riemannian manifold, in the limit of $N_{vert} \rightarrow \infty$, many of the properties of that graph approach their continuous analogs [26].

3.1.1 Surface Adjacency Matrix

By benefit of the vertex-vertex correspondence of the white matter surfaces achieved by the process described in 2.5, the indices of connected vertices are identical for all subjects, although the spatial embedding of vertices is variable. As graphs and their adjacency matrices are invariant to the spatial embedding of their vertices, only one adjacency matrix \hat{A}_{mesh} is necessary to represent the surface mesh of all subjects.

For the surface mesh with N_{vert} , its adjacency matrix \hat{A}_{mesh} is of size $(N_{vert} \times N_{vert})$. Here, the spatial resolution of $N_{vert}=81,924$ represents an unprecedented level of resolution for studies seeking to construct graph models of the structural connectivity of the cortex. \hat{A}_{mesh} is populated according to equation 3, where the edges are between vertices connected by the polygonal connections that define the mesh, and the edge weights $w_{i,j}^s$ are all equal to 1.

It is important to note that \hat{A}_{mesh} describes both hemispheres of the brain and as a result, there are two fully-connected regions (the left and right hemisphere) of $G(\hat{A}_{mesh})$. This is a result of the fact that current standard methods of constructing cortical surfaces from T1 weighted MR images- including those used to construct the surfaces used in this analysis- are only equipped to do so on a per hemisphere basis. When the two hemispheres are represented in a single graph, there exists a physiologically erroneous disconnection between them.

In order to correct for the fictitious division between the two hemispheres, an inter-hemisphere connection adjacency matrix \hat{A}_{IC} was constructed by connecting all vertices lying on the midline structures (Corpus callosum, brain stem, thalamus, etc) of the surface of each hemisphere to their nearest neighboring vertex on the opposite hemisphere with edge weight $w_{i,j}^{IC}=1$. The full surface adjacency matrix \hat{A}_s was then calculated as $\hat{A}_s = \hat{A}_{mesh} + \hat{A}_{IC}$.

The inclusion of \hat{A}_{IC} improved both the left-right symmetry of structural connectome normal modes and the symmetry of resulting parcellations. It is important to note that the construction and inclusion of \hat{A}_{IC} is an entirely novel feature of this model; in the model presented by Atasoy et. al- the only other study seeking to calculate connectome harmonics- no attempt to correct for the false disconnection between hemispheres of white matter surface meshes was made.

3.1.2 Long-Range White Matter Fiber Adjacency Matrix

All long-range white matter connectivity information relevant to this model is encapsulated by the spatial coordinates of the endpoints of each fiber and indices of endpoints that belong to the same fiber. Although the seeding and termination of tracts was constrained to the white matter/gray matter interface, the location of seeds was randomly sampled from a continuous domain representing this interface. Therefore, fiber endpoints are not in general at the precise location of the white matter surface mesh vertices.

In order to encode the connectivity information contained by the fiber endpoints in the space parameterized by the white matter surface mesh, each endpoint was associated with its nearest neighboring surface mesh vertex using SciLearn’s kd-tree nearest neighbors algorithm, where Euclidean distance was the metric. The average distance between fiber endpoints and their identified nearest neighboring surface vertex is less than the average spacing of surface vertices, indicating an effective matching.

The white matter fiber adjacency matrix \hat{A}_f^m for each subject is formed identically to \hat{A}_{mesh} , except that its edges $w_{i,j}^f$ correspond to white matter surface vertices connected by long-range white matter fibers, and is equal to 0.1. The lower magnitude of $w_{i,j}^f$ was determined through exploratory analysis of several different weighting strategies. This weighting strategy was determined to give the most physiologically meaningful representation of structural connectivity.

3.1.3 Total Adjacency Matrix

The number of nonzero elements of \hat{A}_f^m , \hat{A}_{mesh} , and \hat{A}_{IC} are 2,860,971 (group-average), 491,520, and 9,636 respectively. The structural connectome graph proposed is described by its adjacency matrix \hat{A}^m . For subject m , \hat{A}^m is given by

$$\hat{A}^m = \hat{A}_{mesh}^m + \hat{A}_{IC} + \hat{A}_f. \quad (4)$$

It is worth noting that certain elements of $G(\hat{A}^m)$ have weight 1.1 due to its definition as a sum of matrices, however it is overwhelmingly the case that the edges defined by \hat{A}_f^m are not between the same vertices as those defined by \hat{A}_{mesh} and \hat{A}_{IC} , and vice versa. This is shown by the fact that the number of nonzero elements in \hat{A}^m is approximately 3.22 million, whereas the sum of nonzero elements in each of \hat{A}_f^m , \hat{A}_{mesh} , and \hat{A}_{IC} is approximately 3.36 million.

4 Normal Modes of the Structural Connectome

The Laplacian matrix \hat{L} of graph $G(\hat{A})$ is defined as

$$\hat{L}(\hat{G}) = \hat{D}^{1/2} \hat{A} \hat{D}^{1/2}, \quad (5)$$

where

$$\hat{D}_{i,i} = \sum_{j=0}^{N_{vert}} A_{i,j}. \quad (6)$$

As stated in **3.1**, in the continuous limit of graph representations of manifolds, the Laplacian matrix approaches the continuous Laplace-Beltrami operator. This implies that, given a graph formed by a dense sampling of a manifold, the eigenvectors of the Laplacian matrix of that graph give the steady state vibrational modes of that manifold [11, 27]. That is, the eigenvectors correspond to the spatial patterns of motion and the eigenvalues correspond to the energy or frequency associated with those patterns. Here, large N_{vert} gives a good approximation of the continuous Laplacian. Further, the number of fiber connections (2.8 million) gives a dense sampling of structural connectivity in comparison with the only other study seeking to calculate connectome harmonics [13].

