Discrete Parcels and Continuous Gradients:

Algorithmic Analyses of Cortical Areal Architecture and Function

Kristian Matthew Eschenburg

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2022

Reading Committee:
David R. Haynor, Chair
Thomas J. Grabowski, Chair
Paul E. Kinahan

Program Authorized to Offer Degree:
Bioengineering

University of Washington

Abstract

Discrete Parcels and Continuous Gradients: Algorithmic Analyses of Cortical Areal Architecture and Function

Kristian Matthew Eschenburg

Chairs of Supervisory Committee:
David R. Haynor
Department of Neurology
Thomas J. Grabowski
Departments of Neurology and Radiology

The human brain is characterized by a hierarchical system of organization that highlights its complexity at various scales within numerous structural and functional contexts. One mode of hierarchical patterning captures the flow of information through the brain from the sensory cortices, to associative regions, and finally to more integrative cortical areas that underpin complex cognition. Another mode captures increasingly complex physical organization, transitioning from raw genetic material to complex neuronal structure. In the first part of this work, we discuss methods for *mapping* the human cortex in order to identify and delineate putative cortical areas characterized by unique, homogenous functional and structural organization properties. In the second part of this work, we examine the internal organization properties of these cortical areas in more depth, and show that connectivity patterns exist as a *continuum* along the cortical manifold, and that the degree of continuity is related to the functional integration capacity of each respective cortical area.

To my mother, Nilufar, father, Ralf, and sister, Sonja: thank you for your unwavering support, encouragement, and love. My pursuit of the esoteric, the obscure, the fantastical – in short, the sciences – was only possible because of you.

I would like to thank all of the members of the Integrated Brain Imaging Center for their extensive support over the entire duration of my PhD. To my mentors, Dr. David Haynor, and Dr. Thomas Grabowski – thank you for your never-ending encouragement and support. David – one of the most memorable quotes and guiding philosophies that you taught me over the course of my PhD was to "minimize dependencies wherever possible". While science can't progress without integrated efforts, I've incorporated this mantra into my day-to-day life, and I'm considerably better for it. Thank you for sharing in my passion for the outdoors, and for being willing to share your thoughts and experiences about climbing, backpacking, and general philosophy on life. Tom – I appreciate your steadfast approach to managing as large of a lab as IBIC. Along with the rest of the IBIC faculty, you created an inclusive and fair research environment that helped everyone thrive, and shaped my perspective on the definition of a positive and nurturing workplace. You told me once to write my thesis "just a little bit at a time – a paragraph every day". I can't tell you how much time and stress actually doing this saved me.

Liza, you were the bread and butter of my experience at UW – thank you for your ever-positive presence, encouragement, free food, pulled-strings, Husky football tickets, and most importantly, friendship. Physics and Astronomy gained a hell of a Lab Manager, and more importantly, human being. Tina, my PhD would have ended in a fiery mess of doom were it not for your constant technical support, installation of random software packages, and willingness to help out with my late-night questions about why our cluster machines "weren't working". It's only half in jest when I say that you did as much work for my PhD as I did.

I like to joke that the majority of my PhD was completed whilst sitting in coffee shop chairs. I've given more of my life to Herkimer Coffee and Stone Way Cafe than I'd like to admit, but have developed a fondness for the drone of espresso machines, the buttery smell of roasting beans, and the ephemeral white noise of background conversation that, in one form or another, facilitate my creativity and focus.

I would also like to thank Matthew F. Glasser for making subject-level Human Connectome Project multi-modal parcellations available for analyses in Chapter 4, and for his extensive comments and suggestions for the original drafts of the corresponding manuscript.

These projects were all supported by NSF BCS–1734430 grant, titled "Collaborative Research: Relationship of Cortical Field Anatomy to Network Vulnerability and Behavior".

Contents

List of Figures

Abstract

The human brain is characterized by a hierarchical system of organization that highlights its complexity at various scales within numerous structural and functional contexts. One mode of hierarchical patterning is captured by the flow of information through the brain from the sensory cortices, to associative regions, and finally to more integrative cortical areas that underpin complex cognition and thought. Another mode captures increasingly complex structural properties, and represents an axis of brain *scope*, transitioning from raw genetic material to complex neuronal structure.

This sensory-to-integrative flow of information is canonically exemplified in the visual cortex, where retinal cells in the eye project to the thalamic lateral geniculate nucleus (LGN) and onto the primary visual area, V1, which then maps (both in series and in parallel) to downstream higher-order visual areas (Figure 1) [20]. Along this path, different cortical areas are sensitive to increasingly complex visual features that represent the integration of lower-level visual information. For example, the ventral visual stream involved in object recognition follows a gradient of visual cues related to orientation, position, and luminance (V1, V2), color (V4) and movement (V5), pattern and form recognition (BA 19, BA37), and specific object identification (BA 20, BA37). Each of these areas is tuned and partial to increasingly complex visual features, and the spatial organization of these areas on the cortex follows suit, in a graded posterior occipital-to-middle temporal manner.

These smoothly varying spatial gradients are not restricted to the visual cortex, and

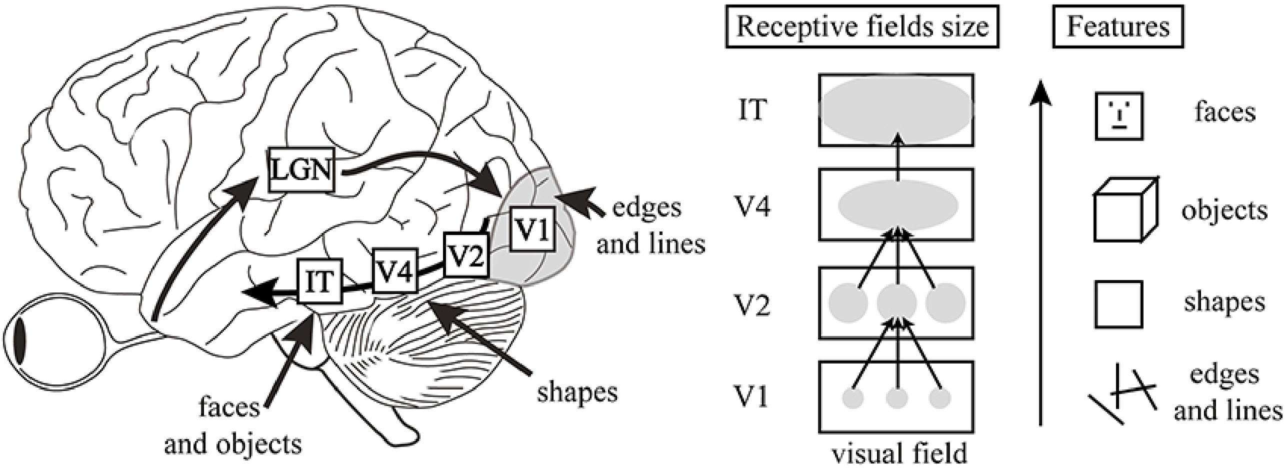


Figure 1: Hierarchical flow of visual information in the ventral visual stream, from [1]. Visual signals received by the retina are processed in both a serial and parallel fashion beginning with the sensory organs that map onto the primary visual cortex, which maps onto more associative areas. Cortical areas more downstream in the pipeline are tuned to more complex visual features and tasks.

have been identified across the whole brain. Functional gradients are mirrored by biological properties like the T1w/T2w ratio that captures relevant information about myelination in the cortex. Studies have identified greater myelination in somatosensory areas and lower myelination in transmodal cortices [18, 24]. Myelination, is related to laminar differentiation, such that increased myelination is associated with increased laminar differentiation – consequently, the ability to discern individual cortical layers follows an analogous gradient to myelin content on a sensory-transmodal axis. In studies relating cortical gradients between monkeys and humans, researchers have identified hierarchical gradients of expression patterns of genes related to synaptic physiology, cell-type specificity, and cortical cytoarchitecture that mirrors T1w/T2w myelination organization patterns [25]. These studies all point to a brain organization paradigm along a sensory-transmodal axis that varies smoothly along the cortical manifold, where integrative capacity of cortical areas increases with each incremental step along this axis.

While we can view hierarchical brain organization from the perspective of areal organization, and the integration of sensory information *along* the cortex, we can also

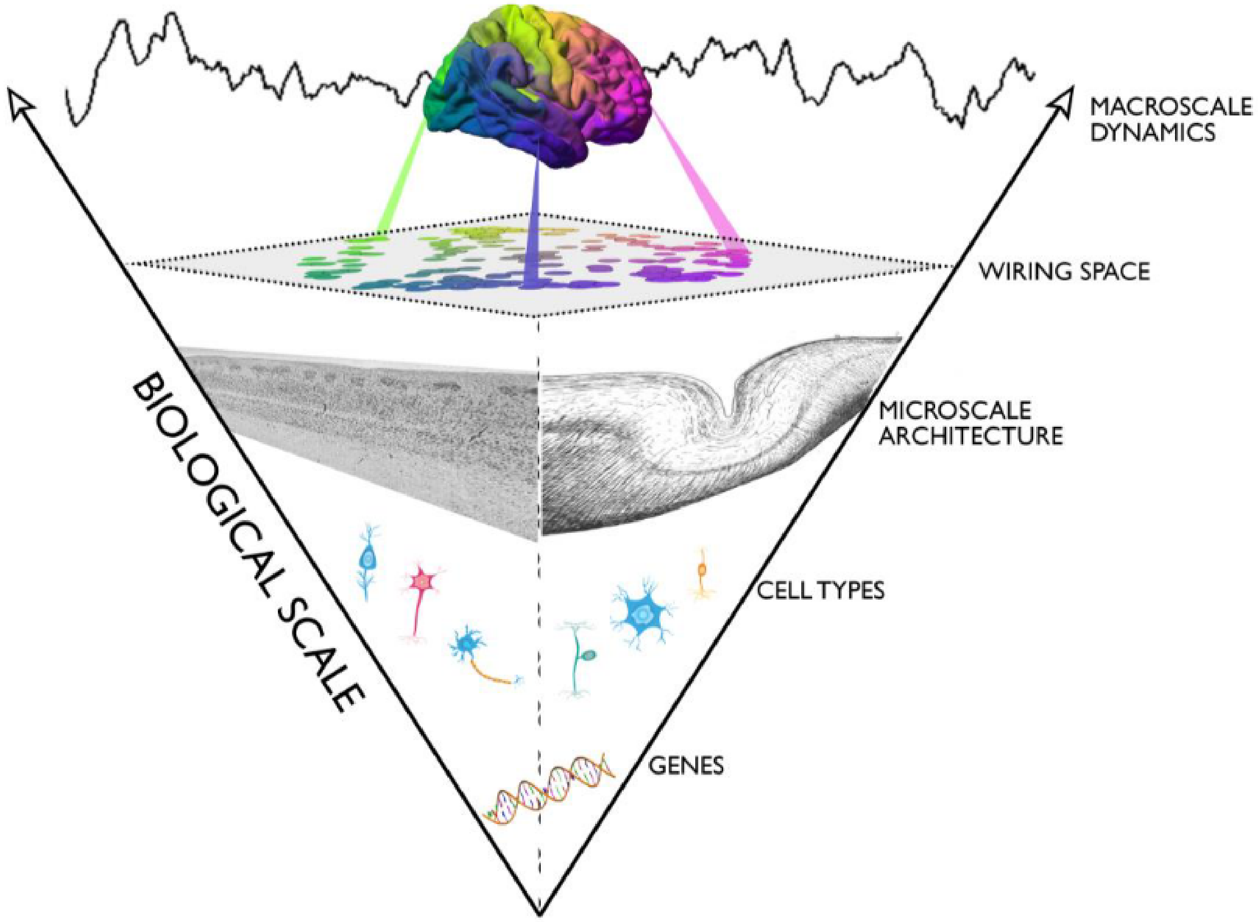


Figure 2: A hierarchy of scope and granularity, as shown in [2]. At the lowest nanoscale resolution, DNA captures variability related to gene expression. At the highest macroscale resolution, we have large-scale lobar patterning and brain network function. In between, we consider the "meso-scale" resolution consisting of cortical parcels and their smoothly varying areal organization ("wiring space").

view hierarchical organization from the perspective of scope, resolution, and granularity *within* the cortex (Figure 2) [2]. For example, DNA and gene expression encode the highest level of granularity in the cortex, and account for structural and functional variation in the brain. We refer to gene expression profiles as the "nanoscale" resolution. As noted above, researchers have identified spatial and temporal gradients of gene expression with respect to various properties that vary smoothly along the cortex. At the next level of granularity, what might be referred to as the "microscale" resolution, we consider spatially varying cell types, which arise due to unique spatio-temporal patterns of gene expression. Homogeneous cell-types give rise to cortical layers and columns, with relatively uniform local characteristics within columns, but that display spatially varying profiles when moving along the cortical manifold.

At the highest ("macroscale") level of brain organization, are large-scale lobes and functional networks that correspond to differential cognitive capacity and specificity. For example, the frontal lobe is shown to have expanded in humans, relative to other species, and this expansion consequently allowed for the evolution of complex thought and self-recognition. These cognitive abilities can be decomposed into large-scale functional networks, that are represented as cohesive spatial areas that co-activate and interact simultaneously.

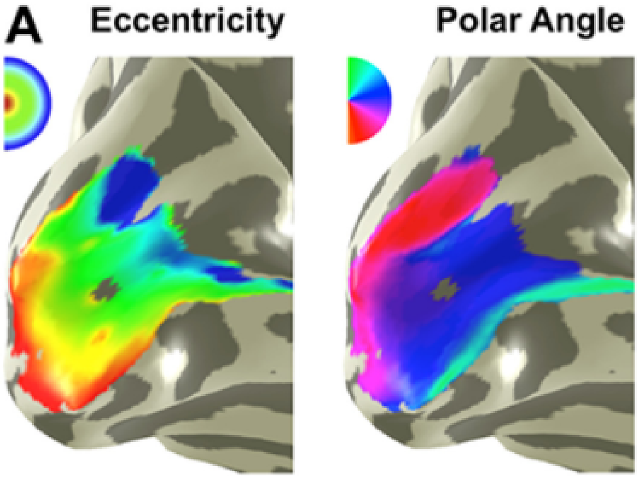


Figure 3: Smoothly varying gradients in the primary visual area (V1) of retinotopic organization, as shown in [3]. Eccentricity relates to the radial distance from the center of the eye's visual field. Polar angle relates to the degree of rotation of an object in the visual field, relative to some fixed coordinate system. We see that both eccentricity and polar angle map onto V1 in a spatially smooth manner, and capture different intrinsic organization patterns in the primary visual area.

Most common in brain imaging research, however, is a mid-range level of granularity, deemed the "mesoscale" organization. The mesoscale level is represented by discrete cortical areas and parcels that are characterized by *relatively* homogeneous functional and structural connectivity patterns. Each of these cortical areas are often thought of as a cohesive "unit", preferentially activating or deactivating during various cognitive tasks. However, even at this mesoscale resolution, organization is not black and white. If we "zoom" in on these cortical areas to examine the intra-regional variation of biological signals – we find that functional and structural organization varies smoothly in space (Figure 3) – this spatial variation is sharp at the boundaries between regions, and gradual within regions [26]. We take these variations to be a marker of the *continuous connectivity* of a region – a smooth representation of local scale trends and transitions in connectivity profiles within discrete cortical areas.

In the analyses that follow, we are interested in the interplay between the meso- and macro levels of brain organization. We present analyses, results, and suggestions for future work on both the discrete and continuous aspects of functional and structural brain connectivity. My dissertation will be divided into two primary segments.

In Chapter 1, we discuss the background and theory underlying various maps and delineations of the human neocortex, and discuss work involved in developing these cortical maps using different modalities. Then, in Chapters 3 and 4, we present work on two related, albeit considerably different, approaches to map the cortex of previously unseen data. Both proposed approaches rely on previously generated maps of the cortex to guide the mapping of new images. However, we'll show that one method is considerably more robust than the other, makes fewer structural assumptions about the data, and takes advantage of the inherent spatial relationships of the cortex. In Chapter 5, we discuss work related to mapping subsections of the cortex in the inferiorparietal area using functional brain connectivity. We apply the methods in Chapter 4 and relate predicted cortical maps to resting-state network organization. In Chapter 6, we discuss work related to the characterization of continuous connectivity gradients distributed across cortical regions. We examine geometric properties of these connectivity gradients and show that these properties are related to the functional specialization of cortical regions, and display patterns of increasingly complex organization. Then, in Chapter 7, we examine topographic organization patterns *between* pairs of cortical areas and relate this organization to the intrinsic functional gradients discussed in Chapter 6. Finally, in Chapter 8, we discuss future avenues of work that apply the topics discussed above, and extend the methodology developed in this dissertation to answers additional questions linking the structure and function of the human brain.

Chapter 1

Introduction

1.1 Mapping the cortex through the ages

The human cortex can be divided into layers (parallel to the cortical surface) on the basis of cell density and cytoarchitecture properties, and into columns (orthogonal to the cortical surface) on the basis of neuronal myelination patterns. Local variation in these measures has been used to subdivide the cortex into discrete units, often referred to as "cortical fields" or "cortical areas". The earliest attempts to map the cortex were based on work by neuroanatomists Theodor Meynert, one of the pioneering researchers of cytoarchitectonics, and Korbinian Brodmann, in the early 1900s. Brodmann identified 44 cortical areas in the human cortex using Nissl staining to characterize the histological features of various cortical areas [27]. Brodmann defined cortical areal boundaries where cytoarchitectural properties – that characterize cell shape and size – began to change, or diverge. Though his initial work was crude in terms of spatial resolution, relative to more modern studies today, some Brodmann areas (notably areas in the sensory cortex) have withstood the test of time and have remained relatively unchanged, while others have been further subdivided, refined, or removed entirely from cortical atlases (Figure 1.1). While he lacked evidence to support his claim at the time, Brodmann was convinced of a relationship between the structural and functional properties of cortical areas, certain that cortical regions with unique cytoarchitectural profiles also served unique cognitive roles. Brodmann's work was later extended and refined by neuroanatomists Constantin von Economo and Georg Koskinas, who subdivided the human cortex into seven large scale cortical lobes, and further delineated these lobes into 102 subregions on the basis of cytoarchitecture.

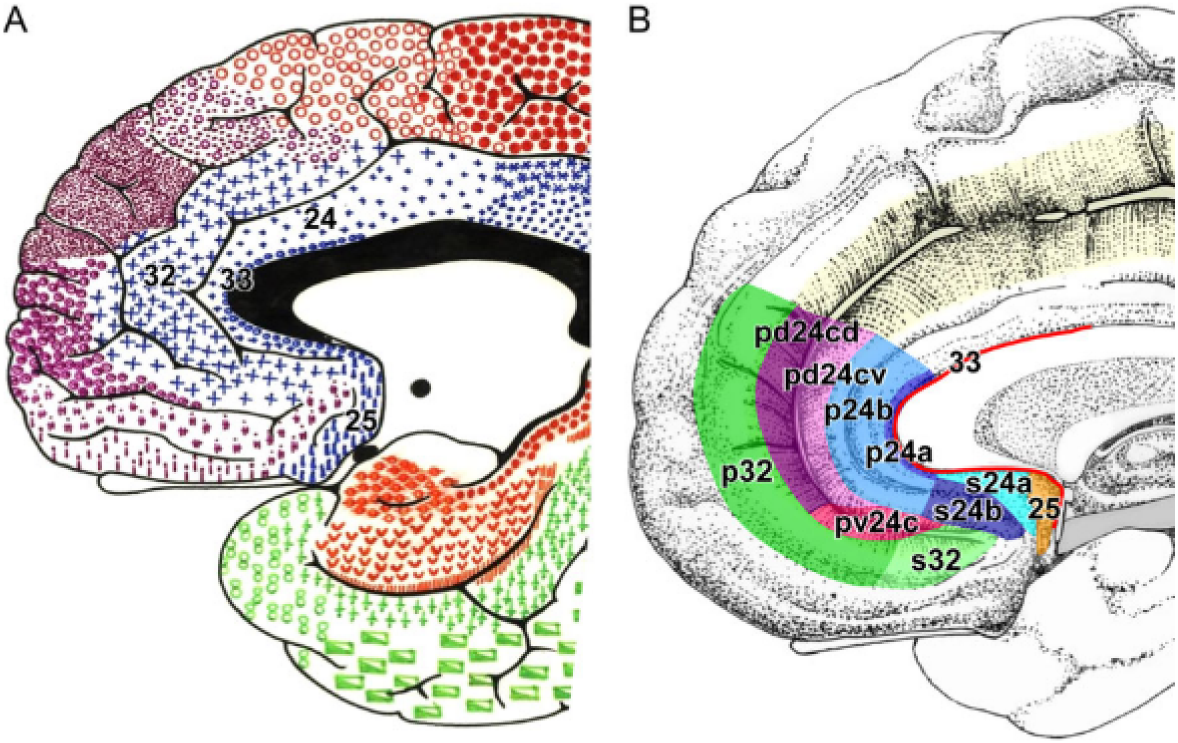


Figure 1.1: Brodmann's original cytoarchitectural (left) map versus a receptor-driven (right) map of the anterior cingulate cortex [4]. Increasing scrutiny by incorporating receptor expression profiles and myeloarcthiecture has enabled researchers to further subdivide and refeine original cortical areal boundaries proposed by Brodmann.

Following Brodmann's hypothesis of a link between the structural and functional organization of the brain, [28] explored this idea in more depth in both humans and non-human primates. They performed electrophysiology mapping in humans and monkeys and compared these results to previously-defined histological maps of both species in order to identify a possible link between electrical neuronal properties and cellular architecture. Their results were later confirmed by [29] who analyzed architectural modularity in the somatosensory cortex and by [30] who identified repeating functional subunits throughout the visual cortices.

The earliest attempts to map the cortex performed *in vitro* analyses, typically with an approach called Nissl staining. Nissl dyes tag negatively charged molecules (e.g. DNA and RNA). Specifically, Nissl bodies, organelles found in neurons, are rich in ribosomal RNA, and are thus easily identified using Nissl dyes. The distribution of Nissl bodies varies by neuronal cell type. Different areas of the cortex are characterized by different distributions of cell types (which display unique cytoarchitectural properties) along cortical layers, which enabled early neuroscientists to identify, roughly, the boundaries between discrete cortical subunits on the basis of cell size and shape.

Advances in immunohistochemistry enabled scientists to target specific proteins in tissues using antibodies, themselves tagged with fluorescent molecules. Neuroscientists adapted these tools to tag neuron receptor proteins, to examine the expression of various receptors throughout the cortex. By coupling different fluorescent molecules to different antibodies, neuroscientists were able to identify the expression levels of many receptors simultaneously, and thereby increase the dimensionality of data available for distinguishing cortical areal boundaries. One of the largest studies to date from the Jülich Research Center has mapped regional and laminar differences in cell body and neurotransmitter receptor distribution patterns for nearly 200 cortical areas by aggregating data from 10 individual human post-mortem datasets [8]. In a recent paper, [31] showed that laminar receptor profiles of 17 unique neuronal receptors did not coincide strongly with boundaries laminar boundaries defined by cytoarchitecture or myeloarchitecture, indicating a possible different mechanism of structural organization beyond that based on cell size and shape. Likewise, different functional areas (e.g. somatosensory vs. visual), and functional areas spanning multiple processing hierarchy levels (e.g. primary vs. association areas) displayed their own receptor profiles, indicating that receptor density profiles reflect unique functional specialization. Using both receptor densities and cytoarchitectonics [5, 32] identified 7 unique subunits in the inferior parietal area, each with their own unique *in vivo* structural and functional connectivity profiles. As such, cortical mapping based on receptor densities has shed new insight into the structural and functional organizational of the brain beyond those result achieved using cyto- and myeloarchitecture alone.

The studies by [32] and [5] are also interesting because they highlight what can be described as the "continuity" or cortical areas (Figure 1.2). While these works were primarily focused on delineating boundaries between putative cortical areas in the inferior parietal region, the studies found that subregions in the IPL also demonstrated considerable within-region receptor expression variability, as well as smoothly-varying spatial profiles of receptor densities as one moved along the cortical manifold. These results demonstrate that, while neuroimaging studies are often interested in segmenting the cortical manifold into discrete subunits and generating aggregate statistics of neurobiological signals within these subunits, the cortex is in fact a *continuous* map where areal boundaries can be defined on the basis of many neurobiological features using a variety of metrics. Further, these neurobiological features vary smoothly in space, where the *local* variability of these features *also* varies in space, and with their own degrees of spatial auto-correlation and covariance with other cortical measures.

Early *in vivo* brain imaging studies proposed defining areal boundaries on the basis of macroscopic landmarks, such as cortical folding patterns in the gyri and sulci. However, relatively few areas of the cortex show correspondence between folding patterns and features computed using in vitro such as cytoarchitecture, myeloarchitecture, and transcriptomics, beyond the border between the primary motor and primary sensory cortices in the central sulcus. Nevertheless, two of the most-commonly used cortical atlases in neuroimaging studies are based on cortical folding patterns [33, 34]. While important

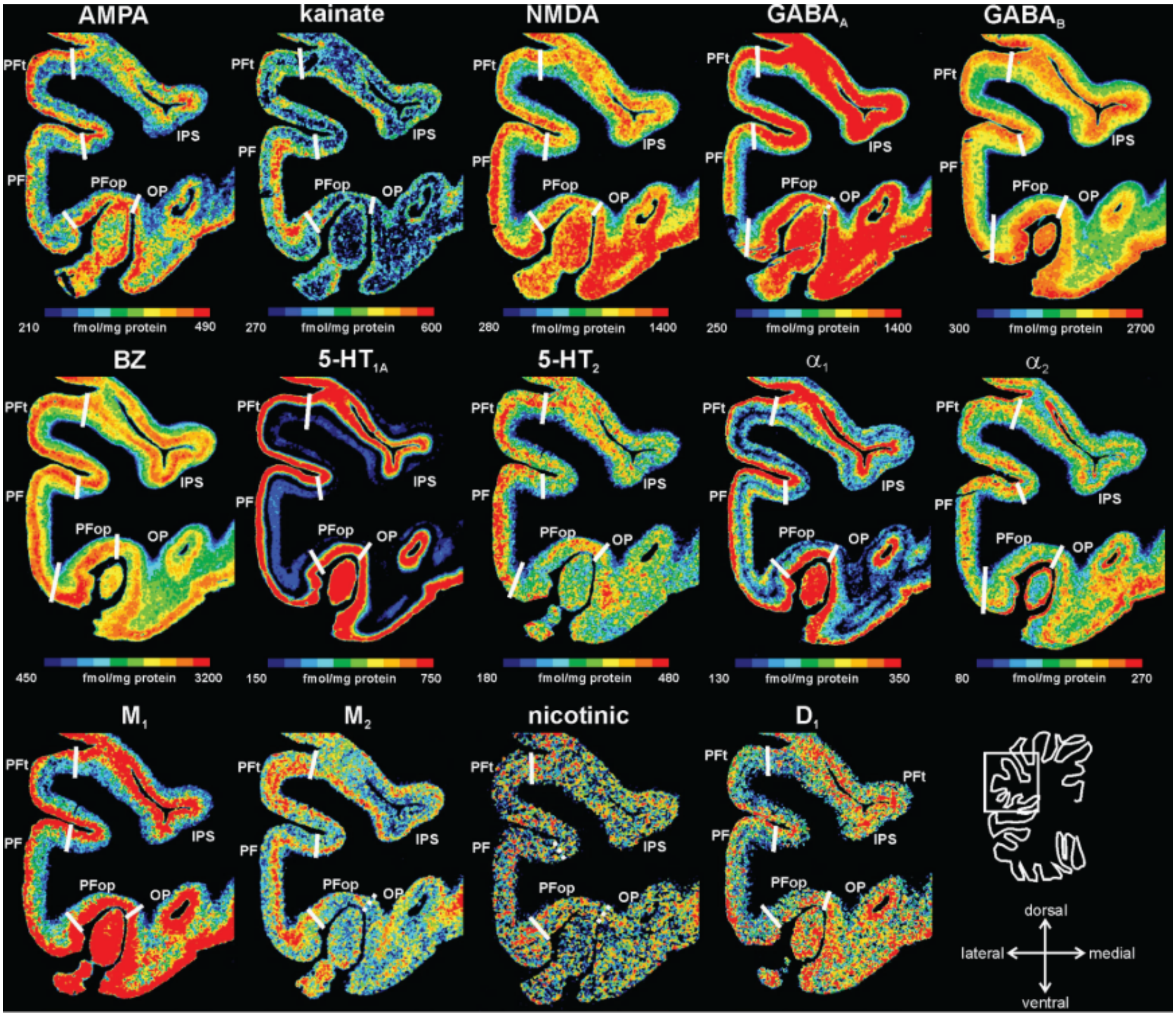


Figure 1.2: Receptor distributions in the inferior parietal lobule for areas PF, PFop, and PFt for 14 unique receptors [5]. Boundaries between putative cortical areas are marked by the white lines.

insights have been gleaned about the structure and function of the brain using these atlases, they represent a somewhat anachronistic approach to neuroimaging analyses, in that the cortical areas defined in these atlases mix signals from what would otherwise be classified as discrete, independent cortical areas on the basis of *in vitro* features.

In [35], the authors refine the definition of a cortical area as: "a distinctive region of cortex that differs reliably from neighboring areas in one or more neurobiological properties from four basic categories: function, architecture, connectivity, and/or topographic organization". Until recently, studies were only able to analyze cortical areas on the basis of architecture, connectivity, and topography. Following advances in computing hardware and software, neuroscientists been able to examine cortical *function*, with attempts to delineate cortical areas on the basis of *in vivo* features such as those computed using electrophysiology, resting-state and task-based fMRI (rfMRI and task-fMRI, respectively), and diffusion tensor imaging (DTI). These studies expand on Brodmann's initial hypothesis that discrete areas of the cortex are associated with specific cognitive functions, and assume that boundaries between cortical regions will be characterized by either changes in histology, structural organization, or functional organization. Task-based fMRI studies typically identify cortical subregions on the basis of linear contrasts, where participants are instructed to perform various tasks while in the MRI scanner. Researchers can then examine where cortical activation profiles are statistically different between tasks, or between task and control, in order to glean insight into which cortical areas are more tuned to specific cognitive functions.

The majority of these recent studies have approached the cortical mapping problem from one of two perspectives: 1) clustering and 2) spatial gradients. Superficially, these approaches are similar, in that they attempt to delineate boundaries between cortical areas by grouping cortical voxels together on the basis of similarity between their features. However, spatial gradient-based methods explicitly define areal boundaries, while clustering-based methods define these boundaries implicitly.

Given a scalar field, $f \in \mathbb{R}^N$ distributed over the cortical manifold, gradient-based approaches compute a multi-dimensional spatial derivative, $\nabla(f)$, of the scalar field at every location on the manifold [6, 9, 26]. The spatial gradient represents how "quickly" the values of the scalar field change in a local region. The hypothesis driving these gradient-based approaches is that, for a signal measured at every location on the cortical manifold (such as myelin density or cortical thickness) the values of this feature will

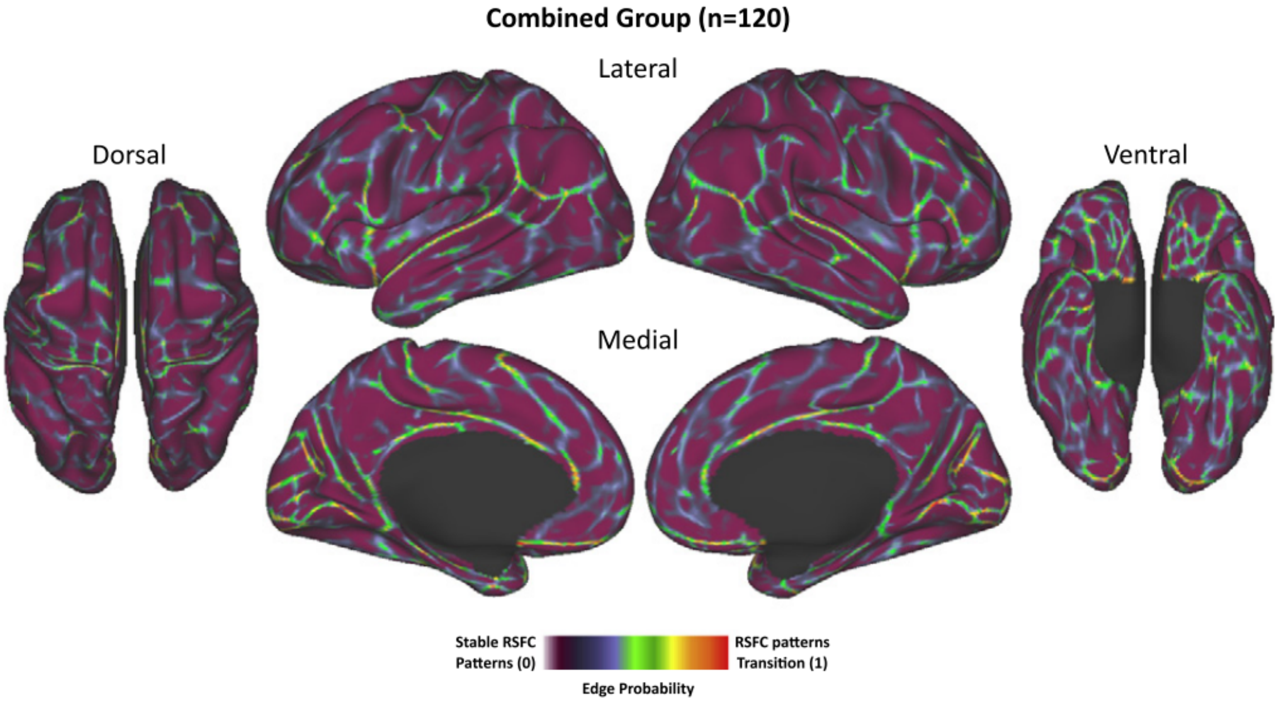


Figure 1.3: A spatial gradient map of resting-state functional correlations [6]. This is a group-average map from 120 subject-level gradient maps, and demonstrates the likely location of cortical boundaries on the basis of resting-state fMRI.

be relatively homogenous in a local area of the cortex. This is similar to the approach that Brodmann, Vogt and Vogt [28], and Mountcastle [36] took via cytoarchitectonics. The magnitude of the spatial gradient within a cortical area will be small – e.g. slowly changing – because the scalar field values are similar to one another. However, at the boundaries of cortical areas, we expect the scalar field values to systematically change in magnitude (see Figure 1.3). As such, these approaches infer the existence of a cortical boundary based on the magnitude of the spatial gradient, and define these boundaries where the gradient is large. Spatial-gradient-based approaches are "spatially aware" in the sense that, in order to compute the gradient of a biological measure of any location on the cortex, the algorithm needs information from the immediately adjacent neighborhood of that cortical location.

Clustering algorithms are unsupervised approaches that "group" individual data points

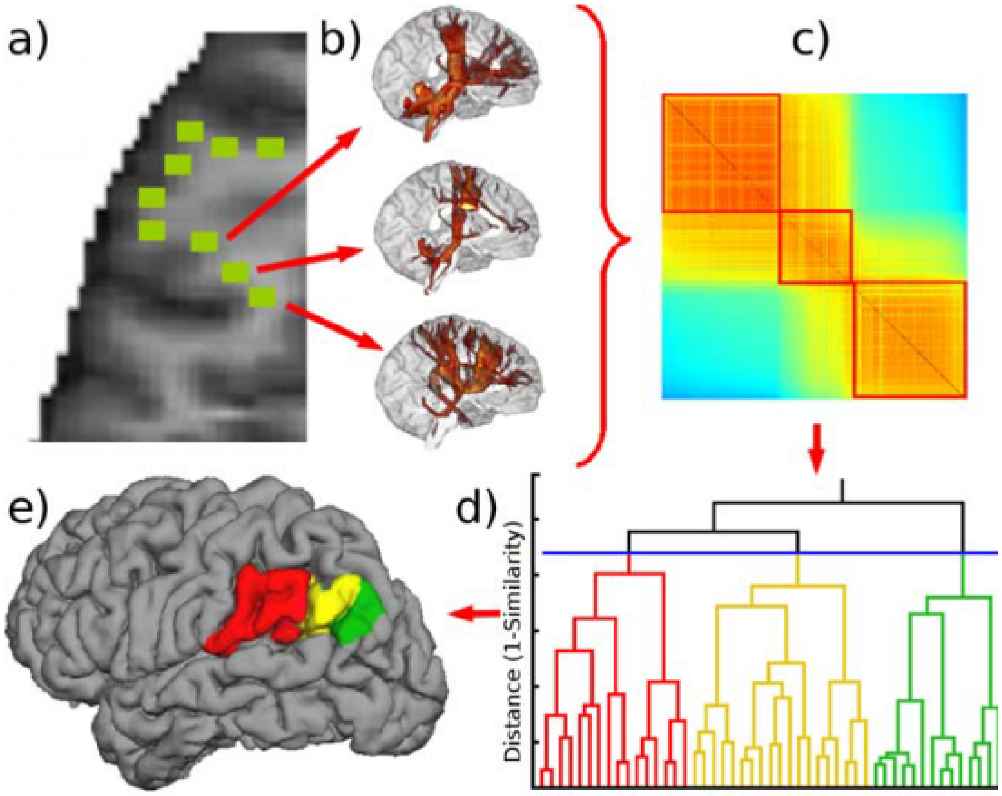


Figure I.

Figure 1.4: Clustering results applied to resting-state functional connectivity, computed using hierarchical clustering by [7].

together on the basis of the distributional properties of the data features. Data points with more similar feature vectors are more likely to be assigned to the same group, or cluster, while data points whose features are more dissimilar are likely to be assigned to different clusters (see Figure 1.4). A variety of clustering-based approaches have been proposed to map the cortical surface, including mixture models [37], hierarchical clustering approaches [7], graph-based methods [38, 39, 40, 41], and non-parametric algorithms [42, 43]. These approaches assume that nearby locations on the cortical manifold will display similar feature profiles. For example, areas in the primary somatosensory cortex are characterized by myelin and cortical thickness profiles that are different from the profile of the directly adjacent primary motor cortex. Clustering algorithms induce a discrete partition of the cortex into subunits, where subunits are composed of voxels or vertices distributed over the cortical manifold. Boundaries between discrete subunits represent implicit boundaries between unique areas. These algorithms can be applied to any feature type, such that the learned subunits can represent partitions of cortical function, architecture, connectivity, or topographic organization.

While generally not enforced by default, spatial contiguity of computed parcels can be used as a form of internal validation to evaluate the merits of a parcellation result. Some clustering algorithms do not enforce any "spatial contiguity" constraints on the resulting partition. When assigning a given sample of data to cluster or cortical area, many clustering algorithms do not account for the fact that points that are *anatomically* close to one another – e.g. in space – also tend to have more similar biological signals. A consequence of this is that a computed cluster might be composed of cortical voxels or vertices that are spread out across disparate areas of the cortex, or might have noisy, discontinuous areal boundaries. In order to conform to the idea that brain signals are locally stationary, many clustering algorithms can be modified to account for spatial proximity of data points along the cortical manifold, and thereby enforce spatially contiguity, or compactness, of the resulting clusters.