The physical interpretation of the eigenvectors of this graph is somewhat ambiguous; although the present model seeks to parameterize the human brain, which is continuous, our graph model could also describe a collection of masses connected in a network by springs. N masses connected by a network of springs with effective spring constants $k_{i,j}^s$ between masses i and j can be represented by a graph whose adjacency matrix \hat{A}_{spring} is populated according to equation 3, where the edge weights are $k_{i,j}^s$. The normal modes of vibration of the physical system described by \hat{A}_{spring} are the eigenvectors ψ_k of the laplacian matrix $\hat{L}(\hat{A}_{spring})$, given by

$$\hat{L}\psi_k = \lambda_k\psi_k. \quad (7)$$

Notably, this formulation is identical to the present formation of the structural connectivity graph, where the ‘effective spring constants’ are 1 for vertices connected by the surface mesh, and .1 for vertices connected by long range fiber connections. We set up these parameters in light of the assumption that local coupling of the cortex (through short fibers and other short-range connections) is stronger than coupling due to long range fibers (association, projection and commissural fibers). The high level of heterogeneity in whole brain functional connectivity profiles under various task regimes gives evidence that this assumption is valid. In parallel, the high proportion of edges defined by fiber connections implies that much of the geometric complexity of the human brain is contained in its white matter geometry, which is trivially obvious.

4.1 Individual

N_{harm} structural connectome normal modes, $\psi_0 \dots \psi_k \dots \psi_{N_{harm}}$ and $\lambda_1 < \lambda_2 < \dots < \lambda_{N_{harm}}$, for each subject m were calculated according to equation 7, with $\hat{L} = L(\hat{A}^m)$.

Here, $N_{harm}=99$. ψ_0 is the trivial solution to the eigen-problem, contains no meaningful information, and is thus excluded from the set of eigenvectors. Several exploratory analyses of the relationship between the eigenvectors of different subjects were conducted. Each subject’s set of eigenvectors is self orthogonal:

$$\begin{aligned} \{i \neq j | 0 \dots N_{harm}\}, \\ \langle \psi_i, \psi_j \rangle = 0, \end{aligned} \quad (8)$$

but they represent only a small subset of the complete set of N_{vert} eigenvectors. Since each A^m is

populated according to a different set of 2.8 million long range white matter connections, there is no guarantee that the subspaces described by each subject’s calculated set of eigenvectors are equivalent. Visual inspection of the eigenvectors of different subjects color mapped onto the white matter surface reveals that similarity between across subjects degrades for wavenumber $k > 10$. Figure 4 demonstrates this by the decay in the normalized dot product between corresponding eigenvectors for a pair of subjects.

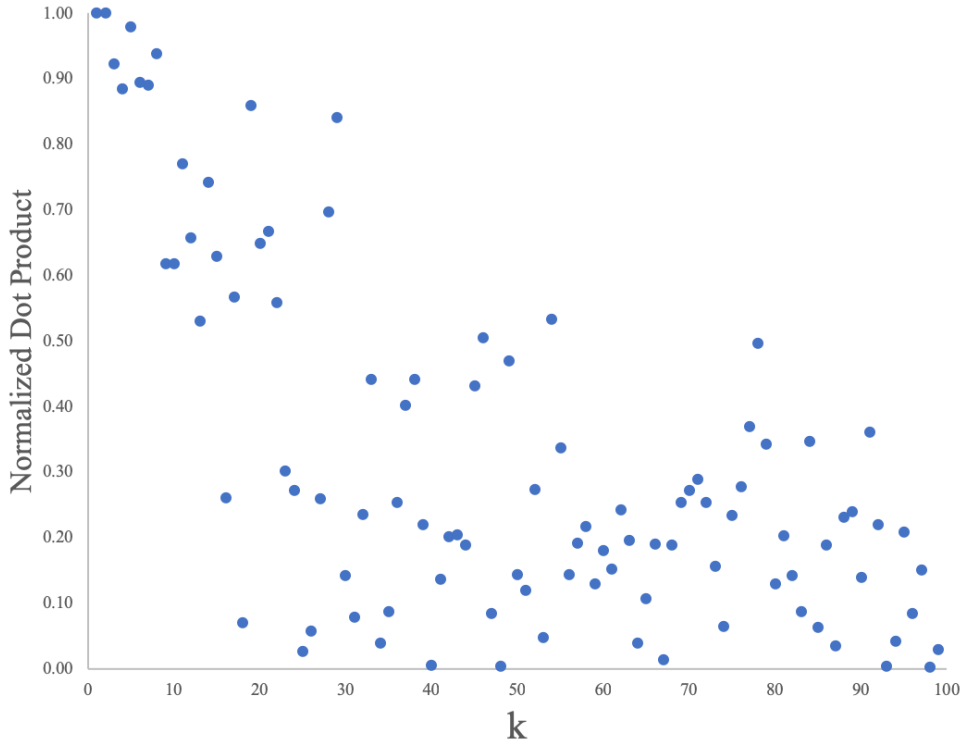


Figure 4: Normalized dot product of corresponding eigenvectors of subjects 1 and 3

In order to investigate the similarity of the subspaces represented by each subject’s eigenvectors, we define subject n ’s reconstruction quality $\Omega_k^{n,m}$ of eigenvector k of subject m as:

$$\Omega_k^{n,m} = \langle \omega_k^n, \psi_k^m \rangle, \quad (9)$$

where

$$\omega_k^{n,m} = \sum_{j=1}^{N_{harm}} \langle \psi_j^n, \psi_k^m \rangle \psi_j^n. \quad (10)$$