While it is understood that the human brain displays considerable variability in terms of size, shape, functional activation profiles, and topographic organization across individuals, there is also remarkable consistency. For example, work by [8] on developing a population-level immunohistochemistry atlas of the human cortex is only possible because the cytoarchitectural and receptor organization properties are reasonably spatially consistent across individuals. This enables researchers to generate accurate spatial probability maps of individual cortical areas [44] (Figure 1.5). Cortical voxels with similar cytoarchitectural and receptor profiles across individuals are assigned high probabilities of belonging to a certain cortical area, while those with more individual variability are assigned low

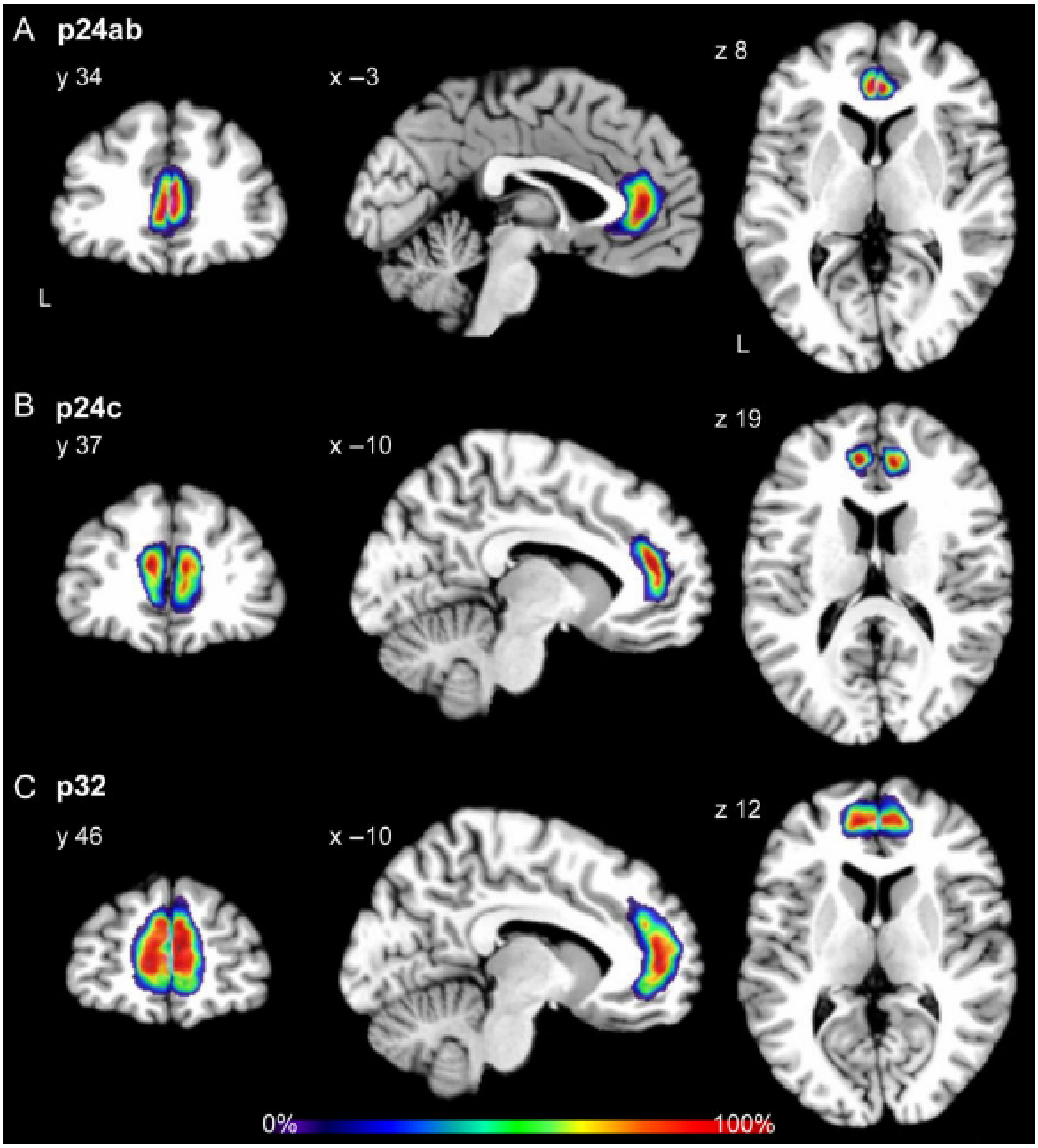


Figure 1.5: Maximum probability maps of anterior cingulate cortex areas p24a, p24b, and p24c, defined using receptor expression profiles from 10 post-mortem brains [8].

probability. In this way, we can begin to analyze and understand which areas of the cortex show homology across individuals, and why some are more variable.

The two general approaches described above, gradient-based algorithms and clustering algorithms, represent only a single step in the cortical segmentation pipeline, in that they provide an avenue for *computing* cortical maps. In their most naïve application, these algorithms are utilized to generate cortical segmentations, *one subject at a time*. However, these tools do not account for redundancy and consistency of neurobiological measures within the population. Each incoming dataset is treated as its own entity, even though, as mentioned above, biological measures are quite consistent across individuals. Similarly, the outputs of these approaches are "standalone" – individual parcels cannot easily be "matched" or aligned across individuals. Most studies perform *ad hoc* post-processing to match computed parcels or areas across subjects. Given this, clustering and gradient-based approaches are quite limiting. Further, many of the described algorithms are computationally intensive – for example, non-parametric approaches such as those presented in [42] [43] rely on Gibbs samplers to estimate the posterior distributions on vertex label assignments, that become nearly intractable as the number of samples increases. Rather than reapplying these algorithms to parcellate the cortices of new data each time a new MRI sequence is acquired, more recent studies have approached the cortical mapping problem from a different perspective. Given a pre-existing map or set of maps, prior work has examined how can one "apply" already-existing maps to unmapped data. The majority of these approaches rely on volumetric or surface registration, whereby new data is spatially normalized to a template space to achieve spatial correspondence between the template data (where the pre-computed maps exist) and the target data (for which a new map is to be computed) [14, 45, 46, 47, 48]. More recent segmentation work has incorporated label-fusion techniques that "aggregate" information from multiple pre-existing templates using statistical priors on label assignments of individual voxels [49]. These approaches are more data-driven, and leverage the similarity of pre-existing images with new images to drive the segmentation. However, these studies have primarily limited to volume-based data, while more recent neuroimaging studies have been leveraging surface-based data representations.

The bulk of this dissertation is devoted to the development of tools to map the cortices of new, previously unseen brain images. These tools leverage the statistical properties of images for which cortical maps have already been generated to "guide" the mapping of new images. As mentioned above, these methods should account not only for subject-level variability in neurobiological features used to compute these maps, but also for similarity of subject-level these features within a population. Fundamentally, this problem can be described as a "prediction" problem: given some pre-computed set of neurobiological features and a set of pre-existing maps, can we "predict" the cortical map for a new subject in such a way that it not only captures the subject-specific variability, but also conforms to a population-level representation? The methods discussed in this dissertation can be broadly decomposed into two groups: methods that are *spatially unaware* and methods that are spatially aware. We show that those methods that are able to utilize spatial information are considerably more robust than those that do not. Similarly, we'll examine how the choice of neurobiological features used to drive the predicted cortical segmentations impacts the final maps – we'll see that some features are inherently more "informative", and that choices made during the pre-processing steps to compute these features have measurable impacts on segmentation quality.

In the second part of this dissertation, we discuss analyses related to the connectivity organization of the cortical parcels themselves. While we can develop tools to discretize the cortex on the basis of any number of biological features, these discrete subunits do not tell the whole story of brain organization, as demonstrated in Figure 1.2. We'll delve into the idea of "continuous" functional connectivity and explore how these continuous maps are related to functional networks. We then examine relationships between pairs of cortical areas and show that we can identify systematic principles of brain organization that are related to core cognitive processes of the human brain.

Chapter 2

Data

For the majority of analyses in this document, we use data from two open-source neuroimaging consortia: the Human Connectome Project (HCP) [11] and the Midnight Scan Club (MSC) [50]. Both projects collected high-quality functional, diffusion, structural, physiological, and demographic data on a set of young, healthy adult participants. Both the HCP and MSC datasets have become de facto datasets for developing and validating novel neuroimaging methods.

2.1 Datasets

2.1.1 Human Connectome Project dataset

The HCP consortium collected data on a set of 1200 young adult subjects 21-35 years of age. The HCP acquired high-resolution 0.7mm isotropic T1w (TI=1000ms, TR=2400ms, TE=2.14ms, FA=8°, FOV=224mm, matrix=320, 256 saggital slices) and T2w images (TR=3200ms, TE=565ms, FOV=224mm, matrix=320). T1w and T2w data were preprocessed using a custom pipeline developed by the HCP [11] using FreeSurfer [51] to generate highly refined cortical surface meshes at the white/gray and pial/CSF interfaces.

The surface meshes were spatially normalized to Montreal Neurological Institute (MNI) space and resampled to have 32k vertices. The pipeline also generated four surface-based scalar maps: cortical thickness, Gaussian curvature of the cortical manifold, sulcal depth of the cortical gyri and sulci, and a myelin density map characterizing the spatially-varying myelin content of the gray matter [52].

For each subject the HCP acquired a single high-resolution diffusion-weighted image: voxel size: TR=5.5s, TE=89ms, multi-band factor=3, $\Delta = 43.1$s, $\delta = 10.6$ms, q-space sampling: 3-shell HARDI, b=1000,2000,3000s/mm^2, voxel size: 1.25 x 1.25 x 1.25 mm. Six b_0 images were collected, three with one phase encoding direction, and the other three with the opposite phase encoding direction. EPI distortion was estimated using FSL's *topup* tool, along with motion parameters, and the eddy-current induced field inhomogeneities. The estimated b_0 EPI distortions and field inhomogenieties were corrected using FSL's *eddy* tool, followed by motion correction, and lastly by gradient non-linearity correction. Images were transformed to the native structural space using a 6-DOF boundary-based registration. Gradient vectors were rotated according to this transformation.

For each subject, the HCP acquired four resting-state functional MRI (rs-fMRI) images: TR=0.720s, TE=33ms, multi-band factor=8, FA=$52°$, FOV=208mm x 180mm, Matrix=104 x 90 x 72, voxel size: 2 x 2 x 2mm. The authors refer to these four acquisitions as: `REST1_LR`, `REST1_RL`, `REST2_LR`, `REST2_RL`. The images were acquired over two separate days, such that `REST1_LR` / `REST1_RL` were acquired on one day, and `REST2_LR` / `REST2_RL` were acquired on another. Each session acquired 1200 time-points, such that each BOLD session was roughly 15-minutes in length. These images were pre-processed using a custom pipeline developed by the HCP [11]. BOLD images were denoised using subject-ICA [53] and FIX [54] to automatically identify and remove spurious noise components, and motion parameters were regressed out. No additional global signal regression, tissue regression, temporal filtering, or motion scrubbing were performed. Denoised voxel time series were interpolated onto the fsaverage_LR32k surface mesh using a barycentric averaging algorithm, and then smoothed at FWHM=2mm to avoid the mixing of signals across gyri. Surface-mapped BOLD signals were brought into register across subjects using a multi-modal surface matching algorithm [14] to the fsaverage_LR32 space and vectorized to CIFTI format, mapping each surface vertex to an index in a vector (towards the end of this work, we learned that different HCP data releases were processed using different versions of this surface registration algorithm; we discuss this in more depth in Section 4.5.5). CIFTI vector indices, referred to as "grayordinates" by the HCP, are in spatial correspondence across subjects (i.e. index i in subjects s and t correspond to roughly the same anatomical location), such that each subject shares the same mesh topology and adjacency structure. Time-series for each session were demeaned and temporally concatenated.

The HCP consortium developed a pipeline to generate high-resolution multi-modal cortical parcellations (MMP) with 180 cortical areas using a spatial derivative based algorithm [9] computed from resting and task-based fMRI signals, cortical thickness, myelin content, and cortical curvature (Figure 2.1. Manual editing was performed on the group-average gradient-based parcellation to ensure that boundaries conformed across feature types. Using a set of 210 independent subjects as training data, the authors trained a 3-layer neural network model to learn these boundary-based regions. The authors trained 180 classifiers, one for each cortical area, to distinguish a single cortical area from its immediately adjacent neighborhood (using a 30mm radius neighborhood size) in a binary classification setting. At test time, the authors compared the probabilities of the predicted areal class across all classifiers in a single find-the-biggest operation. Label predictions were regularized to minimize spurious predictions and "holes" in the final parcellation. Apart from the 30mm radius around each group-level area, the classifiers did not incorporate any spatial information at training or test time. Predictions generated from subjects in the training set were used to compute a group-average multi-modal parcellation which can be freely downloaded here: https://balsa.wustl.edu/DLabel/show/nn6K. The classifier

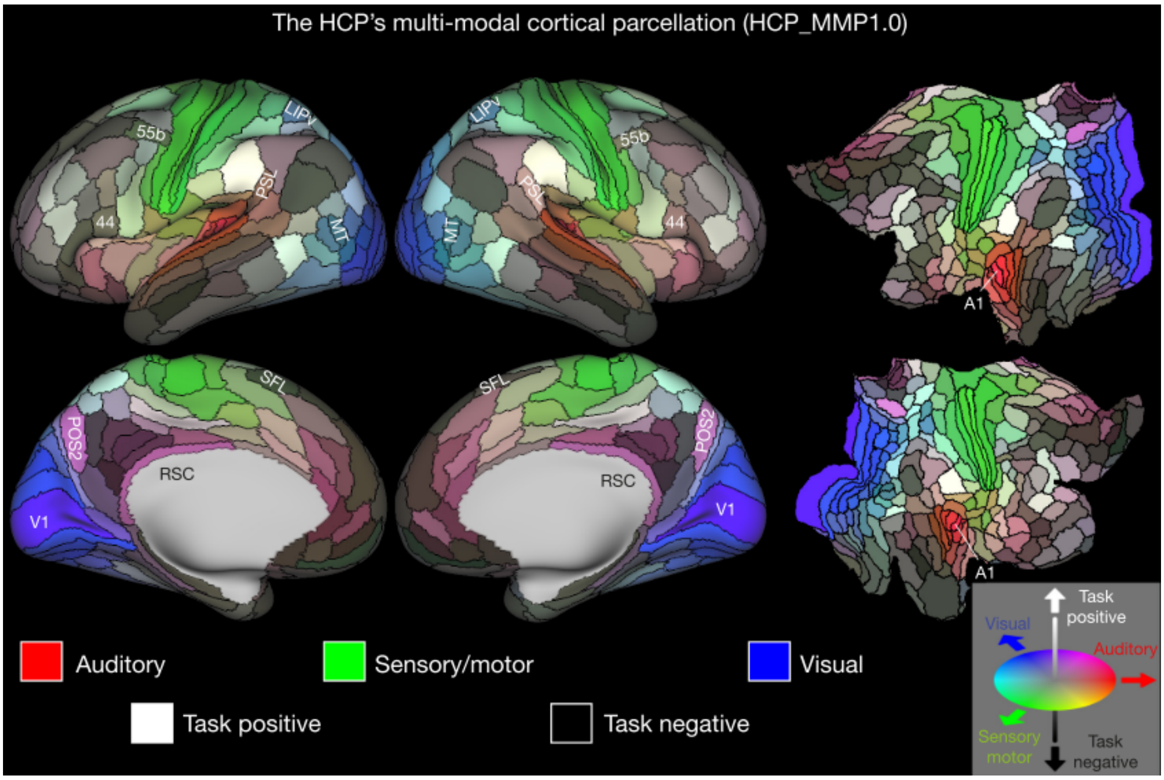


Figure 2.1: HCP-MMP cortical atlas developed by [9].

used to compute this parcellation has not yet been released to the public.

For our analyses in Chapters 3, 4, and 6, we utilize both the group-average HCP-MMP [9], and the subject-level cortical parcellations generated by the HCP. Subject-level parcellations for a subset of 449 subjects were made available by an HCP investigator. However, these subject-level parcellations have not yet been made available for public use.

2.1.2 Midnight Scan Club dataset

The Midnight Scan Club (MSC) study collected MRI data from ten subjects (5 female) ranging in age from 24-34 years of age. The MSC dataset is available for download here [50]. The MSC study acquired 5 hours of resting-state fMRI data on each participant across ten 30-minute acquisitions, with the goal being to develop high-precision, individual-specific functional connectomes to yield deeper insight into the reproducibility and inter-subject differences in functional connectivity.

The MSC dataset preprocessing followed a roughly similar pipeline to that of the HCP dataset. Four 0.8mm isotropic T1w images (TI=1000ms, TR=2400ms, TE=3.74ms, FA=$8°$, matrix=224, saggital) and four 0.8mm isotropic T2w images (TR=3200ms, TE=479ms, matrix=224 slices, saggital) were acquired. T1w images were processed using FreeSurfer to generate refined cortical mesh representations of the white/gray and pial/CSF tissue interfaces, which were subsequently warped to the fsaverage_LR brain surface using the FreeSurfer shape-based spherical registration method, and resampled to 164K and 32k vertex resolutions. The authors performed myelin mapping by computing the volumetric T1/T2 ratio and interpolating the voxel-wise myelin densities onto the 32k surface mesh.

MSC resting-state data were acquired using single gradient-echo EPI sequences with the following parameters: TR=2.2s, TE=27ms, FA=$90°$, voxel size=4 x 4 x 4mm. The MSC applied slice timing correction, and distortion correction using subject-specific mean field maps. Images were demeaned and detrended, and global, ventricular, and white matter signals were regressed out. Images were interpolated using least squares spectral estimation and band-pass filtered (0.009Hz $<$ f $<$ 0.08Hz), and then scrubbed of high-motion volumes. Denoised volumetric resting-state data were then interpolated onto the midthickness 32k vertex mesh. The MSC study did not perform subject-ICA and FIX to remove spurious noise components from the temporal signals.

2.2 FreeSurfer outputs

Freesurfer [51] is a conventional pre-processing tool applied to MRI imaging datasets, that generates a variety of transformed images and metrics computed from raw structural T1 and T2 images. For the analyses described in this dissertation, we primarily used

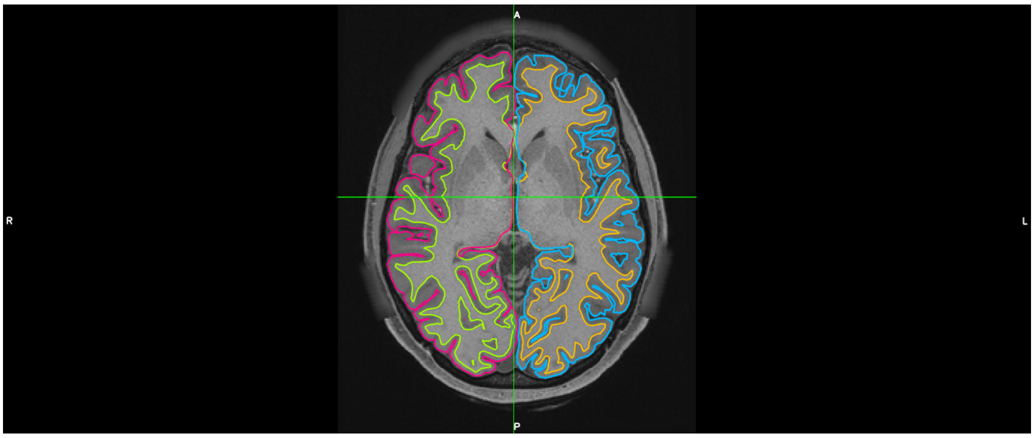


Figure 2.2: Gray/white interface (green, orange) and pial/csf interfaces (red, blue) surface meshes, overlaid on the structural MPRAGE image. In both the left and the right hemisphere, the white, midthickness, and pial surfaces all have 32k vertices. Vertices between meshes are in correspondence with one another.

2 variants of these outputs: folding-based cortical segmentations, and surface-mesh representations of the gray/white and pial/CSF boundaries.

FreeSurfer generates 3 unique cortical atlases, defined by the Brodmann, Desikan-Killiany, and Destrieux cortical atlases [27,33,34], defined in both volumetric-space and in surface-space. For a given atlas, each volume voxel or surface vertex is assigned a label in $\{1 \ldots L\}$, where L is the total number of labels in the parcellation – all vertices assigned to the same label correspond to a discrete, contiguous cortical area. All of these discrete cortical areas delineate the cortex into groups, or "parcels".

FreeSurfer also outputs two unique surface mesh files, corresponding to the gray/white interface, and the pial/CSF interface (Figure 2.2). These files are represented as a list of 3D vertices, e.g. coordinates in 3D Euclidean space, sampled at the tissue interfaces, and a list of faces, e.g. triangles composed of 3 unique, adjacent vertices at the tissue interface. Together, the vertices and faces represent a "surface" rendering of the tissue interfaces of the neocortex. For most signals of interest discussed in this dissertation (such as cortical thickness measures, BOLD signals, mylein density, etc.), volumetric data are interpolated onto these surface mesh vertices by taking a weighted spatial average of

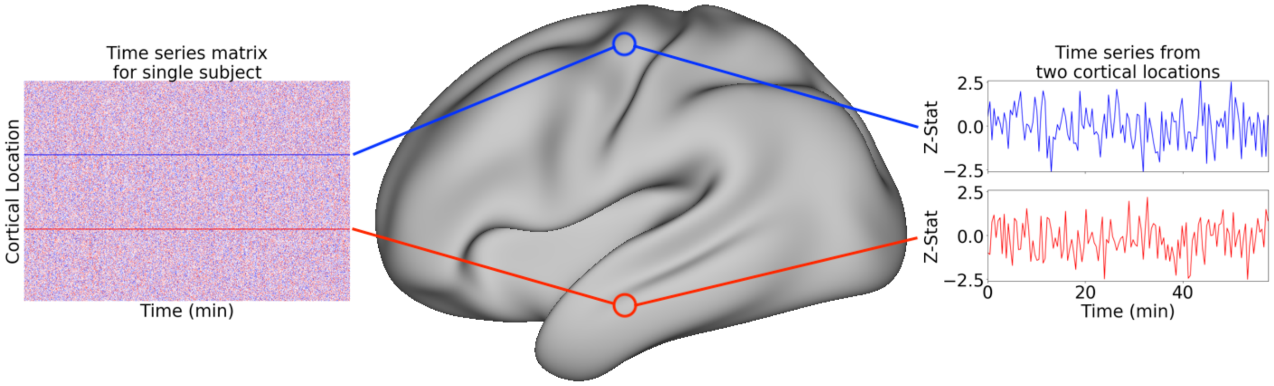


Figure 2.3: Voxel-wise time-series signals are mapped onto the brain cortical surface. Each vertex is represented by a vector of 4800 samples, corresponding to the measured BOLD signal at that vertex.

volumetric voxel signals close to the gray/white tissue interface (Figure 2.3).

2.3 Diffusion image pre-processing

For the analyses in Chapter 3, we utilized diffusion-weighted images (DWI) collected by the HCP consortium to generate probabilistic tractography connectivity profiles. For a single cortical white matter voxel, the DWI signal captures the diffusion of water molecules in response to an applied magnetic field ($b = 1000, 2000, 3000$) – these fields are applied at various directions sampled from the unit sphere. The DWI signal in response to an applied field is related to the amount of Brownian water diffusion, integrated over time. Given a set of spherical directions (angular samples), and the DWI signal for each angular sample, we can model the overall diffusion of water using a 3D probability distribution function that describes the likelihood that a water molecule will move in a certain 3D direction [10, 55, 56] (see Figure 2.4). In our analyses, we use the FSL tool `bedpostx` to model 3D distribution functions of water diffusion in the white matter [57].

Once the 3D probability distributions are fit, we can generate tractography profiles using the FSL tool `probtrackx2`, a probabilistic fiber tracking algorithm based on a

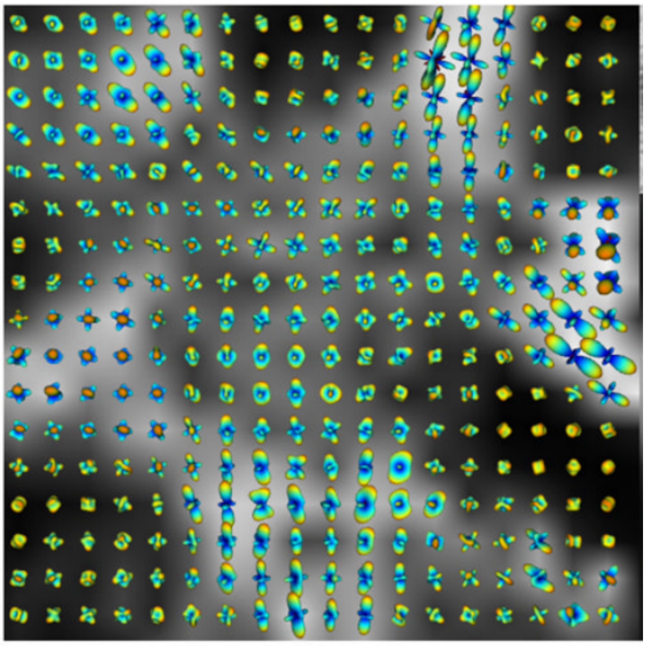


Figure 2.4: Orientation-distribution functions using Constant Solid Angle (CSA), from [10]

Markov Chain Monte Carlo (MCMC) sampling scheme [58].

At each step in Algorithm 1, we sample a direction from the current distribution of diffusion directions, take a step in the sampled direction from our current position, and then update our position and distribution function. This generates a series of random steps, starting at the seed point, and ending somewhere else in the brain. Since we are sampling directions from a probability distribution, it is unlikely that the sampling algorithm will generate the same sequence of steps more than once, meaning that we can explore the space of possible streamlines emanating from a seed point. We repeat this process 5000 times for each seed point, for all vertices in the surface mesh.

This is inherently a stochastic and iterative procedure. Estimated streamlines will vary from run to run of the MCMC algorithm. However, because water diffusion is typically constrained along certain axes (that is, along neurons), the sampled directions will generally not be uniformly random. As such, due to the Central Limit Theorem, we expect to see some limiting distribution of streamline paths, whereby we can determine

Algorithm 1 fiber tracking with FSL's probtrackx2

```
function probtrackx2() :

    initialize seed location $p = (i, j, k)$, diffusion PDF $f_{\text{dir}}(p)$
    sample initial tracking direction $d \sim f_{\text{dir}}(p)$

    while stopping criteria not met do          ▷ criteria include crossing mesh
        $p \leftarrow \alpha * d + (i, j, k)$                    ▷ update position
        $d \sim f_{\text{dir}}(p)$                          ▷ sample new direction
    end while

end function
```

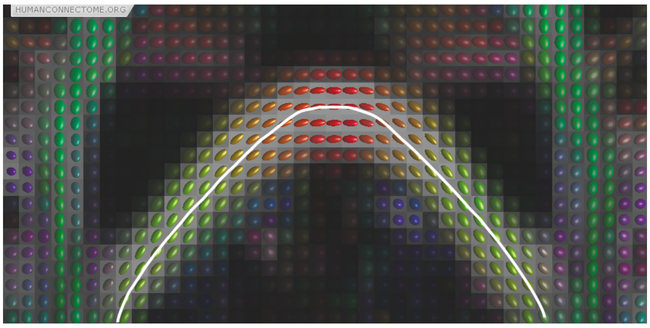


Figure 2.5: Example of a streamline propagated along orientation distribution functions, as shown in [11]. At each step of streamline tracking, a direction is sampled from an orientation distribution function, and propagated a small step forward in this direction.

the strength of connectivity between any two pairs of cortical areas. Analogous to defining the Pearson correlation matrices from the resting state time series, for a given run of probtrackx2, we define an *asymmetric* matrix of streamlines as the real-valued function $\mathbb{T} : \Omega \times \Omega \mapsto \mathbb{R}^+$ by $(x, y) \mapsto T(x, y)$, where Ω is the cortical surface, and $T(x, y)$ is the number of connections (strength of connectivity) from cortical location x to cortical location y.

2.3.1 Functional network assignments of cortical areas

For the remainder of this document, we work directly with the subject-level and group-average lever representations of the Human Connectome Project multi-modal parcellation (HCP-MMP; Figure 2.1) [9]. As described above, [9] computed a group-level parcellation and learned subject-level maps using a connectivity-based linear perceptron model. The group-level map contains 180 cortical areas, while the subject-level maps contain at most 180 cortical areas – not every region identified in the group-level map was "identified" in the subject-level data.

Throughout the remainder of this document, we also refer to an atlas, that we refer to as the "Demirtas atlas", in which each of the 180 cortical areas in the HCP-MMP are assigned to a functional network (Figure 2.6) [16]. We divide these functions into two overarching categories: primary sensory networks (Auditory, Somatomotor, and Visual) and higher-order association network (Cinguloopercular, Dorsal and Ventral Attention, Frontoparietal, Default Mode, and Other). While each cortical area tends to engage and activate in many different functions and settings, each area tends to *preferentially* engage in a limited number of settings. The Demirtas atlas is an assignment of each area in the HCP-MMP atlas to the function of preferential engagement. Regions labeled "Other" were those that could not be readily assigned to any specific functional network. Table 2.1 details the number of HCP-MMP cortical regions assigned to each functional network.

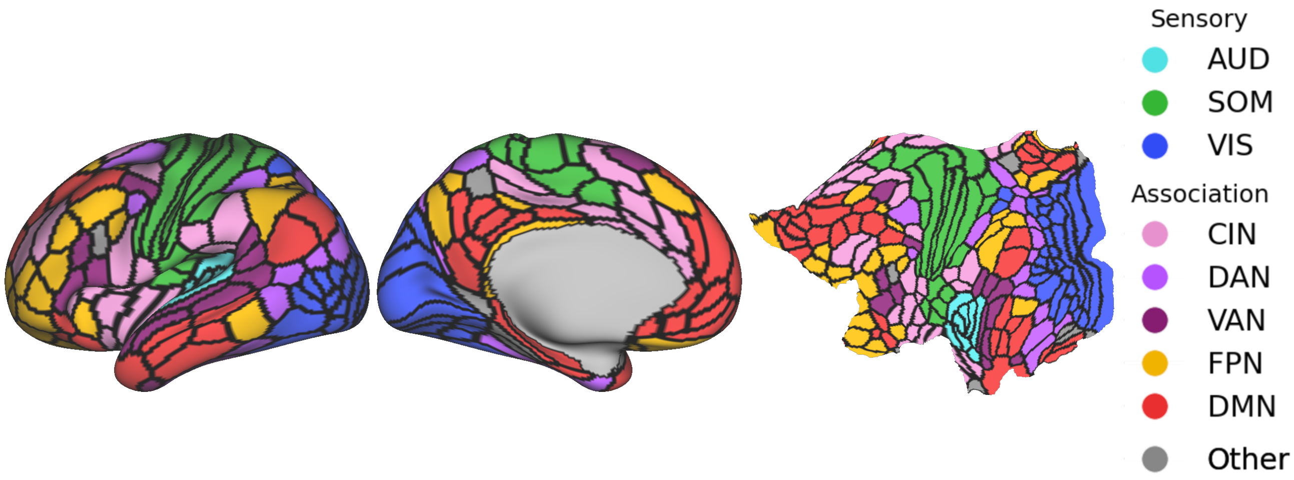


Figure 2.6: Functional network assignments of each cortical area in the HCP-MMP atlas.

Table 2.1: Number of HCP-MMP regions per functional network assignment in the Demirtas atlas. "Other" regions could not reliably be assigned to any specific functional network.

Network	**Number of Regions**
Sensory	
AUD	8
SOM	20
VIS	28
Association	
CIN	27
DAN	16
VAN	13
FPN	26
DMN	36
Other	6

Chapter 3

Cortical mapping using spatially-unaware classifiers

Eschenburg, K, Haynor, D, Grabowski, T. Automated connectivity-based cortical mapping using registration-constrained classification. SPIE, 2018.

Initiated by Kristian M. Eschenburg (KME) as a part of, and funded by, NSF grant BCS-1734430 for Principal Investigator Dr. Thomas Grabowski (TG): "Relationship of Cortical Field Anatomy to Network Vulnerability and Behavior". KME conceptualized this study, developed the code, performed the analyses, and wrote the bulk of the document. DRH and TJG provided neuroscientific and technical guidance for this work and contributed to the editing and organizational structure of the manuscript.

3.1 Introduction

As described in Chapter 1, a fundamental assumption in neuroscience is that the human neocortex can be compartmentalized into discrete subunits on the basis of various broad feature types including functional, architectural, connectivity, and topographic profiles [5, 59, 60]. This idea is based on seminal work by [36] in the 1950s that identified cortical columns in the feline somatosensory cortex. Cortical columns are "tubes" of neurons that are oriented perpendicularly to the cortical surface and spanning the width of the cortex from layers 2 to 5. Cortical columns act as independent computing units tuned to the same receptive field input [36]. These columns, by nature of receiving input from the same receptive fields, often cluster together to form small contiguous processing units [36, 61, 62]. Evidence has increasingly pointed to the idea that the cortex can be grouped into discrete, contiguous parcels. An early study by [27] examined cortical cytoarchitecture using Nissl staining and identified unique spatially varying histological properties across the cortex that could be grouped into larger relatively homogeneous units – these results are still a reference for cortical discretization for many research projects today.

With the rise of machine learning and high-performance computing, researchers have developed sophisticated computational methods to automate the formation and identification of these cortical groupings, a process termed "segmentation" or "parcellation". These algorithms are based on a variety of biological data, including structural and functional brain connectivity signals [7, 9, 39, 41, 51, 63]. Various groups have applied data-driven approaches to discovering the hidden substructure of the cortex using unsupervised clustering [26, 38, 43, 64]. These methods attempt to find hidden structure in the cortical data by grouping cortical voxels with similar properties into higher-order latent clusters.

Here, as well as in Chapter 4, we addressed a tangential, albeit related, problem: given a set of pre-computed cortical maps for a set of subjects (whether these were generated independently or through group-informed means is irrelevant here, so long as the parcels are in correspondence across subjects), can we map the cortex of a new subject, using the previously generate maps as training data? We proposed three approaches for mapping the brain surface of a new subject, using a training dataset of previously mapped brains and data regularly acquired in standard functional MRI scans, i.e. diffusion and resting-state data. Here, we generated a training data library in order to build a statistical model of the parcellated cortical surface that acts as spatial prior for mapping new data. We trained classifiers on training data sampled from local neighborhoods on the cortical surface, using features derived from training brain connectivity information, and apply these classifiers to map the surfaces of previously unseen brains. We demonstrated the performance of 3 different classifiers, each trained on 3 different types of training features, to accurately predict the map of new brain surfaces.

We processed a set of high-quality MRI imaging data acquired by the Human Connectome Project (HCP) [11] along with a set of labels that were derived for the HCP data using resting-state fMRI, task-based fMRI, and various cortical surface scalar measures and vetted by trained neuroanatomists [9]. We treated the HCP fMRI-connectivity-based labels as the dependent variables, and trained classifiers to distinguish between each label, based on the characteristics of the functional and structural connectivity data associated with it. We applied these classifiers to previously unseen subjects in the HCP dataset, and characterized the accuracy of the new parcellations as a function of classifier type, input feature data, and various model parameterization choices.

3.2 Data

We used a set of 85 randomly chosen subjects acquired by the HCP in their Young Adult cohort [11]. We were specifically interested in the resting-state and DWI data acquired for each subject. These data were preprocessed in accordance with steps described in Section 2.1.1.

The minimal preprocessing steps performed by the HCP included running FreeSurfer [51, 65] to generate a white matter surface mesh, which was then spatially resampled to a standard space with 32k vertices for all subjects. We used this mesh throughout our analysis. The voxel-wise resting-state BOLD signals were interpolated onto the mesh. The HCP also released a group-average connectivity-derived parcellation (CBP), with roughly

180 regions per hemisphere, computed from boundary maps derived from resting-state, task-based fMRI, and various cortical scalar metrics [9]. We used these labels, taking values in 1, 2, . . . , 180, as the dependent variable in our three classifiers.

3.3 Methods

3.3.1 Generating connectivity matrices

As mentioned in Sections 2.1.1 and 2.1.2, the HCP studies aligned cortical surfaces to the fsaverage_LR surface space. The result is such that, given two meshes S and T, the anatomical location of grayordinate i in mesh S corresponds to generally the same anatomical location as grayordinate i in mesh T, allowing for direct comparisons between the same grayordinates across individual surfaces.

In cases where spatial normalization of surfaces has not been performed, it would be incorrect to assume that two grayordinate indices correspond to the same anatomical locations across subjects. In order to alleviate the requirement of explicit vertex-wise correspondence across training, validation, and testing datasets, we assumed that most imaging studies will first run FreeSurfer to generate subject-specific folding-based cortical parcellations [33, 34]. We could then aggregate the high-dimensional vertex-wise connectivity features over one of these cortical atlases, and simultaneously reduce the feature vector dimension. This guarantees that column indices of feature vectors represent anatomically comparable variables across individuals corresponding to connectivity to whole cortical areas rather than explicit vertex-vertex connections. These low-dimensional vectors are agnostic to the original mesh resolution and degree of spatial normalization. As long as resting-state and diffusion-weighted data are collected for a given study, and that good spatial correspondence between the T1w, DWI, and BOLD images can be achieved, we can apply our processing steps to this data.

Given a BOLD time series matrix $B \in \mathbb{R}^{32k \times t}$ and cortical atlas with k regions, we considered the set of vertices assigned to region k and compute the mean time-series of region k as:

$$\hat{B}_{k,t} = \frac{1}{|k|} \sum_{i \in k} T_{i,t} \tag{3.1}$$

where $\hat{B} \in \mathbb{R}^{K \times t}$ is the matrix of mean regional time-series. We computed $R \in \mathbb{R}^{32k \times K}$, the Pearson cross-correlation between B and $\hat{B}$, where $R_{i,k}$ represents the temporal correlation between a vertex i and cortical region k.

Similarly, for the diffusion data, we applied FSL's `bedpostx` and `probtrackx2` tools [57, 58] to each DWI to model water diffusion in each voxel in the brain. We then fit the cortico-coritco structural probabilistic tractography matrix, S. Index i, j in this matrix describes the number of streamlines (strength of structural connectivity) from vertex i to vertex j. We reduced the dimensionality of the structural connectivity profiles over the same cortical atlas as the resting-state data, by computing the total streamlines between vertex i and region k. We took the log-transform generate log-normal count distributions in order to reduce the bias towards short-range streamline connections present in tracking algorithms.

These cross-correlation vectors and aggregate streamline count vectors are used as features to train our models.

3.3.2 Surface registration

The goal of our method is to parcellate the cortical surface of any subject, given its resting state or structural connectivity profiles – we wished to predict which cortical area any given vertex belongs to, on the basis of its connectivity. In the following analysis, we attempted to assign vertices to one of the 180 labels in the HCP-MMP parcellation. In this proposal, we predicted the cortical map of a new subject based on a set of template maps derived for the HCP parcellations with labels in 1,2. . . 180 – however, we could apply this same methodology to any set of template cortical maps with any number of unique labels. For example, we could also predict a new cortical map based on the Gordon parcellation with 178 areas per hemisphere [26], the Yeo cortical map of 7 intrinsic network functional areas [63], the Destrieux parcellation scheme with 75 unique regions [33], the Desikan-Killiany parcellation scheme with 35 unique regions [34]. One possible classification approach would be to train a classifier to distinguish all 180 HCP labels from one another in a one-vs.-all classification scheme. However, we could also incorporate prior information into this process. We constrained the label search space to a small, localized set of labels « 180. Using a non-rigid, spectral-feature based surface registration method [12, 66], we matched vertices on one surface to vertices on another on the basis of local similarity in surface geometry. We computed a weighted adjacency matrix of each surface, compute its graph Laplacian, L, and diagonalize L. The first k eigenvectors of L, along with the reduced-dimension connectivity vectors, are used as features in the point-set registration between the two surfaces. In Figure 3.1, the first few Laplacian eigenvectors are shown as spatial functions on the brain surface.

The registration step produces two directed matches, where for each brain pair (s_1, s_2), every vertex in s_1 maps to a single vertex in s_2, and vice versa, though the two mappings are only approximate inverses of each other (see Figure 3.2).