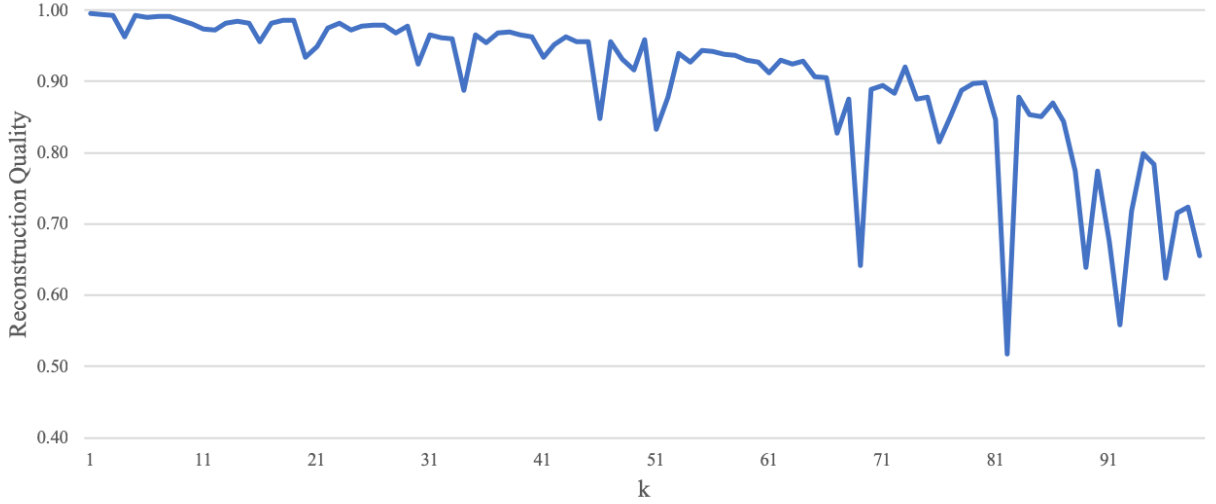


Figure 5: Overlap of subspaces. $\Omega_k^{3,1}$ is plotted versus k . The reconstruction quality is high throughout the spectrum, with few individual exceptions and a slight decay near $k = 85$. This indicates a high degree of overlap between the subspace represented by subject 3’s eigenvectors with the subspace of subject 1’s eigenvectors.

4.2 Group Average

In addition to calculating connectome harmonics from each subject’s $\hat{L}(A^m)$, N_{harm} group average harmonics, $\overline{\psi}_k$, were calculated from $\overline{L} = L(\overline{A})$, where

$$\overline{A} = \frac{1}{N_{sub}} \sum_{m=1}^{N_{sub}} \hat{A}^m, \quad (11)$$

where in this study, the number of subjects $N_{sub} = 10$. Figures 6 and 7 display selected average eigenvectors color mapped onto the averaged and smoothed white matter surface. There are no large qualitative disagreements between individual subject eigenvectors and group average eigenvectors, however there is not in general correspondence between individual and average eigenvectors of the same wavenumber.

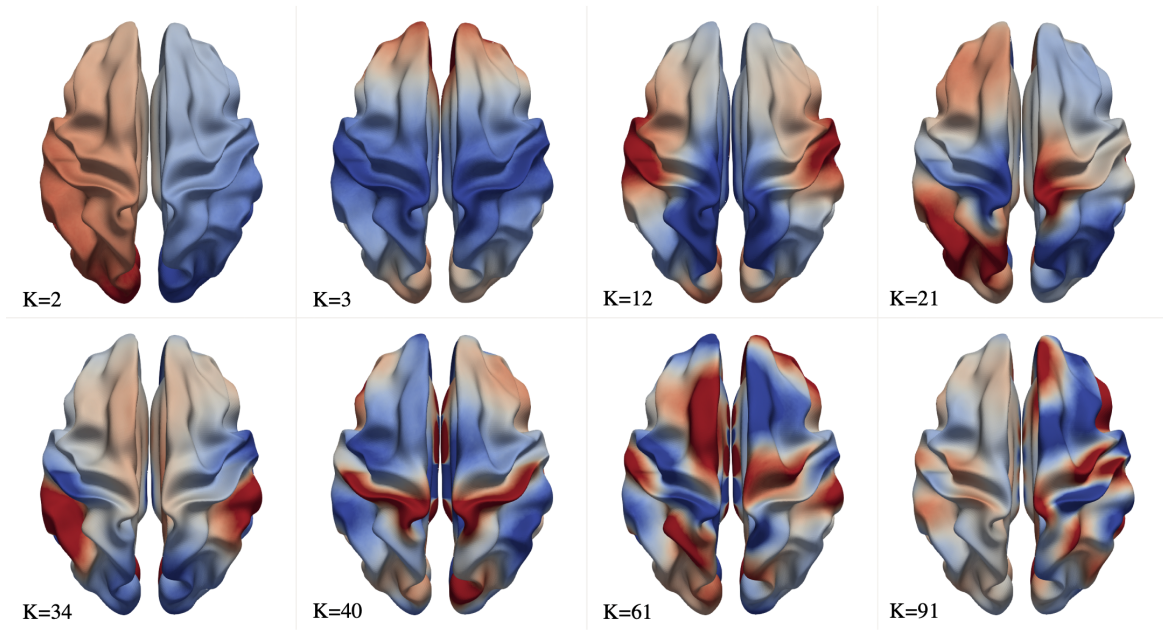


Figure 6: Selected group-average connectome harmonics viewed in $-z$ direction. Red indicates positive values and blue indicates negative values.

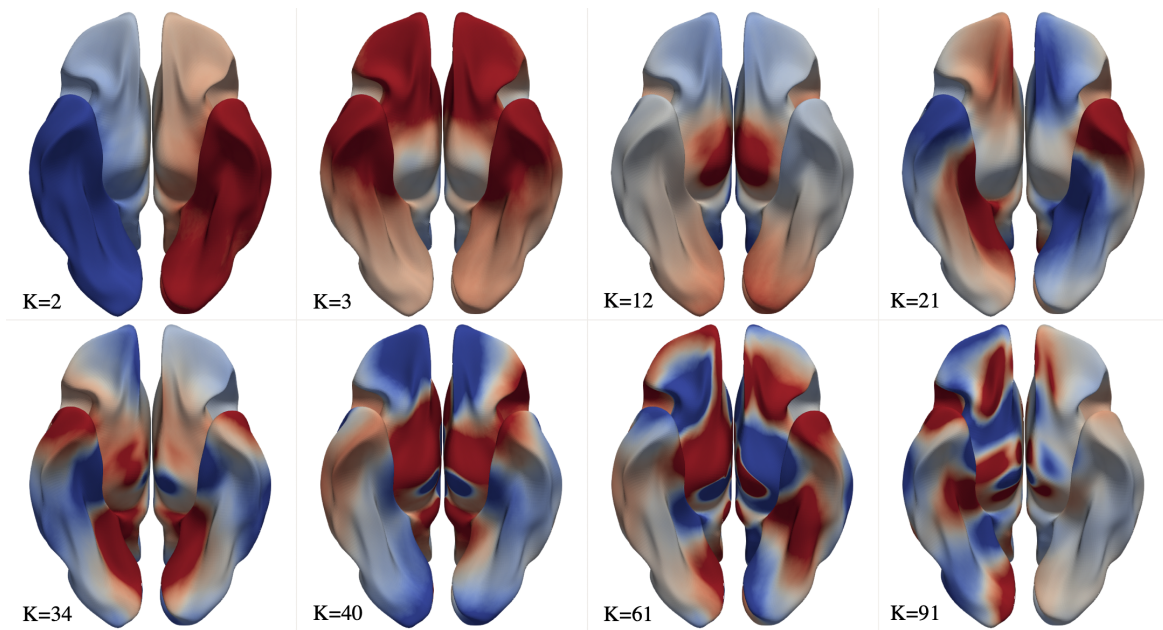


Figure 7: Selected group-average connectome harmonics viewed in $+z$ direction. Red indicates positive values and blue indicates negative values.