For two arbitrary surface meshes, s and t, we defined $f_{s,t} : s \mapsto t$ as the mapping of vertices from mesh s to mesh t. Without loss of generality with regards to cortical parcellation, ρ, of interest (here, the HCP-MMP map), we defined $L(v, s, p)$ as the "known" label of vertex $v \in V(s)$ in mesh s for template map ρ. For a training dataset with g training images, we could compute a total of $\eta = g \times g$ pairwise surface mappings. We computed the set of all pairwise mappings $F = \{f_1 \dots f_\eta\}$, between all pairs of images in the training set, where f_i is a mapping for one pair of training images (see Figure 3.3)

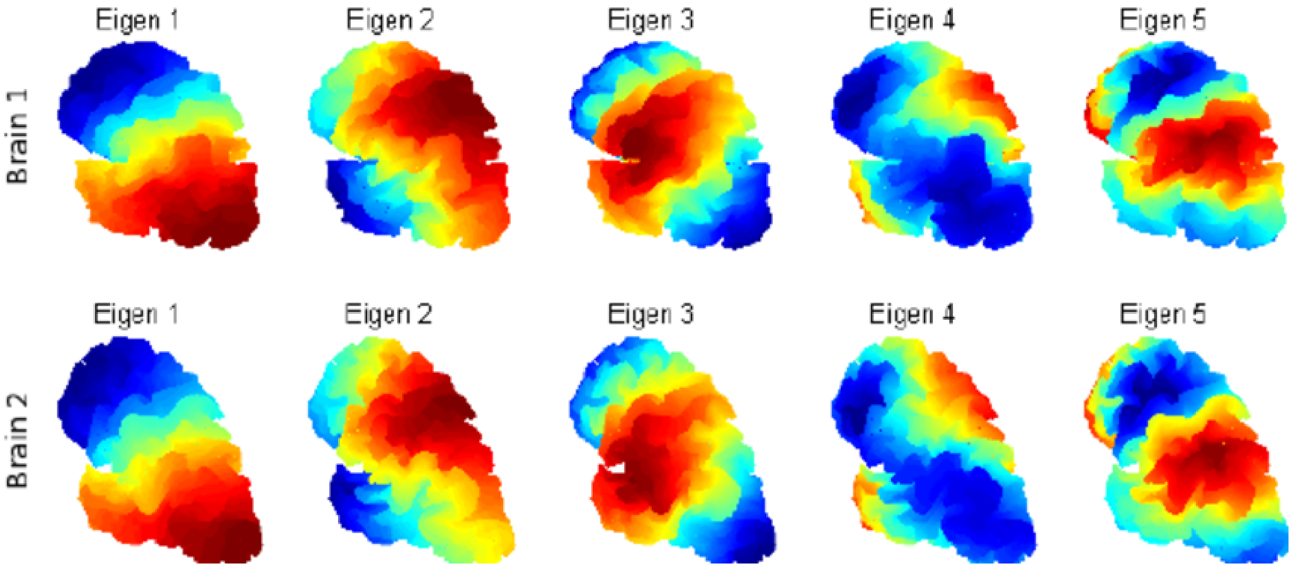


Figure 3.1: Laplacian embedding of cortical surfaces characterizing the intrinsic geometry of the mesh, as described in [12]. *Eigen 1* corresponds to the Fiedler vector, that attempts to assign constant values to areas with similar features. Higher-order vectors capture increasingly-complex geometric features of that surface.

3.3.3 Prior probabilities on label assignments

We defined two types of prior distributions:

1. in the **test** data, on the possible labels to which a new dataset might be assigned
2. in the **training** data, on how often any single label is likely to be incorrectly *confused* with any another.

To address **(1)**, we took the following approach. Given a new test dataset τ with its own surface mesh, the set $S = \{s_1 \dots s_g\}$ of g training surface meshes and set $P = \{\rho_1 \dots \rho_g\}$ of g training label maps, we registered the test data to each individual training image, generating g test-to-training surface mappings. Each vertex, v, in the test mesh maps to g (possibly non-unique) labels in the training set. We defined the set of labels to which a test vertex maps to as

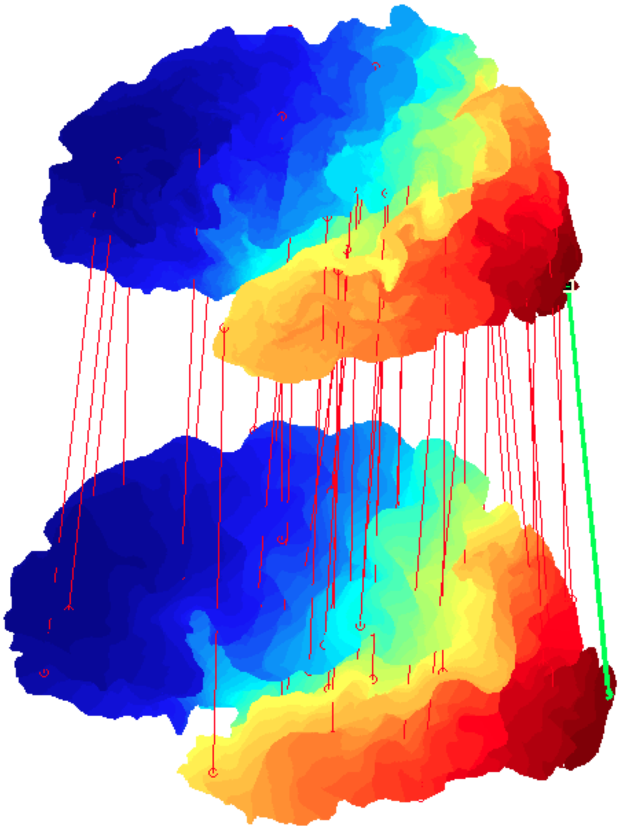


Figure 3.2: Illustrating pairs of corresponding nodes (red links) between two cortices. The mesh coloring is the magnitude of the Fiedler vector. The correspondence is found by identifying the closest nodes in the spectral embedding space [12]

$$st(v, S, F) = L(f_i(\tau(v)),\ s_i,\ \rho_i)\ \ \forall\ \ i \in 1 \ldots g \tag{3.2}$$

where $f_i(\tau(v))$ is the vertex in training image i to which the test vertex maps to. For each vertex v in the test brain, we computed a discrete probability distribution describing the empirical frequencies with which v maps to any label in cortical parcellation, ρ.

Figure 3.4.**a** shows the empirical prior distribution for a single vertex v from an exemplar test image (in this case, the true label of $v = 52$ is known *a priori*). For this test vertex, the empirical PMF assigns a high probability for area 52 i.e. test vertex v tends to

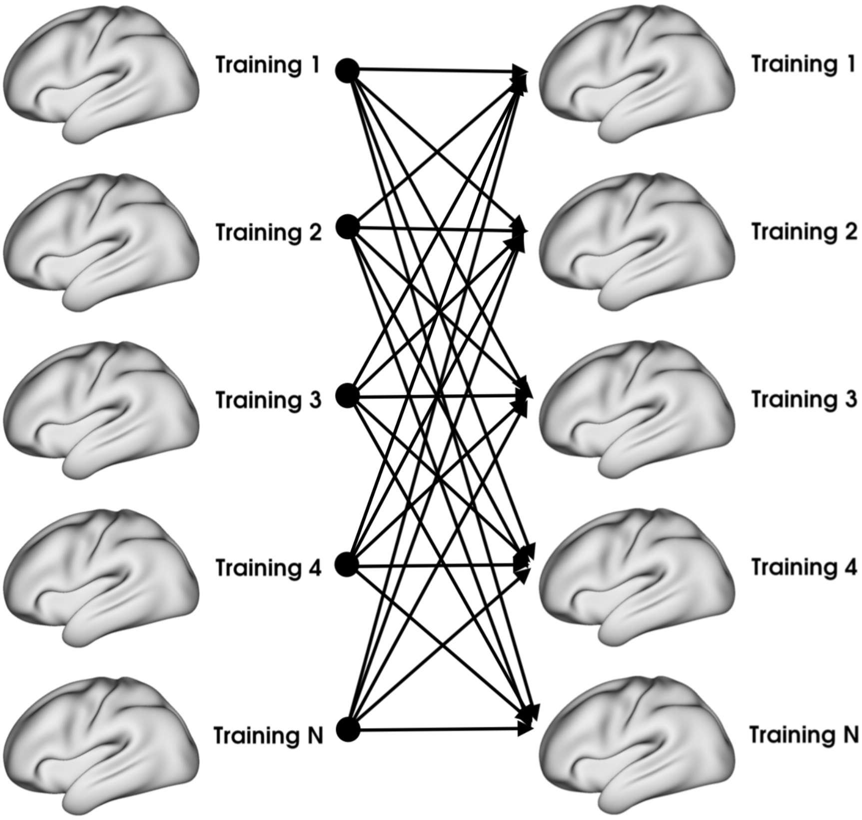


Figure 3.3: Pairwise mappings between all training image pairs. We used these mappings to compute empirical prior distributions on which labels are most frequently confused with each other.

map onto vertices with the same label as itself.

Most of the probability mass is centered on a small set of labels. Therefore, it would be inefficient to consider all labels as equally-likely candidates. For example, if we knew a vertex was in the temporal lobe, it would not make sense to consider labels associated with cingulate cortex as viable candidates. As such, $st(v, S, F)$ acts as a spatial prior on possible candidate labels that a test vertex might be assigned to.

In this analysis, there exist up to 180 possible regions in parcellation $\rho = \{1, 2 \ldots 180\}$ to which a test vertex could be assigned. This number varies, depending on which cortical atlas we are trying to predict (i.e. Gordon [26], Yeo [63], Desikan-Killiany [33],

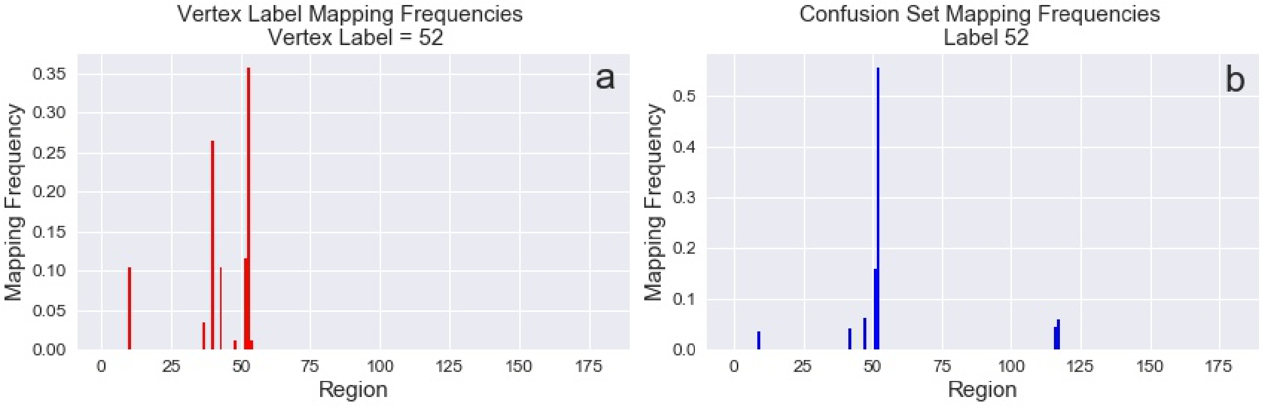


Figure 3.4: Vertex-to-label mapping frequencies for a vertex in a training brain. Here, the vertex label assignment is known *a priori*. **a)** probability mass function describing the frequency with which a single exemplar vertex with known label 52 from an example training image maps to other labels. **b)** probability mass function describing the frequency with which **any** vertex with known label 52 maps to other labels, aggregated across all training images.

Destrieux [34], etc.). It would be inefficient to train a classifier to distinguish all 180 regions from one another.

We proposed to compute a prior distribution of "confusion labels" – for a given label m, the confusion set defines the set of possible labels with which m might be incorrectly confused with based on surface registration alone.

$$vtx(m) = \{v \in V(s) \;\forall\; s \in S \;\mid\; \mathbb{L}(v, s, \rho) = m\} \tag{3.3}$$

as the set of vertices across all training images with a known label m. Given all pairwise registration mappings F between training images (Figure 3.3), for each vertex $v \in vst(m)$, we computed the empirical prior distribution of label assignments of v, as described above in **(1)**. We then aggregated these single-vertex prior distributions across all $v \in vtx(m$ as

$$lst(m) = \{st(v, s, F_s)\ \forall\ v \in V(s),\ s \in S\ \mid\ v \in vtx(m)\} \tag{3.4}$$

as the set of all labels to which any vertex in the training data set with label m maps to. The value $lst(m)$ is the aggregation of vertex-level prior distributions described in Equation 3.2, for all training vertices with label m. $lst(m)$ is a property of m and describes the likelihood of label m being misclassified as any different label.

Given a set of training data, we computed $lst(m)\ \forall\ m \in \rho$ (here $m = \{1 \ldots 180\}$). The m-th classifier is trained only on the training data corresponding to those vertices with labels in $lst(m)$. That is, we trained as many classifiers as there are labels in the parcellation map. This is analogous to work in [9], where the authors trained 180 independent connectivity-based multi-layer perceptron models to classify cortical areas. For each classifier, the training data consists of only those vertices belonging to the confusion set of that particular label.

At test time, we registered the test dataset t to all training images $s \in S$. For each vertex $v \in V(t)$, we computed $st(v, t, F_t)$ – i.e. the prior distribution on possible label assignments for v from Equation 3.2. We defined $m_{max} = argmax(st(v, t, F_t)$ as the label to which vertex v maps to most frequently. m_{max} is the label that maximizes the prior distribution density i.e. the label to which v maps to most frequently in the training data. We then assigned a label to v based on the output of the model for m_{max}. Additionally, we retained the probability that a test vertex v maps to any cortical region – we used these probabilities as a spatial prior on label assignments for a test vertex.

3.3.4 Models

We generated 3 sets of training features for each test vertex: 1) – regionalized resting state correlations, 2) regionalized log-normal tractography streamline counts, and 3) the concatenation of (1) and (2). For each of the 3 types, we appended 3 scalar maps computed using FreeSurfer: sulcal depth, myelin density, and Gaussian curvature. In total, we generated three feature sets: two of size $32k \times 75 + 3$ and one of size $32k \times 150 + 3$. We examined how model performance varies by feature type.

In the current analysis, we considered three unique algorithms to learn cortical maps from connectivity data: mixture models, decision trees, and linear feed-forward neural networks. Each of these approaches makes unique assumptions about the state of the feature data, but all three share one feature in common: none of the above approaches take into account the spatial relationships between data points. All methods assume that individual samples are spatially independent and statistically unrelated.

Gaussian Mixtures Models

Mixture models are a class of parametric statistical models that assume data points are drawn from a discrete set of independent subpopulations. In this analysis, we considered the case of Gaussian Mixture Models, where data points modeled as samples from independent multivariate Gaussian distributions.

Assume we are given N data points $x_i \in \mathbb{R}^p$ for $i = 1 : N$, such that $x_i = [x_{i,1}, x_{i,2} \dots x_{i,p}]$. A Gaussian Mixture Model (GMM) is parameterized by K, the number of subpopulation distributions. Each subpopulation, C_k, is distributed according to a p-dimensional Gaussian distribution, $\mathcal{N}_k$, parameterized by a mean vector $\mu_k \in \mathbb{R}^p$ and covariance matrix $\Sigma_k \in \mathbb{R}^{p \times p}$. The collection of these parameters are represented as vector $\Theta = [(\mu_1, \Sigma_1) ... (\mu_K, \Sigma_K)]$. For a given data point x_i, consider the likelihood that x_i is sampled from component k as:

$$p(x_i; C_k) = \mathcal{N}(x_i | \Theta_k) = \Big(\frac{1}{(2\pi)^p |\Sigma_k|} \Big)^{\frac{1}{2}} exp\Big(-\frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \Big) \tag{3.5}$$

and the probability of observing any sample x_i, given the mixture model, is characterized as

$$p(x_i;\Theta) = \sum_{k=1}^{K} \alpha_k p(x;C_k) \tag{3.6}$$

the weighted sum of subpopulation likelihoods. The goal of mixture modeling is to estimate the set of model parameters, Θ, that maximize the likelihood, $p(X)$ of a given dataset $X = [x_1, x_2 ... x_N]$

$$p(X;\Theta) = \sum_{i=1}^{N} p(x_i;\Theta) \tag{3.7}$$

which can be solved using the expectation maximization algorithm.

We were interested in identifying the class label, z_i, of data point x_i, which amounts to identifying the subpopulation component k from which x_i is sampled. This can be solved using a maximum likelihood approach, e.g. identify the mixture component that maximizes the likelihood of x_i as

$$z_i = \operatorname*{argmin}_k p(x_i;\Theta_k) \tag{3.8}$$

It is immediately evident that GMMs make an important assumption about the structure of the data, namely, that individual sample likelihoods $p(x_i;\Theta)$ are independent of one another i.e. that

$$p(X_1 = x_1 | X_2 = x_2 \ldots X_N = x_n;\Theta) = p(X_1 = x_1;\Theta) \tag{3.9}$$

In the current context of cortical segmentation, this indicates label assignments of individual data points are also independent – that is, knowing the label assignment of one data point does not influence that label assignment of another data point.

Random Forests

In contrast to Gaussian Mixture Models, decision tree classifiers make no assumptions about the underlying distributional properties of individual samples and their corresponding classes.

A decision tree classifier is structured as a binary search tree. Given a data vector $x_i = [x_{i,1}, x_{i,2} \dots x_{i,p}]$, the vector is pushed through the tree, until a leaf node is reached At each node, beginning at the root node, a binary decision (generally, comparing the value of a vector feature against a threshold value, $x_{i,j} > t$) is made regarding a single feature in the vector. Training a decision tree classifier consists of learning the feature threshold and value (j, t) at each node to best identify individual classes in the dataset. This is achieved by minimizing the current node class impurity (e.g. the heterogeneity of class labels present at a given branch node or leaf node in the tree) using a variety of optimization functions including Gini impurity and information gain. When a data vector reaches a leaf node, that data sample is assigned the maximum probability class label of the leaf node.

Random forests are a form of ensemble learning that consists of training multiple decision trees. Additionally, while a single decision tree has access to all available features at every node, in a random forest, each composing tree randomly samples a subset of features at each node ("feature bagging"), and thus learns a unique decision function at each branching point, and consequently, a unique prediction function. Probabilities for each class are summed across individual trees in the forest and then normalized – the final predicted class label of a test data point is the maximum probability class across all trees in the random forest.

Multi-layer perceptron

Neural networks are a class of algorithms that iteratively perform a series of linear transformations, followed by non-linear transformations of the input samples that are used in both prediction and classification settings. These algorithms have been shown to be useful in a variety of image segmentation and classification settings. Briefly, given an input feature matrix $X \in \mathbb{R}^{N \times F}$ of independent variables and output matrix Y representing the dependent variable(s), a single layer neural network is represented as:

$$\hat{Y} = \sigma(XW) \tag{3.10}$$

where $W \in \mathbb{R}^{F \times K}$ is a linear transformation applied to the feature matrix, σ is a non-linear function (typically a rectified linear unit (ReLU), tanh, or logit function), and $\hat{Y}$ is an estimate of Y. Given a loss function $L(\hat{Y}, Y) \mapsto \mathbb{R}$, the parameters of W are estimated via backpropagation, using the Chain Rule. The matrix W represents the fitted parameters of the network that are optimized over the course of learning. Let us assume that z_l is the weighted input of a layer (after applying W_l to X), and a_l is the "activation" of a layer (after applying the non-linearity, σ). For a single layer network (l=1), backpropagation is applied as

$$\frac{\delta L}{\delta a_1} \cdot \frac{\delta a_1}{\delta z_1} \cdot \frac{\delta z_1}{\delta x} \tag{3.11}$$

where L is the loss function. A multi-layer network stacks these single-layer components together. For example, 2-layer network is defined as follows:

$$\hat{Y} = \sigma_2(\sigma_1(XW_1)W_2) \tag{3.12}$$

where the parameters of each W_j are learned via backpropagation. The non-linear transformations are critical for increasing the flexibility of neural network models. If the non-linearity functions σ_i as $Y = (XW_1)W_2 = XZ$ are removed, the individual linear transformations collapse into a single matrix, Z leaving a simple one-layer network.

3.3.5 Model predictions

We performed 3 types of map predictions on the test data:

1. **Control**: label assignments are based on surface registration alone
2. **Unweighted**: label assignments are based on the raw classifier probabilities
3. **Weighted**: label assignments are based on raw classifier probabilities, scaled by the frequency with which a test vertex maps to a cortical region

Control predictions

As a control, we assigned labels based only on the surface registration, excluding all connectivity information. For a test vertex $v \in V(t)$, we computed $f_v = st(v, t, F_t)$, the prior probability mass function on label assignments for vertex v, and represent these probabilities as a vector f_v of length L. For each $l \in L$, $f_v(l)$ is the probability of vertex v mapping to label l. We defined the predicted label of vertex v as

$$Q_{reg}(v) = l_{max} = \arg\max_{l} f_v(l) \tag{3.13}$$

Unweighted and weighted predictions

For the weighted and unweighted approaches, for test vertex v, we computed prediction probabilities only for labels in $st(v, t, F_t) \ \cap \ lst(l_{max})$ – we considered those labels both with non-zero probabilities in the vertex-level prior distribution and the confusion set of l_{max}. Model probabilities are represented as p_v, a vector of length L. The value $p_v(l)$ is the probability assigned to label $l \in L$ for vertex v if $l \in \{st(v, t, F_t) \ \cap \ lst(l_{max})\}$, and 0 otherwise. In the unweighted approach, we defined the predicted label of vertex v as

$$Q_{unweighted}(v) = \arg\max_{l}(p_v(l)) \tag{3.14}$$

In the weighted approach, probabilities of label m, $p_v(l)$ were weighted by the mapping frequencies $f_v(l)$. A vertex is assigned to the label with the highest weighted probability. We defined the predicted label of v as the label that maximizes the following expression:

$$Q_{weighted} = \arg\max_{l}(f_v \times p_v) \tag{3.15}$$

3.3.6 Model performance metrics

Dice coefficient

The Dice coefficient is often used to assess the degree of overlap between two different classification outputs, and is defined as

$$Dice(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{3.16}$$

where X in our case is the true cortical map (given by the HCP labels), and Y is the predicted cortical map. We compared the spatial overlap of predicted cortical maps with the ground truth maps.

Regional homogeneity

Based on the cortical field hypothesis [29,59,60,67], we expected the connectivity profiles of vertices within a region to be more similar to each other than to vertices in other regions. We used a measure of regional homogeneity (ReHo) metric to characterize this intra-regional similarity, and compare the homogeneity estimates of predicted maps across classifier models. We also compared estimated homogeneity to estimates of ground truth maps.

For a given cortical map L, and parcel l, each vertex v assigned to parcel l is characterized by a feature vector $f_v \in \mathbb{R}^D$ characterizing the unique connectivity profile(s) of that vertex. We computed the mean pairwise correlation matrix P_l of the features vectors $f_v \; \forall \; v \in l$, and define regional homogeneity of parcel l as the mean of P_l. We estimate ReHo for all cortical regions in a cortical map, and take the average of all regional estimates.

3.4 Results

In this analysis, we treat the subject-level maps computed by [9] as "ground truth" maps used to train and test our classification models. We computed classification accuracy of predicted parcellations relative to these ground truth maps as a function of classifier algorithm, spatial prior weighting scheme, and feature data type (resting-state correlations, structural connectivity, and both combined). As a control, we also report prediction accuracy based only on the surface registration. We average the classification accuracy across 5 rounds of cross-validation. We report regional homogeneity of the features of each cortical map, and attempt to characterize the misclassified vertices using a shortest-path approach.

3.4.1 Classifier accuracy

Figure 3.5 shows accuracy of each classifier, stratified by training data type and weighting scheme. Accuracy is measured as the proportion of correctly-labeled vertices. In all models, weighting the classifier probabilities by the mapping frequencies improves results. However, this improvement is smaller in magnitude for the neural network model. However, in the unweighted case, the neural network outperforms both the random forest and Gaussian mixture model. Model performance is comparable across datatype – we hypothesized that model performance would be improved by including both the resting-state and tractography together in the training model, but we found this not to be the case. When classifying test vertices using only surface registration, prediction accuracy was 66.4%, indicating that the registration-only predictions outperform all classification models considered in this analysis.

Classification Accuracy

	Unweighted			Weighted		
	Struct.	Rest.	Combined	Struct.	Rest.	Combined
RF	48.2%	42.0%	47.4%	64.7%	62.7%	64.8%
GMM	32.1%	37.5%	35.2%	65.6%	66.0%	64.2%
NN	48.1%	40.5%	47.2%	54.9%	52.0%	54.9%

Table 3.1: Model accuracy as a function of training data type and weighting scheme. struct = structural tractography, rest = resting-state correlations, comb = combined. Model accuracy is defined as the percentage of correctly labeled vertices.

Figure 3.6 shows an example of an HCP map, and a predicted cortical map produced by the Random Forest model trained on the combined resting-state and tractography connectivity features. The predicted map strongly resembles the true map in terms of label topology. However, the predicted map is quite noisy, especially at areal boundaries.

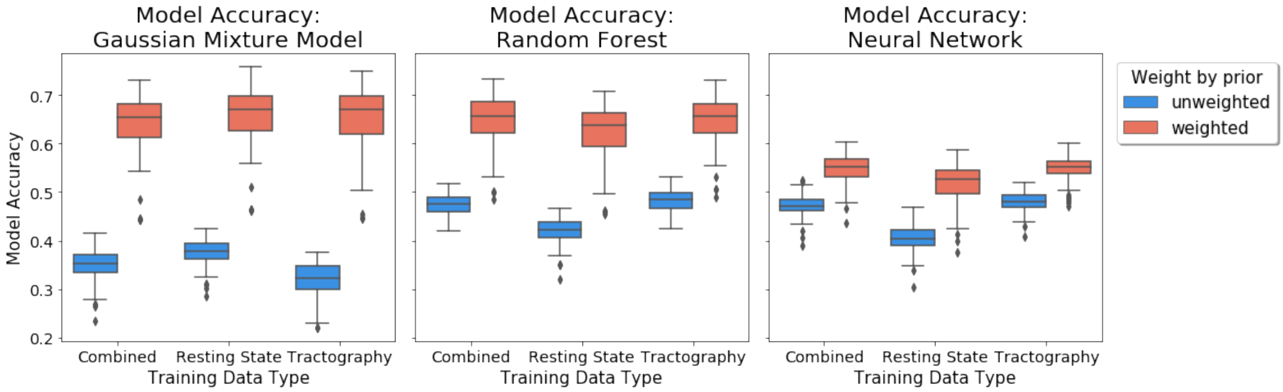


Figure 3.5: Model performance, measured via proportion of correctly-assigned labels. In order of columns, results show Gaussian mixture model, random forest, and neural network performance. Colors show classification scheme: **blue** refers to classification scheme **2**, where we use the raw classifier probabilities, **red** refers to classification scheme **3**, where we weight the probabilities output by the classifier, by the empirical vertex mapping frequencies.

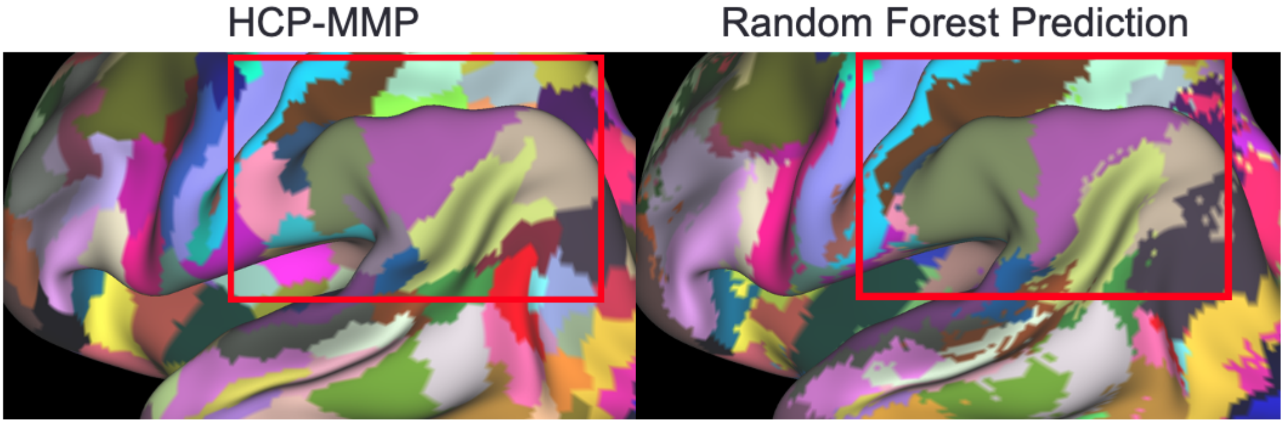


Figure 3.6: A visual comparison of a known HCP-MMP map generated by the HCP consortium versus a map predicted using the random forest classifier with weighting scheme **3**. While the topology of the labels is generally correct, the boundaries of predicted regions are quite noisy.

Figure 3.7 examines mean classification accuracy across all subjects. Models tend to correctly classify vertices near the center of cortical areas, and misclassify vertices near areal boundaries. Additionally, each model tends to perform well across different areas of the cortex. For example, GMMs tend to perform better in the parietal area, relative to the Random Forest and Neural Network models, but performs poorly in the temporal occipital junction and in the primary visual area. Random forests perform well in the primary

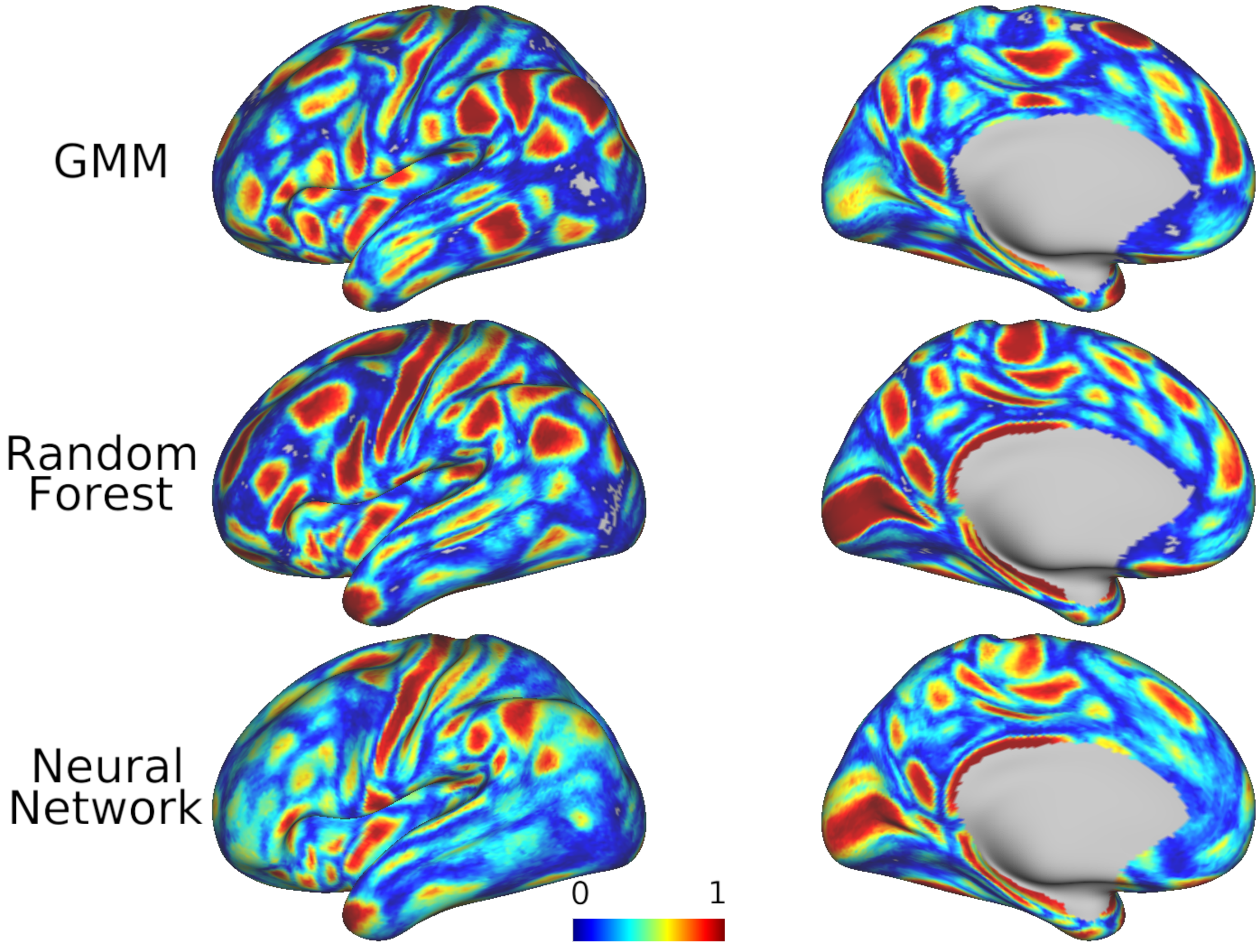


Figure 3.7: Mean classification accuracy maps for GMM (top row), Random Forest (middle row), and Neural Network (bottom row) models for models trained on regionalized resting-state connectivity features. For each model, for each subject's predicted parcellation, we computed the vertex-level accuracy. Maps are averaged across all subjects, for each model.

visual area, and poorly in the lateral frontal area. However, as we show in Chapter 4 in Figure 4.5, these accuracy rates are considerably lower for the spatially unaware algorithms, relative to methods that leverage spatial information.

3.4.2 Characterization of misclassified vertices

We were interested in determining how close the label of a misclassified vertex was to the true label of that vertex as a measure of how accurate a given parcellation was. We constructed an adjacency graph $N = (\rho, E)$ of the HCP-MMP map (see Figure 3.8), where ρ, again, is the set of possible labels, and E are edges connecting those labels. If there exist two adjacent vertices (v, u), one with label m_i and the other with label m_j , we considered labels m_i and m_j to be adjacent in N. Using Dijkstra's algorithm, we computed the shortest path length between all pairs of nodes in graph N. For two nodes (i, j) in a graph, we defined a function $sp : (i, j) \mapsto \mathbb{R}^+$ that computes the shortest path between the two nodes. For a predicted map $Q(v)$ and known map p, we computed the following:

$$\frac{1}{N}\sum_{i=1}^{N} sp(Q(v), L(v,p)) \ \ \forall \ \ v \in V \ \text{ if } \ Q(v)! = L(v,p) \tag{3.17}$$

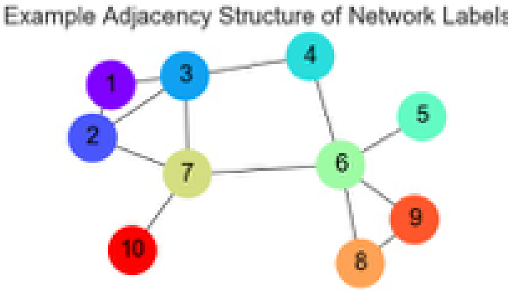


Figure 3.8: An example of label adjacency graph, for labels 1...10. The label adjacency graph capture which labels in the HCP-MMP parcellation are adjacent to one another, and describes this structure as a graph. We used this structure to calculate distances between labels, using geodesic paths computed from this network. Directly adjacent labels would have a distance of 1.

where $L(v, p)$ is the known label of vertex v. This metric represents the mean shortest path length between the predicted and known label, for all misclassified vertices. On average, the mean path length is close to 1, indicating that if a vertex is misclassified, the predicted label is typically directly adjacent to the known label (see Table 3.2).

Shortest Path Lengths Between Misclassified Labels

	Unweighted			Weighted		
	Struct.	Rest.	Combined	Struct.	Rest.	Combined
RF	1.076	1.100	1.081	1.044	1.058	1.044
GMM	1.104	1.090	1.097	1.044	1.038	1.051
NN	1.067	1.092	1.068	1.059	1.078	1.059

Table 3.2: Shortest path distances between the incorrectly predicted labels and known labels. The predicted labels are, at a minimum, a distance of 1 away from the true label, in the case where the true label and the predicted label are directly adjacent.

3.4.3 Regional homogeneity of parcels

We computed the regional homogeneity estimates for every predicted map, based on the feature type used to train each model. Estimates are averaged over the left and right hemisphere.

Regional Homogeneity

	Struct.	Rest.
HCP	0.763	0.623
Destrieux	0.719	0.522
Desikan	0.645	0.431
SurfReg.	0.769	0.622

Table 3.3: Mean regional homogeneity estimates for ground truth and predicted cortical maps. Homogeneity estimates are computed from the known HCP-MMP, Destrieux, Desikan-Killiany maps, the predicted maps from each classifier, and the map predicted by surface-registration alone. Struct. = structural tractography, Rest. = resting-state.

Table 3.3 shows that homogeneity estimates of the surface registration resemble those for the HCP-MMP maps. For comparison purposes, we also show the homogeneity estimates for the known Desikan-Killiany (35 regions) and Destrieux (75 regions) atlas maps. Homogeneity estimates for both the Destrieux atlas and the Desikan-Killiany atlas were under-estimated, relative to the HCP-MMP and surface registration predictions. Because there are fewer regions in the Desikan (35) and Destrieux (75) atlases, each cortical region spans a larger surface area – larger areas will tend to be more spatially variable than smaller areas, resulting in reduced intra-regional similarity of connectivity features. Figure 3.9 shows that parcellation predictions computed using the unweighted approach tended to show decreased functional homogeneity, relative to the weighted model approaches. Likewise, tractography signals tended to be more spatially variable than corresponding resting-state features.

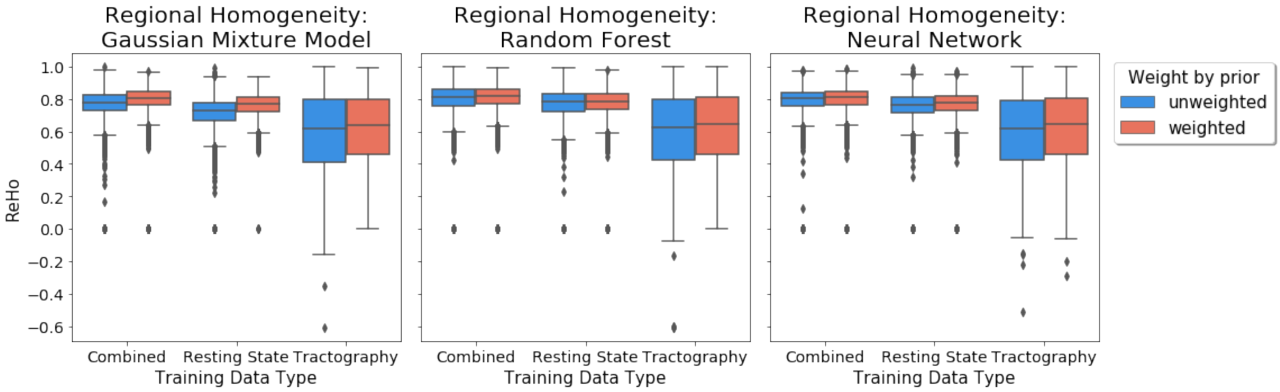


Figure 3.9: Regional homogeneity estimates for each classification model. Tractography data was more variable than resting-state or combined data, as indicated by the larger boxplot spread, and the lower mean estimates.

We computed the Dice coefficient to examine the relative degree of overlap between the predicted and known maps (Figure 3.10). In correspondence with our previous results, unweighted models resulted in less spatial overlap of predicted labels with the HCP-MMP parcels, relative to the weighted schemes. The Random Forest models tended to produce maps with higher Dice coefficients in both the unweighted and weighted approaches. However, cortical maps generated using only the surface registration results produced larger Dice coefficient estimates than any of the model-based predictions (mean Dice: 0.796)

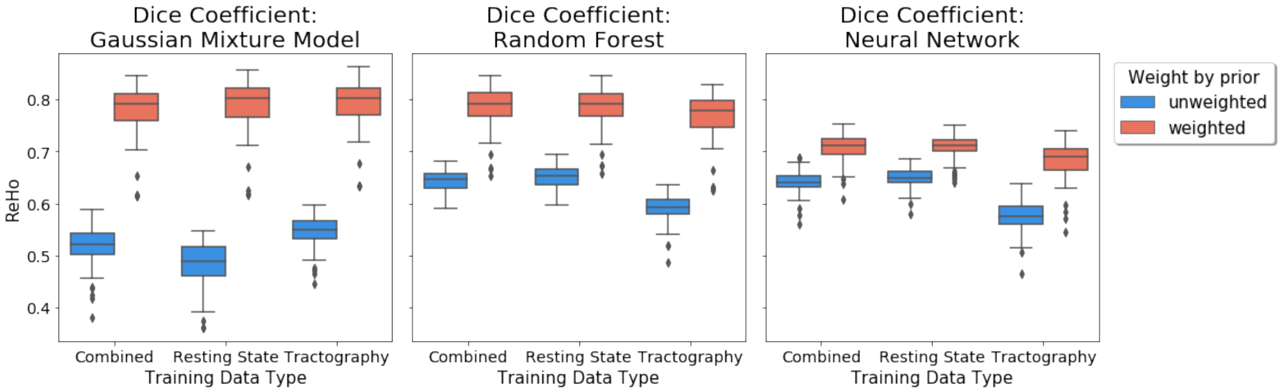


Figure 3.10: Dice coefficient estimates for each classification model. These results largely mirror the classification accuracy results, as in Figure 3.5. Gaussian mixture models and random forests outperform neural networks in terms of spatial overlap of predicted parcels.