4.3 Multi-Layer Graph

Given M graphs $G(A^m)$, each with N vertices, a multi-layer graph can be represented by the super-adjacency matrix \mathbf{A} , a block-diagonal matrix whose diagonal blocks of size $(N \times N)$ are the adjacency matrices A^m , and all other blocks are the identity matrix multiplied by an inter-layer weighting strength γ .

When the vertices of the subgraphs $G(A^m)$ represent the same features, and properties common to all M are of interest, $\mathbf{G}(A^m)$ provides useful information regarding both the common features and the differences between the set of $G(A^m)$'s [28]. Multi-layer approaches have been employed in many neuroimaging studies, including in the work of Lefevre et al [14], which uses spectral clustering of the eigenvectors of the laplacian matrix of a fully surface-based graph to coarsely parcellate the cortex into 5 major regions. These studies demonstrate the efficacy of multi-layer approaches to establish group-level correspondence between features of individuals while preserving individual differences [28, 14].

Here, a super adjacency matrix of size $(819,240 \times 819,240)$ was constructed using the individual adjacency matrices A^m . The eigenvectors of the super adjacency matrix \mathbf{A} , ψ , are of length $N_{vert} \times 10$. A subject's, individual multi-layer eigenvectors $\tilde{\psi}_k^m$ are defined as the length N_{vert} sections of ψ_k with indices corresponding to those of that subject's A^m within \mathbf{A} . In contrast to the individual eigenvectors, there is high similarity between $\tilde{\psi}_k^m$ across m . In subsequent analyses, the effect of varying the inter-layer weighting strength γ is explored. As expected, the magnitude of γ was found to correspond to the degree of coherence of eigenvectors across subjects.

5 Clustering and Parcellation

As a test of the hypothesis that the eigenvectors of the Laplacian matrix of the present graph model of the structural connectomes reflect physiologically meaningful spatial patterns of the function of the human cortex, vertices of the white matter surface were hierarchically clustered by their 'eigenprofile', or spectral coordinates. We define the eigenprofile of v_i as x_i , where

$$x_i = [\psi_1(v_i), \psi_2(v_i), \dots, \psi_{N_{harm}}(v_i)]. \quad (12)$$

Hierarchical clustering of vertices based on their spectral coordinates was performed using SciKit-Learn's Agglomerative clustering algorithm. The matrix $\hat{A}_{mesh}^m + \hat{A}_{IC}$ was also inputted to the algorithm to speed the tree construction initialization. The output of the algorithm when initialized with N_{clust} clusters is a cluster label vector c of length N_{vert} , with indices identical to the vertex indexing, and with an integer valued entries (labels) between 0 and N_{clust} indicating the cluster to which the vertex was assigned.

Clustering was carried out for each set of individual eigenvectors, ψ^m , for the group-average eigenvectors, $\bar{\psi}$, and for the super eigenvectors ψ calculated using 6 different inter-layer weighting strengths. In each case, the clustering was performed at various levels of coarseness. Also in each case, vertices lying within the mid-brain structures were not included in the clustering, and are given a uniform and distinct cluster label.

5.1 Metrics of Variability

To measure the similarity between different clustering results, we employ two different metrics of similarity: the normalized mutual information score and the dice score. Mutual information between pairs of cluster label vectors was calculated as described in 2.2.4 with the two cluster label vectors as input. Dice score between two clusters was calculated by first binarizing the cluster labels, and calculating the dice score as described in 2.2.3 with the two binary cluster vectors as input. Both of these metrics take on a value between 0 and 1.

Dice score was used to investigate the distribution of similarity between pairs of clustering results across clusters, which is only meaningful when there is correspondence between the location of vertices given a particular label between the two clustering results being compared. Because the clustering algorithm was initialized independently for the clustering of individual eigenvectors, the resulting labels are not spatially coherent across subjects, and therefore the dice score was not used to evaluate their consistency.

5.2 Color Mapping

It is of interest to view parcellation results on the cortical surface in order to qualitatively compare them to existing parcellations in the literature. The scalar values representing cluster labels contained in given cluster label vectors are agnostic to the spatial distribution of the clusters on the cortex, as they are calculated solely from the spectral coordinates given by the set of x_i 's. Therefore, if the scalar values are directly mapped to the white matter surface, each scalar value must be assigned a color through some method. For a relatively small number of clusters, it is not overly time-consuming to manually construct the color map (as is done in figure 8), however for large N_{clust} this method becomes implausible.

In order to efficiently assign cluster labels to a color based on spatial criteria, a python script called `ColorMap.py` (available upon request to the author) that assigns each cluster label a unique red, green, blue (R,G,B) coordinate was written and employed for parcellation results where $N_{clust} > 80$. `ColorMap.py` works by assigning each cluster label an (R,G,B) coordinate based on its barycentric coordinates, where the reference simplex (here, a triangle) is defined by three "major lobe centers" corresponding to the approximate cartesian centers of the frontal, temporal, and parietal lobes of the cortex. Under this formulation, clusters that are closer to or within the frontal lobe are more red, clusters closer to the parietal lobe are more green, and clusters closer to the temporal lobe are more green. In order to allow for better visual discrimination between clusters, each (R,G,B) coordinate for each cluster was modulated by a small factor determined using a random number generator.

5.3 Individual and Group-Average Parcellation

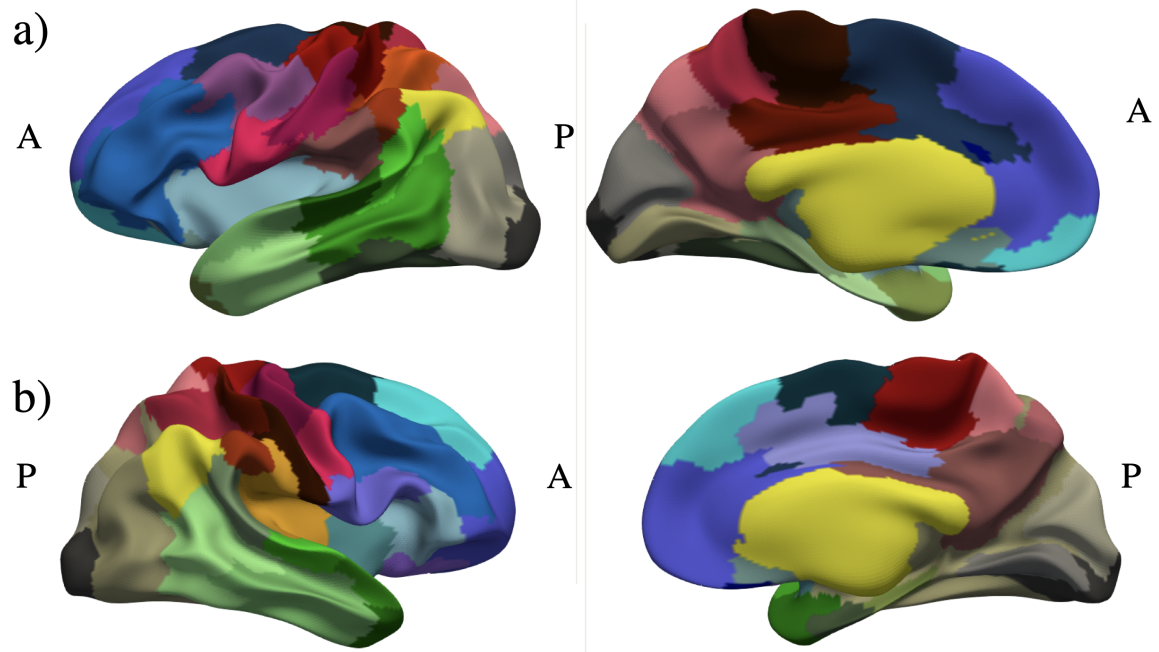


Figure 8: Parcellation result calculated using $\bar{\psi}$'s with $N_{clust} = 80$ manually color-mapped onto the a) left hemisphere and b) right hemisphere of the smoothed group average white matter surface.

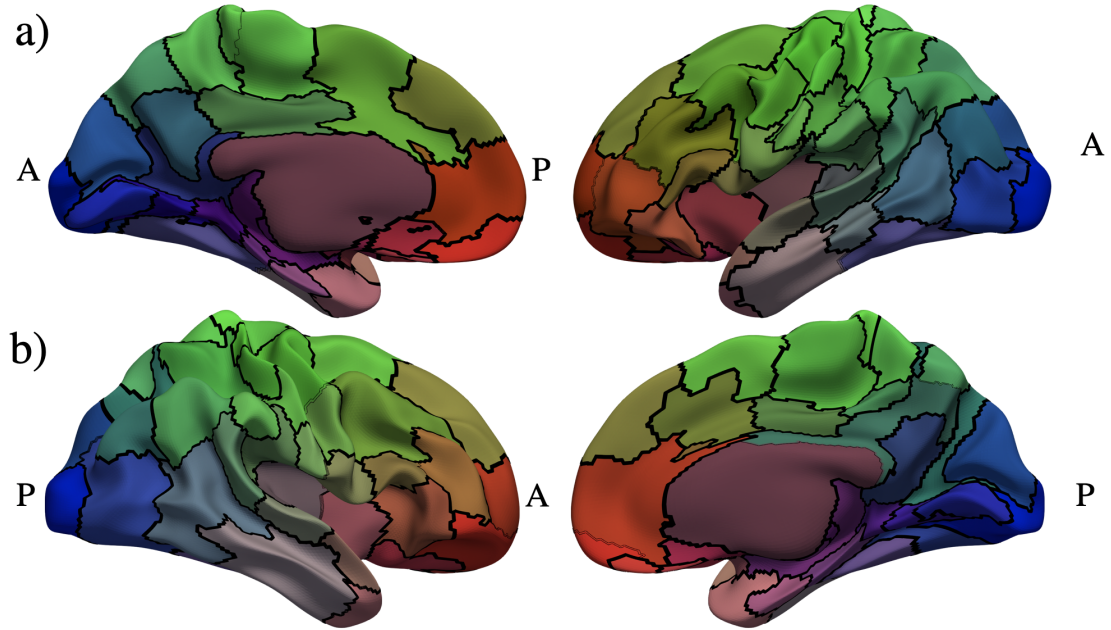


Figure 9: Parcellation result calculated using $\bar{\psi}$'s with $N_{clust} = 120$ color-mapped using barycentric coordinates onto manually color-mapped onto the a) left hemisphere and b) right hemisphere of the smoothed group average white matter surface. Parcel borders appear in black.

In order to quantitatively evaluate the consistency of individual parcellation results, the mutual information between each subject's cluster label vector and the group average cluster label vector was calculated for 4 values of N_{clust} (figure 10).

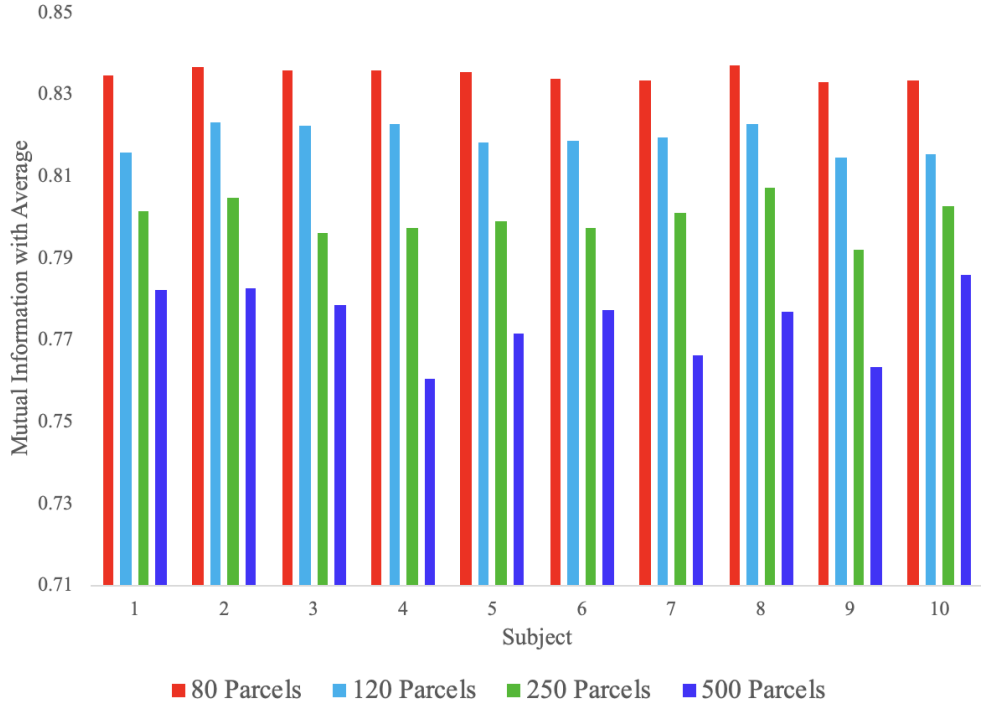


Figure 10: Mutual information between individual cluster label vectors c^m and group-average cluster label vector \bar{c} .