3.5 Discussion

In this chapter, we developed a framework to compute cortical parcellations for new, unmapped MRI images on the basis of functional and structural connectivity, local surface geometry, and cortical scalar measures. We examined segmentation performance as a function of three unique classifiers, and three different feature sets. We assessed the performance of these models using raw classification accuracy and regional homogeneity of predicted cortical maps.

We compared Gaussian mixture models (GMM), random forests, and neural network classifiers against a control model based on surface registration alone. We found that the surface registration model outperformed all three classifiers in terms of classification accuracy and regional homogeneity estimates. Random forests and GMM models performed comparably to one another, and outperformed the neural network classifiers. We hypothesize that the weaker performance of the neural network in terms of raw prediction accuracy could be due to over-fitting of noise in the connectivity profiles. Prior research has shown that, for image classification and segmentation tasks, deeper networks often outperform shallower networks. We did not perform extensive parameter optimization of our models – as such, future work might consider examining neural network performance as a function of network depth. Similarly, it has been shown that certain network architectures, such as encoder-decoder networks, often perform better on biological data [68].

One limitation of this work concerns the fact that none of the models studied in this analysis are spatially aware. That is, at training time, mixture models, random forests, and simply feed-forward neural networks do not leverage that fact that cortical voxels that are spatially close to one another on the cortical manifold are correlated with one another. Given two cortical voxels adjacent to one another, if we know the feature vector or connectivity profile of one of those voxels, we have gained a significant amount of information about the connectivity profile of the other voxel. Similarly, if we know the label or areal assignment of one voxel, we have also gained information about the label assignment of the other voxel. In this analysis, we tried to impose a spatial prior on label assignments at test time by weighting raw model probabilities. These weights were computed on the basis of surface registration of a test image to a library of target images. However, we found that this prior is quite rigid, and tends to yield predicted cortical maps that look more like the group-average HCP-MMP parcellation than the specific subject-levels maps. One possible approach to incorporate spatial information into the classification procedure would be to utilize convolution neural networks. These models learn translationally invariant kernel filters that convolve image signals across patches of the data domain, thereby incorporating local information into the training and testing procedure. However, conventional convolutional networks are useful only for data sampled from a regular domain, such as 2D images, 3D volumes, and 4D video streams. While these approaches have been used for 3D MRI volume segmentation and disease classification, they are not readily applicable to data distributed over networks, such as data considered in this analysis where data are interpolated onto the surface mesh. Recent work by [69, 70] extended the idea of convolution kernels to the graph domain, via graph neural networks (GNN). In Chapter 4, we examine cortical segmentation using three variants of graph neural networks, and compare their performance to a simple feed-forward network architecture.

One additional limitation of this current analysis (as well as the analysis considered in Chapter 4) concerns our approach to incorporating multi-modal imaging features into the classification procedure. Specifically, we utilized resting-state correlations, structural connectivity streamline counts, and cortical scalar metrics as features for our models, and aggregated these features using simple column-wise concatenation of the respective feature matrices. As discussed in more depth in Chapter 8, many neurobiological features, like cytoarchitecture, cortical activation patterns, and connectivity features, are significantly correlated with one another, while also displaying their own unique structure. While column-wise concatenation is the most straightforward approach to incorporating multiple feature types into a statistical model, recent advances in deep learning have enabled more intelligent "fusion" of feature types to leverage the inherent structure of each data modality. We discuss some of these approaches in Chapter 8.

Chapter 4

Cortical mapping using spatially-aware classifiers

Eschenburg, K, Grabowski, T, Haynor, D, Learning cortical parcellations using graph neural networks. Frontiers of Neuroscience, 2021.

Initiated by Kristian M. Eschenburg (KME) as a part of, and funded by, NSF grant BCS-1734430 for Principal Investigator Dr. Thomas Grabowski (TG): "Relationship of Cortical Field Anatomy to Network Vulnerability and Behavior". KME conceptualized this study, developed the code, performed the analyses, and wrote the bulk of the document. TJG provided comments and neuroscientific insight into the analysis, and contributed to the editing and organization structure of the manuscript. DRH provided extensive neuroscientific and technical guidance for this work and contributed to the editing and organizational structure of the manuscript.

4.1 Introduction

In the previous chapter, we proposed three different classification schemes for learning maps of the cortical surface using spatially unaware algorithms: Gaussian mixture models, random forests, and simple feed-forward neural networks. Neural network approaches, such as the multi-layer feed-forward network in Chapter 3, have been applied to a wide variety of tasks in medical imaging, ranging from disease classification to tissue segmentation. However, these networks do not always take into account the true spatial relationships between data points. Convolutional neural network approaches, such as those applied to static images or dynamic video streams, learn translationally-invariant, multidimensional kernel filters over the data domain. Both these methods assume that the data is sampled regularly in space, allowing convolution and pooling of information from fixed neighborhood topologies. However, real-world data, such as graph-structured data, is often sampled on irregular domains. Data sampled from graph domains often contains non-uniform topology – individual data points can vary in their neighborhood structure, and notions of direction (e.g. up, down, left, right) do not generalize well to graphs. This makes learning filters to process graph-structured data very difficult with conventional neural network approaches.

Graph neural networks are a class of neural network models that operate on data distributed over a graph domain. Data are sampled from a graph with an explicit structure defined by a set of nodes and edges. These models have been shown to be useful for graph and node classification tasks, along with learning generative models of data distributed over graphs [71, 72, 73, 74]. Graph convolution networks (GCN), proposed in [69], generalized the idea of convolutional networks on grid-like data to data distributed over irregular domains by applying Chebyshev polynomial approximations of spectral filters to graph data. Graph attention networks (GATs) are based on the idea of an attention function, a learned global function that selectively aggregates information across node neighborhoods. The attention function maps a query and set of key-value pairs to an output [75]. The output is defined as a weighted sum of the values, where weights are computed using some similarity or kernel function of the key-value pairs.

It is believed that biological signals distributed over the cortical manifold are locally stationary. Given a small cortical patch, voxels sampled from the patch will display similar functional and structural connectivity patterns, cortical thickness and myelin density measures, and gene expression profiles, among various other signals [8, 52, 76]. Prior studies have attempted to delineate and map the cortex by identifying contiguous cortical subregions that are characterized by relative uniformity of these signals [7, 26, 38, 43]. This work is based on the fundamental idea that contiguous regions of the cortex with similar connectivity and histological properties will tend to function as coherent units. Biological signals distributed over the cortex exhibit local but not global stationarity, so any attempt to parcellate the cortex must take both properties into account.

Here, we developed an approach to perform cortical segmentation – a node classification problem – using graph neural networks, that leverages this biological spatial stationarity. The cerebral cortex is often represented as a folded sheet, and a usable parcellation approach must be applicable to this sort of data. Neural networks can be extended to account for non-stationarity in MRI volumes by incorporating 3D-volumetric convolution kernels. However, these approaches are not easily applied to data distributed over 2-D manifolds like the cortical surface. Additionally, more recent large-scale studies interpolate neurological signals, like cortical activation patterns or various histological scalar measures, onto the cortical manifold to mitigate the potential for mixing signals from anatomically close yet geodesically distant cortical regions, e.g. across sulci [11, 63]. These studies could also benefit from methods that operate directly on graphs.

With the growth of large-scale open-source brain imaging databases (ADNI [77], ABCD [78], HCP [11]), neuroscientists now have access to high-quality data that can be used for training models that can then be applied to new datasets. We leveraged the statistical properties of these high-quality datasets to inform the segmentation of new data using multiple variants of graph neural networks. We considered graph convolution networks (GCN) and two variants of graph attention networks: standard attention networks (GAT) [79], and attention networks with adaptive network depth weighting (a.k.a. jumping-knowledge networks; JKGAT) [80]. We examined how algorithm choice and network parameterization affect cortical segmentation performance. We trained our classification models on high-quality open-source imaging data, and tested them on two datasets with unique spatial and temporal resolutions and different pre-processing pipelines. Other methods have been proposed for delineating the cortex using various registration [15,51], neural network [9,81], label fusion [49,82,83], and even graph neural network approaches [84,85]. To the best of our knowledge, this is the first attempt to examine the performance of common variants of graph neural networks in a whole-brain cortical classification setting and explore their ability to generalize to new datasets using functional magnetic resonance imaging (fMRI). While other studies have proposed the use of graph neural networks to delineate cortical areas, these studies did not perform in-depth analyses on how network architecture, algorithm parameter choices, feature type, and training and testing data parameters impact the predicted cortical maps [84,85]. To this end, we studied how each of these different variables impacts model performance and prediction reliability.

4.2 Background

4.2.1 Graph convolution networks

Convolution filters over graphs using spectral graph theory were introduced by [69]. For a graph $G = (V, E)$ with N nodes and symmetric normalized graph Laplacian, L, define the eigendecomposition of $L = U\Lambda U^T$, where the columns of U are the spectral eigenfunctions of G. Given a graph signal $x \in \mathbb{R}^N$ distributed over G, the graph Fourier transform of x is defined as $\tilde{x} = U^T x$, and its inverse graph Fourier transform as $x = U\tilde{x}$. Graph filtering of x is then defined as $g_\theta(L)x = Ug_\theta(\Lambda)U^T x$, where g_θ is an arbitrary function of the eigenvalues.

Because these filters are not localized in space, [69] proposed to use a Chebyshev polynomial approximation to learn spatially localized filters directly from the Laplacian, reducing the filtering operation of a x to

$$g_\theta(L)x = \sum_{k=0}^{K-1} \theta_k T_k(L)x \tag{4.1}$$

where $T_k(L)$ and is the k-th polynomial and θ_k the k-th learnable Chebyshev coefficient. The polynomial order, K, determines the local spatial extent of the filter. If two nodes i and j are more than K hops apart, the filter value $g_\theta(L)_{i,j} = 0$.

In [70], the polynomial order is set to $K = 1$ so that the spatial extent of the filter is limited to directly adjacent nodes and only one coefficient weight is learned per feature component in each layer of the network. Given $H^l \in \mathbb{R}^{N \times k_l}$, the input feature matrix for layer l, the model learns k_l Chebyshev coefficients, in addition to any additional mixing weights. The model incorporates signals from the l-ring neighborhood into the update of a node – each layer implicitly aggregates over a larger neighborhood than the previous layer (Figure 4.1).

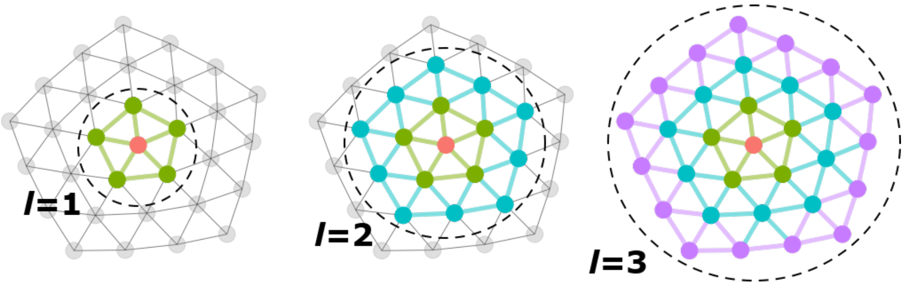


Figure 4.1: Each layer, l, implicitly aggregates more distant neighborhood signals into a node update. The first layer (left) aggregates only over immediately adjacent neighbors, while the second layer, third layer, and so on incorporate signals from larger and larger neighborhoods.

4.2.2 Graph attention networks

Whereas GCNs uniformly aggregate local neighborhood signals, attention networks learn optimized weights for each node neighbor using an attention mechanism. Assume we have data distributed over a graph with N nodes. Inputs to the network are characterized by matrix $X \in \mathbb{R}^{N \times F}$, where F is the number of features. Assume that at any given layer, the inputs to layer l are represented as $H^l \in \mathbb{R}^{N \times k_l}$, where $H^0 = X$. We define the immediate neighborhood of node i as $\mathcal{N}_i$. For two vectors $\vec{n}, \vec{p} \in R^k$, we define their feature-wise concatenation as $n||p \in R^{2k}$. In [79], the attention paid by node i to node $j \in \mathcal{N}_i$ at layer l is computed using a single-layer perceptron as

$$\alpha_{i,j} = \sigma(\vec{a}^T(\mathbf{W}^l \vec{h_i^l} || \mathbf{W}^l \vec{h_j^l})) \tag{4.2}$$

where σ is a fixed non-linearity, $\mathbf{W}^l \in \mathbb{R}^{k_{l+1} \times k_l}$ is a learned layer-specific global linear projection matrix and $\vec{a}$, the attention function, is also learned. The attention weights for $j \in \mathcal{N}_i$ are then normalized by a softmax operation. To update the features of node i at the $(l+1)$-st layer, we compute the weighted sum over the neighborhood $\mathcal{N}_i$ with weights defined by the normalized attentions.

[79] propose an ensemble ("multi-head") attention mechanism, such that, for each layer, M different attention functions are learned, each with their own weight vector $\vec{a}_m^l$. The outputs of each attention head are concatenated feature-wise. In the last layer, the number of hidden channels is the number of output classes, C – rather than concatenating across attention heads, the outputs of all attention heads are averaged to generate the final network output.

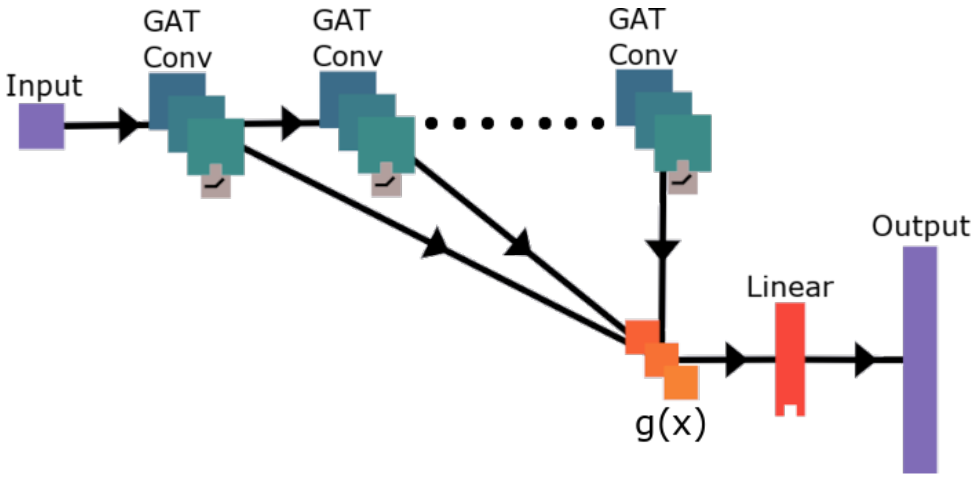


Figure 4.2: Graph attention network employing jumping-knowledge representation learning (JKGAT) with an aggregation function, $g(x)$, that learns a convex combination of the layer-wise embeddings.

4.2.3 Jumping-knowledge networks

While graph neural networks have been instrumental in applying principles of deep learning to graph-structured domains, they are not without pitfalls [70, 79, 80, 86]. GNNs are prone to over-fitting of model parameters and over-smoothing of learned embeddings as network depth increases [86]. One approach to alleviate this over-smoothing is to adaptively learn optimized network depths for each node in the graph, a method [80] describe as "jumping-knowledge networks".

Suppose we have a network with L layers, such that the l-th layer embedding h_i^l for node i is learned by incorporating signals from up to l hops away from node i. The layer aggregation function described by [80] learns a unique output embedding by optimally combining the embeddings of each hidden layer as

$$y_i = \sigma(g(h_i^1, h_i^2, \ldots, h_i^L)) \tag{4.3}$$

[80] propose three permutation-invariant aggregation functions for $g(x)$: concatenation, max-pooling, and long-short term memory (LSTM) [87]. The output, y, is then passed through a linear feed-forward layer to generate the network probabilities. Concatenation is a *global* aggregator (i.e. the same function is applied to all graph nodes) whereas max-pooling and LSTM both learn node-specific aggregations. Further, by utilizing a bi-directional LSTM layer, jumping-knowledge networks learn layer-specific attention weights *for each node* which can then be interrogated *post-hoc* (Figure 4.2). In this analysis, we incorporate the jumping-knowledge mechanism into an attention network framework (JKGAT) and examine cortical segmentation performance using both the LSTM and the concatenation functions.

Given a sequence of samples $x_1, x_2, \ldots x_t$, an LSTM layer maintains a memory of previously observed samples in the sequence in order to learn dependencies between elements. Here, the "sequence" consists of the embeddings learned at each consecutive hidden layer, $h^1, h^2 \ldots h^L$, representing increasingly-abstract representations of functional connectivity. We hypothesized that, because the JKGAT networks learn optimized node-specific network depths, these networks would be able to more-accurately segment the cortex of new data.

4.3 Data

The data used in these analyses come from the Human Connectome Project (HCP) [9, 11] and from the Midnight Scan Club (MSC) [50] (see Sections 2.1.1 and 2.1.2). We were specifically interested in examining how models trained on one dataset would perform on another dataset. Specifically, we trained models on data from the HCP [11], one of the highest quality MRI datasets to date in terms of spatial and temporal sampling of brain signals. We then tested our models on images from both the HCP and MSC datasets.

4.4 Methods

Here, we describe processing steps applied to the HCP and MSC fMRI datasets for this analysis. We begin with the minimally pre-processed BOLD and scalar data interpolated onto the 32k surface mesh.

4.4.1 Markers of functional connectivity

For this work, we considered two variants of functional connectivity as inputs to our segmentation models. These variants each capture unique representations of temporal dependencies between cortical locations.

Regionalized functional connectivity

As mentioned in Sections 2.1.1 and 2.1.2, both the MSC and HCP studies aligned cortical surfaces to the fsaverage_LR surface space. We applied the same methodology described in Chapter 3, Section 3.3.1 to generate regionalized functional connectivity features between each cortical surface vertex and all regions in a cortical atlas.

Briefly, given a BOLD time series matrix $T \in \mathbb{R}^{32k \times t}$ and cortical atlas with k regions, we consider the set of vertices assigned to region k and compute the mean time-series of region k as:

$$\hat{T}_{k,t} = \frac{1}{|k|} \sum_{i \in k} T_{i,t} \tag{4.4}$$

where $\hat{T} \in \mathbb{R}^{K \times t}$ is the matrix of mean regional time-series. We compute $R \in \mathbb{R}^{32k \times K}$, the Pearson cross-correlation between T and $\hat{T}$, where $R_{i,k}$ represents the temporal correlation between a vertex i and cortical region k. These cross-correlation vectors are used as features to train our models.

In this analysis, we generated connectivity features using the Destrieux atlas ([34]) with 75 regions per hemisphere, as it is computed by FreeSurfer and represents a reasonably high-resolution partition of the cortical surface that we hypothesize captures vertex-to-vertex functional variability well. In Section 4.5.5, we show how classification performance depends on which cortical atlas we regionalize over, and on which representation of functional connectivity models are trained on.

Resting-state networks

Prior research has shown that resting-state networks can be used to segment the neocortex on the basis of homogeneous activation profiles ([9, 63]). While regionalized connectivity features described above represent discretized functional connectivity profiles, in that they represent the temporal linear dependence between a vertex and whole discrete cortical regions, matrix decomposition approaches like independent component analysis identify statistically independent whole-cortex activation profiles, such that any given vertex can participate in multiple activation maps [53]. We hypothesized approaches computing what might be called "continuous" connectivity features would yield more informative features for learning cortical parcellations than regionalized connectivity approaches. For this analysis, we examined two methods for generating subject-level resting-state networks: group-ICA coupled with dual regression [53, 88], and template-ICA [89].