As expected, the similarity between individual parcellations and the group average parcellations decreases with increasing N_{clust}

5.4 Multi-Layer Parcellation

Clustering of the super eigenvectors gives a super cluster label vector \mathbf{c} of length $N_{vert} \times 10$, and like the multi-layer individual eigenvectors, the individual multi-layer cluster labels \tilde{c}^m are given by the length N sections of the super clustering label corresponding to the positions of the individual adjacency matrices in the super adjacency matrix.

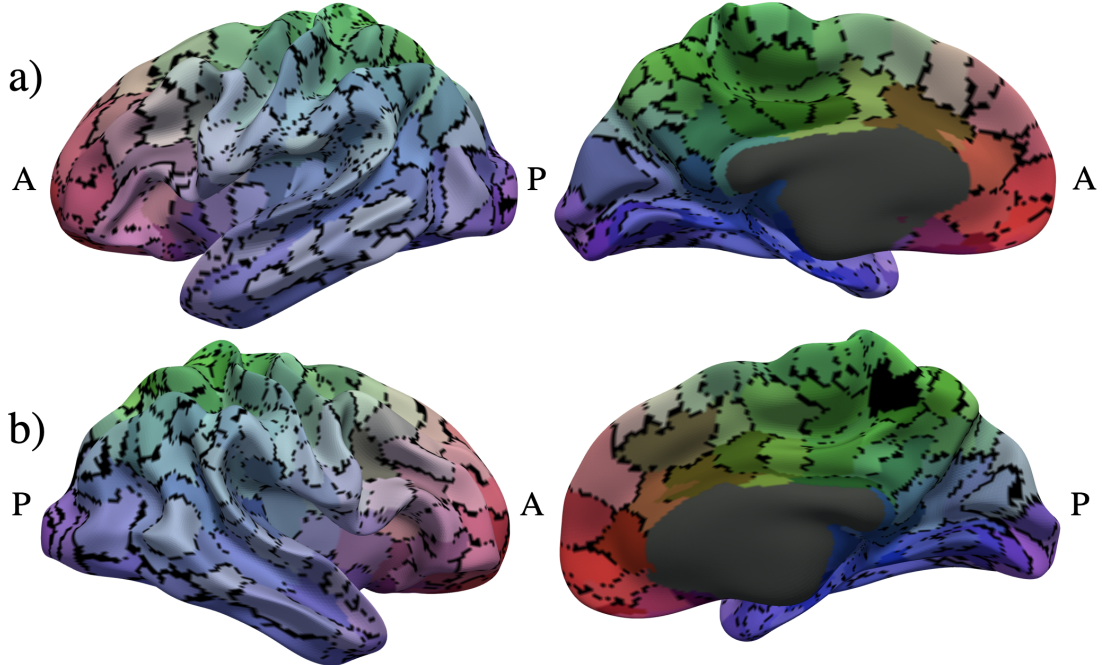


Figure 11: Consensus parcellation calculated using $\tilde{\psi}$'s with $N_{clust} = 500$ and $\gamma = 0.8$ color-mapped using barycentric coordinates onto the a) left hemisphere and b) right hemisphere of the smoothed group average white matter surface. Vertices for which all \tilde{c}^m do not agree are colored black.

In contrast to individual cluster labels c^m , there is spatial correspondence between cluster labels across m for the individual multi-layer clustering results \tilde{c}^m . To evaluate the effect of varying N_{clust} and γ on the consistency of clusters both across subjects and across clusters, we first compute the consensus cluster labelling for a given set of individual multi-layer clustering results. The consensus labelling of a vertex is given a particular label only if the labels of that vertex agree across all subjects. Otherwise, it is given a label outside of the range of $0 - N_{clust}$ designated as the disagreement label. Figure 12 demonstrate the effect of inter-layer weight γ and number of clusters on the cluster-wise agreement between individual multi-layer clustering results \tilde{c}^m and the consensus cluster labelling. As expected, agreement decreases with both γ and with N_{clust} .

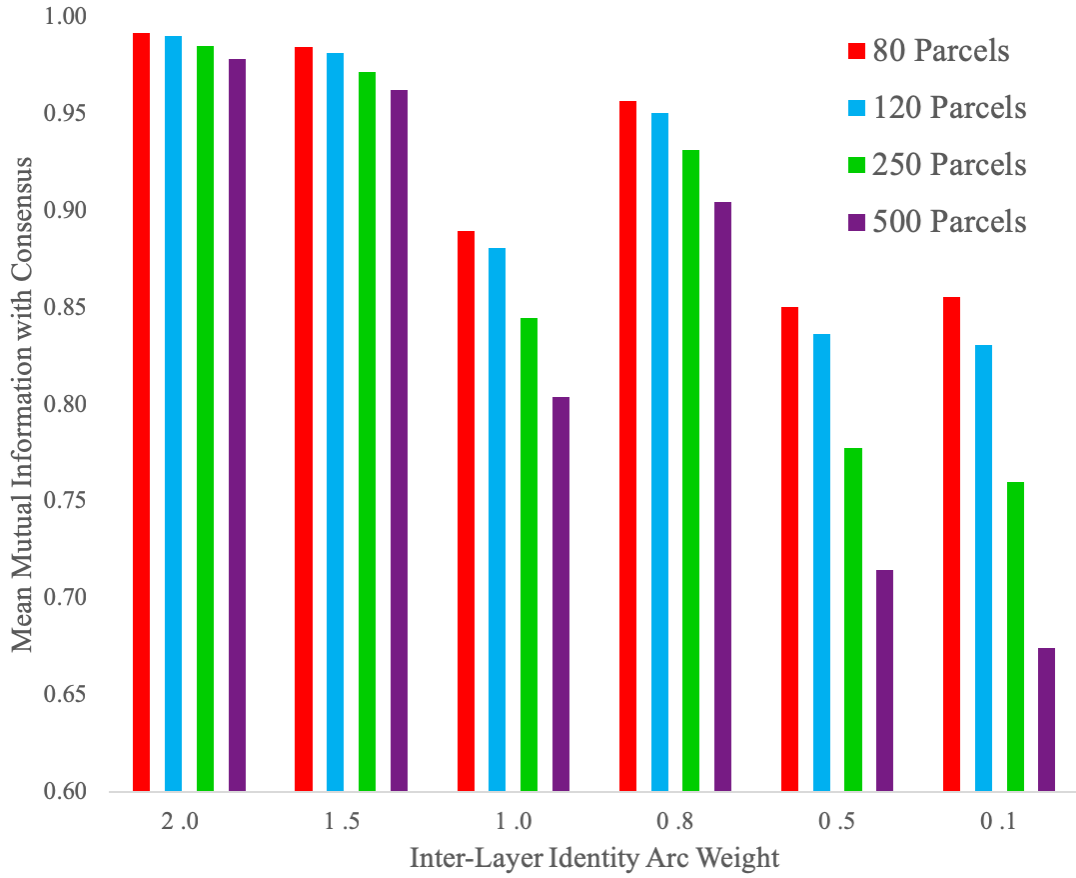


Figure 12: Mutual information with consensus parcellation for varying number of parcels and varying interlayer weighting strength.

The consistency of parcellation results was also evaluated for the multi-layer clustering results on a per-parcel basis using the dice score. Figure 13 displays the bottom half of the cross-subject average distribution of per-parcel dice scores with the consensus parcellation for $N_{clust} = 80$ and $N_{clust} = 500$.

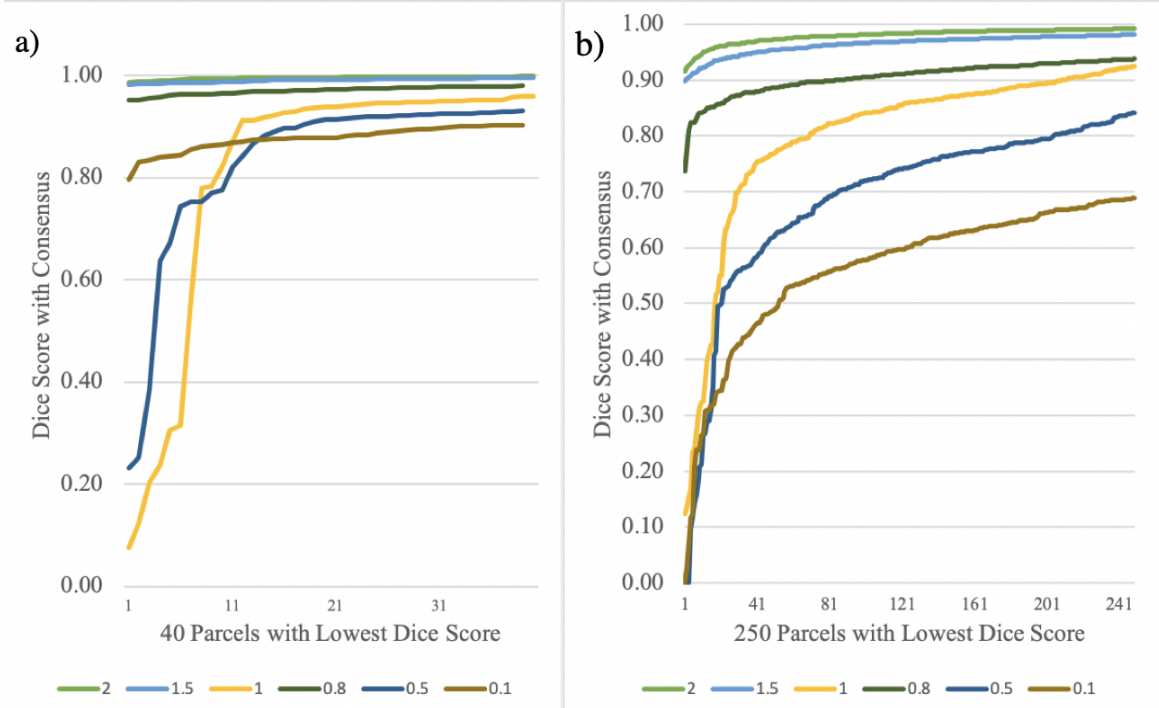


Figure 13: Group average per-parcel dice score with consensus parcellation for a) 40 parcels with lowest average dice score from clustering with $N_{clust} = 80$ b) 250 parcels with lowest average dice score from clustering with $N_{clust} = 500$. Different values of γ are plotted as separate lines, with the legend indicating the color of the line and the value of γ .

From figure 13, it can be seen that for high values of γ , there is high similarity between all parcels and the consensus parcels, and that for lower values of γ , there are individual multi-layer clusters with little or no similarity to the consensus parcels of the same label. The number of these minimally-overlapping parcels increases with increasing N_{clust} and with decreasing γ . This suggests that the value of γ can be varied in order to obtain a desired level of group-level correspondence of parcellations, and that the optimal value of γ depends on the desired N_{clust} .

5.5 Qualitative Comparison with Existing Parcellations

All calculated parcellations show qualitative similarity with various existing parcellations in the literature, including but not limited to those in [29, 30, 18]. This similarity is variable across N_{clust} , which literature parcellation is used as reference, and the particular clusters that show similarity. In general, higher N_{clust} is necessary to achieve similarity with literature between frontal-lobe parcels than for temporal, parietal, and occipital lobe parcels. Conversely, at high N_{clust} , it is almost always possible to achieve similarity between combinations of calculated parcels with larger parcels in literature parcellations.

Taken together, these results indicate the model presented in this work produces structural connectome

normal modes that meaningfully reflect brain function.

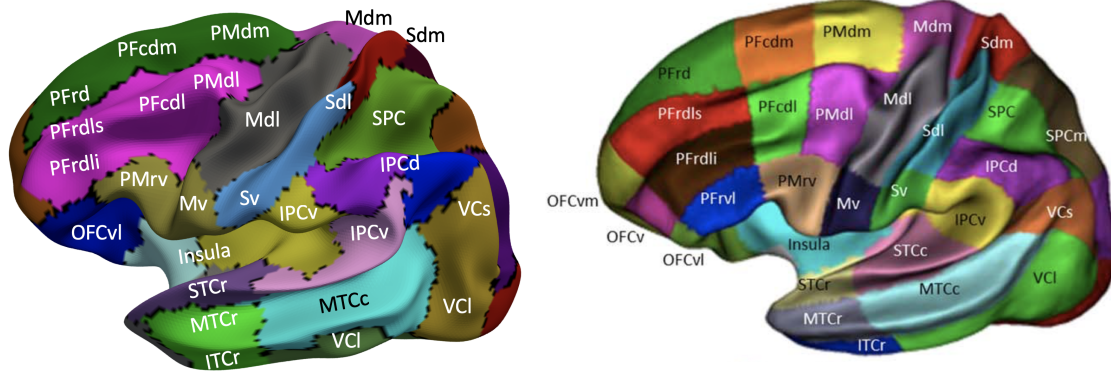


Figure 14: Left hemisphere parcellation result calculated using $\bar{\psi}$'s with $N_{clust} = 80$ manually color-mapped and labelled to resemble the MarsAtlas parcellation [30].

References

- [1] Herculano-Houzel, S.: The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience* (2009)
- [2] Sardi, S., Vardi, R., Sheinin, A., Goldental, A., Kanter, I.: New Types of Experiments Reveal that a Neuron Functions as Multiple Independent Threshold Units. *Scientific Reports* (2017)
- [3] Slack, J.M.: *Molecular Biology of the Cell*. In: *Principles of Tissue Engineering: Fourth Edition*. (2013)
- [4] Saraste, M.: Biomembranes: Molecular structure and function. *Trends in Biochemical Sciences* (2003)
- [5] Mercer, T.R., Mattick, J.S.: Structure and function of long noncoding RNAs in epigenetic regulation (2013)
- [6] Lee, D., Redfern, O., Orengo, C.: Predicting protein function from sequence and structure (2007)
- [7] Van Den Heuvel, M.P., Mandl, R.C., Kahn, R.S., Hulshoff Pol, H.E.: Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human Brain Mapping* (2009)
- [8] Meuli, R., Honey, C.J., Sporns, O., Gigandet, X., Hagmann, P., Cammoun, L., Thiran, J.P.: Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences* (2009)
- [9] Koch, M.A., Norris, D.G., Hund-Georgiadis, M.: An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage* (2002)

- [10] Damoiseaux, J.S., Greicius, M.D.: Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity. *Brain Structure and Function* (2009)
- [11] Lévy, B.: Laplace-beltrami eigenfunctions towards an algorithm that "understands" geometry. In: *Proceedings - IEEE International Conference on Shape Modeling and Applications 2006, SMI 2006.* (2006)
- [12] Niethammer, M., Reuter, M., Wolter, F.E., Bouix, S., Peinecke, N., Ko, M.S., Shenton, M.: Global Medical Shape Analysis using the Laplace-Beltrami-Spectrum. In: *MICCAI07, 10th International Conference on Medical Image Computing and Computer Assisted Intervention.* (2007)
- [13] Atasoy, S., Donnelly, I., Pearson, J.: Human brain networks function in connectome-specific harmonic waves. *Nature Communications* (2016)
- [14] Lefèvre, J., Pepe, A., Muscato, J., De Guio, F., Girard, N., Auzias, G., Germanaud, D.: SPANOL (SPectral ANalysis of Lobes): A spectral clustering framework for individual and group parcellation of cortical surfaces in lobes. *Frontiers in Neuroscience* (2018)
- [15] Gao, Y., Schilling, K.G., Stepniewska, I., Plassard, A.J., Choe, A.S., Li, X., Landman, B.A., Anderson, A.W.: Tests of cortical parcellation based on white matter connectivity using diffusion tensor imaging. *NeuroImage* (2018)
- [16] Ganepola, T., Nagy, Z., Ghosh, A., Papadopoulo, T., Alexander, D.C., Sereno, M.I.: Using diffusion MRI to discriminate areas of cortical grey matter. *NeuroImage* (2018)
- [17] Tittgemeyer, M., Rigoux, L., Knösche, T.R.: Cortical parcellation based on structural connectivity: A case for generative models (2018)
- [18] Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M.: The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* (2013)
- [19] Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K.: The WU-Minn Human Connectome Project: An overview. *NeuroImage* (2013)
- [20] Smith, R.E., Tournier, J.D., Calamante, F., Connelly, A.: SIFT: Spherical-deconvolution informed filtering of tractograms. *NeuroImage* (2013)
- [21] Smith, R.E., Tournier, J.D., Calamante, F., Connelly, A.: Anatomically-constrained tractography: Improved diffusion MRI streamlines tractography through effective use of anatomical information. *NeuroImage* (2012)
- [22] Tournier, J.D., Calamante, F., Connelly, A.: MRtrix: Diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology* (2012)

- [23] Ahrens, J., Geveci, B., Law, C.: ParaView: An end-user tool for large-data visualization. In: Visualization Handbook. (2005)
- [24] Oliphant, T.E.: SciPy: Open source scientific tools for Python (2007)
- [25] Varoquaux, G., Buitinck, L., Grisel, O., Louppe, G., Mueller, A., Pedregosa, F.: Scikit-learn. GetMobile: Mobile Computing and Communications (2017)
- [26] Singer, A.: From graph to manifold Laplacian: The convergence rate (2006)
- [27] van Dam, E.R., Haemers, W.H.: Developments on spectral characterizations of graphs. Discrete Mathematics (2009)
- [28] Zhang, H., Stanley, N., Mucha, P.J., Yin, W., Lin, W., Shen, D.: Multi-layer large-scale functional connectome reveals infant brain developmental patterns. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). (2018)
- [29] Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., Fox, P.T., Eickhoff, S.B., Yu, C., Jiang, T.: The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. Cerebral Cortex (2016)
- [30] Auzias, G., Coulon, O., Brovelli, A.: MarsAtlas: A cortical parcellation atlas for functional mapping. Human Brain Mapping (2016)