Independent component analysis (ICA) is a statistical approach to separate a multivariate signal into constituent components, such that the constituent components are statistically independent of one another. ICA has been used for the analysis of resting-state and task-based fMRI research to decompose a single-subject BOLD time-series into subject-level resting-state networks. Given the low signal-to-noise ratio in individual resting-state images, independent components estimated at the single-subject level (sICA) are noisy and often unreliable. Similarly, when applying sICA to multiple subjects inde-

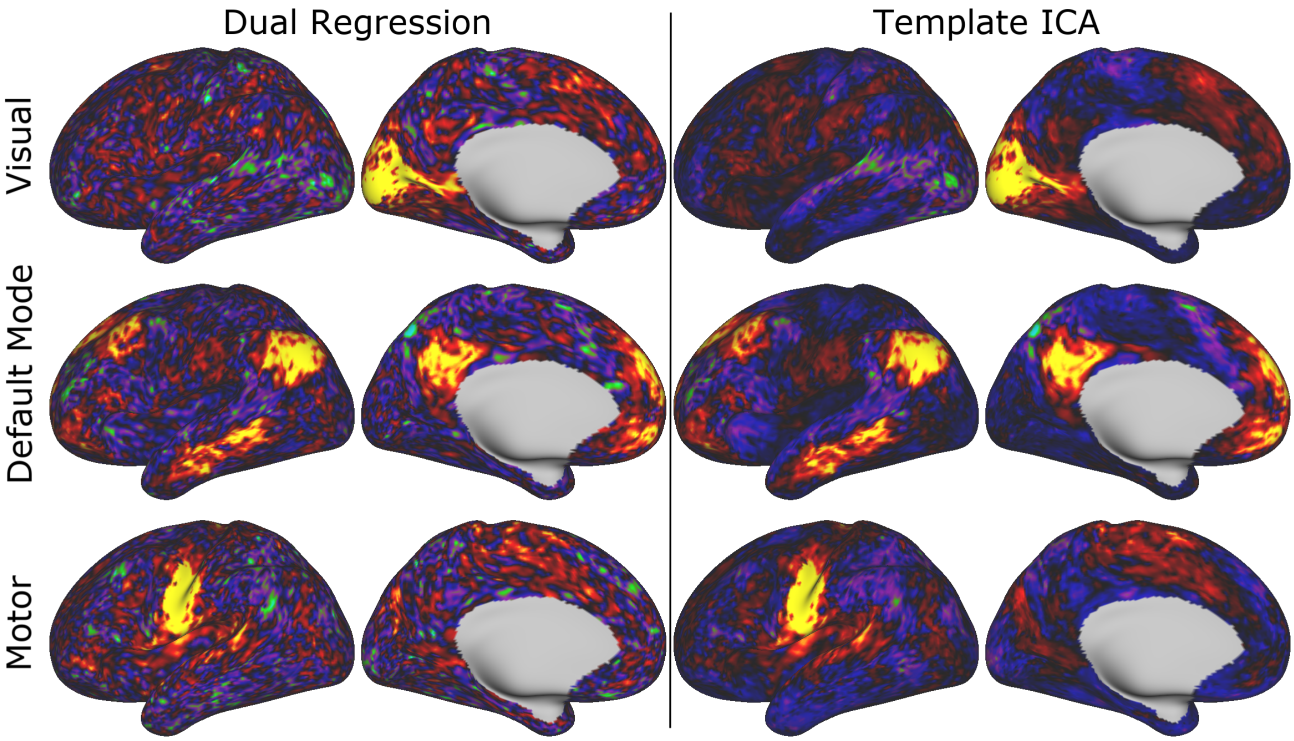


Figure 4.3: Comparison of subject-level resting-state networks estimated through group-ICA and dual regression (left) versus template-ICA for the visual network (top), default mode network (middle) and motor network (bottom). While topologically similar, resting-state networks estimated through group-ICA and dual regression show considerably more spatial noise, relative to those estimate by template-ICA. This is especially clear when examining the lateral premotor cortex in the visual network. We hypothesize that this spatial noise would have detrimental effects on model performance, by corrupting the learned embeddings and thereby reducing model efficacy in mapping the cortex.

pendently, there is no way to enforce the identification of analogous components across subjects, making inference across subject-level networks difficult. For a single subject, ICA proceeds as follows:

$$Y_i = A_i S_i \tag{4.5}$$

where $Y_i \in \mathbb{R}^{T \times V}$ is the subject-level BOLD time series matrix, $A_i \in \mathbb{R}^{T \times L}$ is a mixing matrix, and $S_i \in \mathbb{R}^{L \times V}$ is the matrix of spatial subject-level ICA components.

Group-ICA (gICA) is an approach to generate group-level resting-state networks by aggregating the BOLD signals across multiple subjects [88]. gICA approaches typically employ iterative algorithms that retain a running average of group-level spatial components. These spatial components represent group-level resting-state networks, and are less-noisy than subject-level components because they have incorporated the BOLD signals from many individuals. Components identified via gICA can be analyzed alone, or can be further processed using dual regression to generate subject-level resting-state networks. Given group-level components S_0 estimated using gICA, and a subject level BOLD matrix, Y_i, dual regression proceeds in two steps, via two multivariate linear models, as follows:

$$Y_i = A_{G,i} S_0 \quad \text{(Step 1)} \tag{4.6}$$

$$Y_i = A_{G,i} S_{G,i} \quad \text{(Step 2)} \tag{4.7}$$

where we first estimate the subject-level mixing matrix $A_{G,i}$ using the group-level components (Step 1), and then estimate the subject-level ICA components, $S_{G,i}$ using $A_{G,i}$ (Step 2). Dual regression first estimates the mixing fractions of each group component in the subject-level time-series data. Given these mixing fractions, the algorithm identifies corresponding subject-level networks. In this way, dual regression accounts for the temporal and spatial variation in subject level data *and* maintains correspondence of networks across subjects. However, subject-level components estimated through dual regression and gICA are still influenced by low signal-to-noise ratio in the subject-level BOLD data.

Template-ICA is a statistically-driven modeling approach for estimating subject-level resting-state networks, that is implemented via a hierarchical Bayesian framework. Briefly, the algorithm assumes the following nested structure:

$$Y_i = A_i S_i + A_{nuis} S_{nuis} + \epsilon \tag{4.8}$$

$$S_i = S_0 + \Delta \tag{4.9}$$

where $\Delta_l(v) \sim \mathcal{N}(0, \sigma_l^2(v))$. Again, A_i and S_i are the subject-level mixing and component matrices. However, the model assumes that the subject-level components are sampled from a distribution with mean centered at the group-components S_0 with some between subject-variance Δ. The model treats A_i and S_i as random variables with added noise, ϵ, and estimates nuisance / noise components S_{nuis} that are removed from the estimated components S_i. In this way, template-ICA enforces correspondence of networks across subjects, and applies denoising to the estimated subject-level components.

Given that graph networks perform a series of spatial convolution steps defined by spectral filters (as many convolution steps as there are layers in the network), we hypothesized that reducing spatial noise in the training data would improve model performance considerably by minimizing propagation of noise through the learned embeddings. We sought to examine whether models trained on features computed using template-ICA could outperform models trained on features computed using dual regression.

4.4.2 Markers of global spatial position

We also included measures of position in grayordinate space (global spatial position) as model features [84,85]. Surface mesh Laplacian eigenvectors represent a spatial variance decomposition of the cortical mesh into orthogonal bases along the cortical manifold. We retained the first three eigenvectors corresponding to eigenvalues $\lambda_1, \lambda_2, \lambda_3$. The eigenfunctions represent an intrinsic coordinate system of the surface that is invariant to rotations and translations of the surface mesh.

The eigendecomposition computes eigenvectors up to a sign flip (that is, the positive/negative direction of an eigenvector is arbitrary), and eigenvector ordering is not guaranteed to be equivalent across individuals. We chose a template subject and flipped (multiplied by -1) and reordered the eigenvectors of all remaining subjects with respect to this template subject via the Hungarian algorithm, to identify the lowest cost vector matching for every template-test pair (here, we minimized the Pearson correlation distance).

4.4.3 Incorporating a spatial prior

The models trained in this analysis represent multi-class classifiers. By default, each vertex considers every label (out of a total of 180 possible labels) as a viable candidate. This approach, however, does not take advantage of the fact that training and testing data are in spatial correspondence with one another. For example, if we know a vertex is likely to be assigned a label in the occipital lobe, we can restrict the set of candidate labels for this vertex to a subset of the possible 180 areas e.g. only those areas in the primary and higher-order visual areas. We restricted the label search space of a test vertex to only those labels with non-zero probabilities in the training set. If a given vertex i is never assigned label k in the training data, we set the estimated network probability of label k for vertex i to 0, such that it is never assigned label k in the test set. We implemented the application of the spatial prior by multiplying the network logits with a binary masking matrix *at test time* (e.g. the prior is not included in the model training phase).

Applying the spatial prior is only feasible if the test image surface mesh has been spatially normalized to the fsaverage_LR32 space. Given that many studies will be interested in performing multi-subject inference over surface-based maps, we believe this is a reasonable assumption to make. We examine classification performance when excluding and including the spatial prior.

4.4.4 Regional homogeneity

We examined whether our models learned parcellations in which the features of each parcel were homogenous. We defined homogeneity for a given parcellation as in [26]. Assume we are given a resting-state fMRI BOLD time series matrix $T \in \mathbb{R}^{32k \times t}$ and precomputed cortical parcellation with L cortical areas. For each parcel $l \in L$ with n_l vertices, we computed the Pearson correlation matrix, $R_l \in \mathbb{R}^{n_l \times 32k}$, between the parcel BOLD signals with the BOLD signals of the entire cortex. We then applied the singular value decomposition as $R = USV^T$, where S is the diagonal matrix of singular values $\sigma_1, \sigma_2 \ldots \sigma_N$. [26] defined homogeneity as $\rho_l = 100 * (\sigma_1^2 \Big/ \sum_{i=1}^{k} \sigma_i^2)$, the percent of variance explained by the first principal component. The variance captured by the first component describes how well a single vector explains the functional connectivity profiles of a given cortical parcel – the larger the variance explained, the more homogeneous the parcel connectivity. We computed an estimate of functional homogeneity for each parcel and averaged the estimates across all parcels.

For scalar features (e.g. myelin density), we estimated homogeneity as the ratio of within-parcel variance to between-parcel variance. For each parcel $l \in L$ and feature $F \in \mathbb{R}^{32k}$, we computed the mean, μ_l, and variance, σ_l^2 of the parcel-wise features. Homogeneity is estimated as $\sum_{i=1}^{L}(\sigma_l^2 - \bar{\sigma^2}) \Big/ \sum_{i=1}^{L}(\mu_l - \bar{\mu})^2$, where $\bar{\sigma^2}$ and $\bar{\mu}$ are the average variance and average mean estimates. This measure of homogeneity is a dimensionless quality that allows us to compare estimates across datasets and features. A smaller value represents more homogeneous parcels.

4.4.5 Model training and parameter selection

We implemented each model using the Python package Deep Graph Library (DGL) and PyTorch [90]. Code developed for this analysis for training these models can be found here: https://github.com/kristianeschenburg/parcellearning/. For a detailed description of the necessary input files and how to use this package, see Appendix 10.

We split the 268 HCP subjects into 100 training samples, 20 validation samples, and 148 test samples. For parameter optimization, we trained models on three types of datasets: 1) 100 15-minute images (`REST1_LR` session for each subject), 2) 100 60-minute images (temporal concatenation of all four rfMRI sessions), and 3), 400 15-minute images (four independent rfMRI sessions per subject). We used a validation dataset of 20 subjects of the same scanning duration as the training data to determine when to stop training. We examined the performance of each model on test hold-out test set of different scanning durations: 15-minute (four independent rfMRI sessions), 30-minute (concatenation of two 15-minute rfMRI sessions acquired on the same day), and 60-minute (temporal concatenation of all four 15-minute rfMRI sessions). The outcome variable to be predicted was the subject-level parcellation provided to us by MG. We performed similar temporal concatenation of the MSC data, concatenating the original ten 30-minute sessions into five 60-minute sessions, two 150-minute sessions, and one 300-minute session.

The features used for parameter optimization were the regionalized functional correlations between each cortical vertex and all regions in the Destrieux atlas, the first three Laplacian eigenvector embeddings capturing global location information, and four scalar maps corresponding to sulcal depth, Gaussian curvature, myelin density, and cortical thickness for a total of 81 features at each vertex. We concatenated these features column-wise into a matrix for each subject.

We optimized model performance over network depth, number of hidden channels per layer, feature dropout rate, number of attention heads (in the case of attention networks), and aggregation function (in the case of jumping-knowledge networks). In all cases, the "default" network is a model with 3 layers, a feature dropout rate of 0.1, 32 hidden channels, 4 attention heads per layer (GAT and JKGAT only), and an LSTM aggregation function (JKGAT only). We varied one parameter type at a time: for example, when comparing networks with 3 layers versus 6 layers, keeping all other parameters equal to the default settings. In figures, we refer to the graph convolution network, graph attention network, and jumping knowledge network as GCN, GAT, and JKGAT, respectively. We compared the performance of these algorithm variants to a simple linear feed-forward baseline neural network where only the features at each vertex were used to classify cortical nodes (no adjacency information is incorporated into the learning process), which we refer to as Indep-NN.

For training, we used a cross entropy loss, a LeakyReLU activation function with a negative slope of 0.2, and Adam optimization with a weight decay rate of 0.0005 and L2 weight regularization of 0.005. We trained in mini-batches of size $s = 10$ graphs and accumulate the gradients for each batch before computing the gradient update. We trained for 1000 epochs using an early stopping criteria evaluated on the validation loss. At each iteration, we retained the model if the current validation loss was lower than the previous validation loss. If validation loss did not decrease for 150 epochs, training was terminated and the best performing model was saved. In practice, we found that few of the models trained for 1000 epochs.

4.5 Results

We first examine the best performing model of those we considered in our analysis, and discuss the classification accuracy and reproducibility of parcellations predicted by this model in relation to parcellations computed by [9]. We define classification accuracy as the percentage of correctly predicted vertex labels relative to the ground truth maps. We then show broadly how algorithm choice, network architecture, and training and testing image scan duration affect overall model performance. Finally, we illustrate how classification performance is related to the features used during model training and testing.

4.5.1 Prediction accuracy in the best performing model

Network optimization was performed using labels provided by MG (see: Acknowledgements) using subject data from the S500 HCP release. As mentioned in Section 2.1.1, the S1200 data release uses a different surface registration algorithm, producing subject-level resting-state data that is better aligned with the labels provided by MG. Final model evaluation was performed using the S1200 data. The best performing model was the 6-layer graph attention network, with 4 attention heads per layer, 32 hidden channels per layer, and a dropout rate of 0.1, and incorporated a spatial prior at test time. When trained on features computed using ICA, this model achieved a mean classification accuracy of 79.91%. We henceforth refer to this model as the "optimal" model, and discuss results associated with this model below.

In Figure 4.4.**A**, we show predicted parcellations computed using this model for exemplar HCP and MSC test subjects. Predicted subject-level parcellations topologically resemble the "ground truth" maps generated by [9] (see Supplementary Results for example predictions generated by each model). By taking advantage of the local network topology of the data domain, graph neural networks account for spatial dependencies in the input features and thereby encourage more contiguous parcels. Subjects from the MSC dataset do not have corresponding ground truth maps against which to compare their predictions. In Figure 4.4.**B**, we show consensus predictions for each dataset, compared against the publicly released HCP-MMP atlas. Consensus predictions were

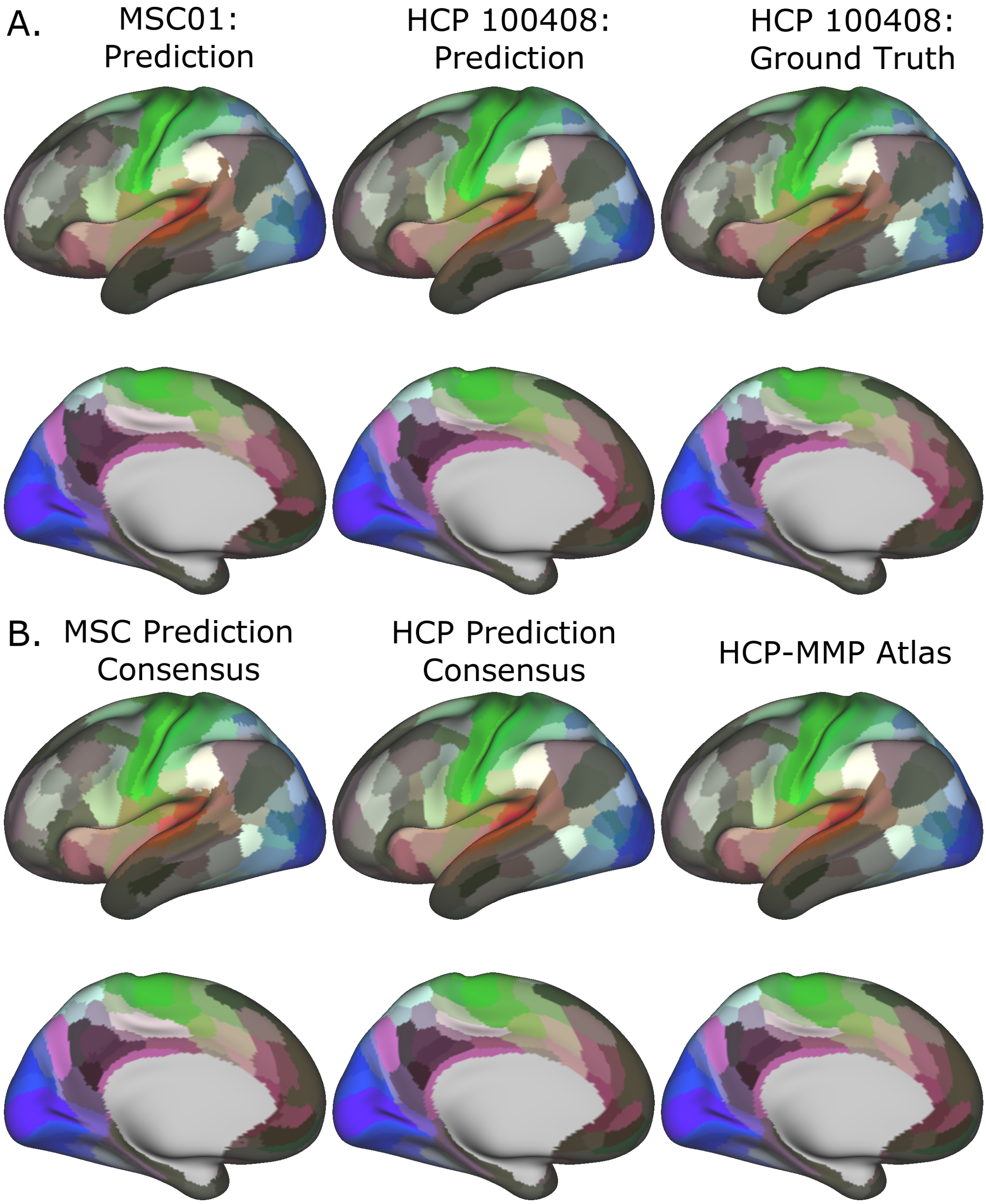


Figure 4.4: Subject-level (**A**) and group-level (**B**) predictions generated by the optimal model in the MSC (left column) and HCP (middle column) datasets.

computed by assigning a vertex to the label most frequently assigned to that vertex across the individual test subject predictions. We see that both consensus predictions closely resemble the HCP-MMP atlas – however, the consensus map derived from the MSC subjects shows noisy parcel boundaries and disconnected areal components (lateral and medial prefrontal areas).

Figure 4.5 shows the spatial distribution of classification accuracy rates averaged across all subjects in the HCP test set. Average accuracy is shown as a map distributed over the cortex, with values ranging between 0 (blue; vertex incorrectly classified in all subjects) and 1 (red; vertex classified correctly in all subjects). Vertices near the centers of cortical regions were classified correctly more frequently, while prediction errors tended to be distributed near the boundaries of cortical regions. To some degree, this effect can be attributed to the idea that boundaries between putative cortical areas represent segments of the cortex with changing biological properties. In developing a statistical model to assign a vertex to one cortical area or another, vertices at region boundaries will have more ambiguous label assignments simply due to the fact that their feature vectors are sampled from a space with greater distributional overlap across various cortical areas. However, another possible explanation is that modern magnetic resonance imaging resolution is low with respect to cortical functional features like cell columns. Consequentially, this means that voxel-wise measurements reflect mixtures of biological effects due to partial voluming, thereby reducing the ability of a statistical model to distinguish between two cortical areas.

While errors globally tended to be concentrated at region boundaries, some cortical areas showed higher error rates than others. In Figure 9.1, we show *regional* classification accuracy rates for the optimal model. The model performs best at correctly classifying vertices in the primary visual cortex, as well near the angular and supramarginal gyri. Interestingly, in the primary sensory cortex, the model performs notably worse at classifying area 3a, relative to immediately adjacent areas 3b and 4. The model performs less

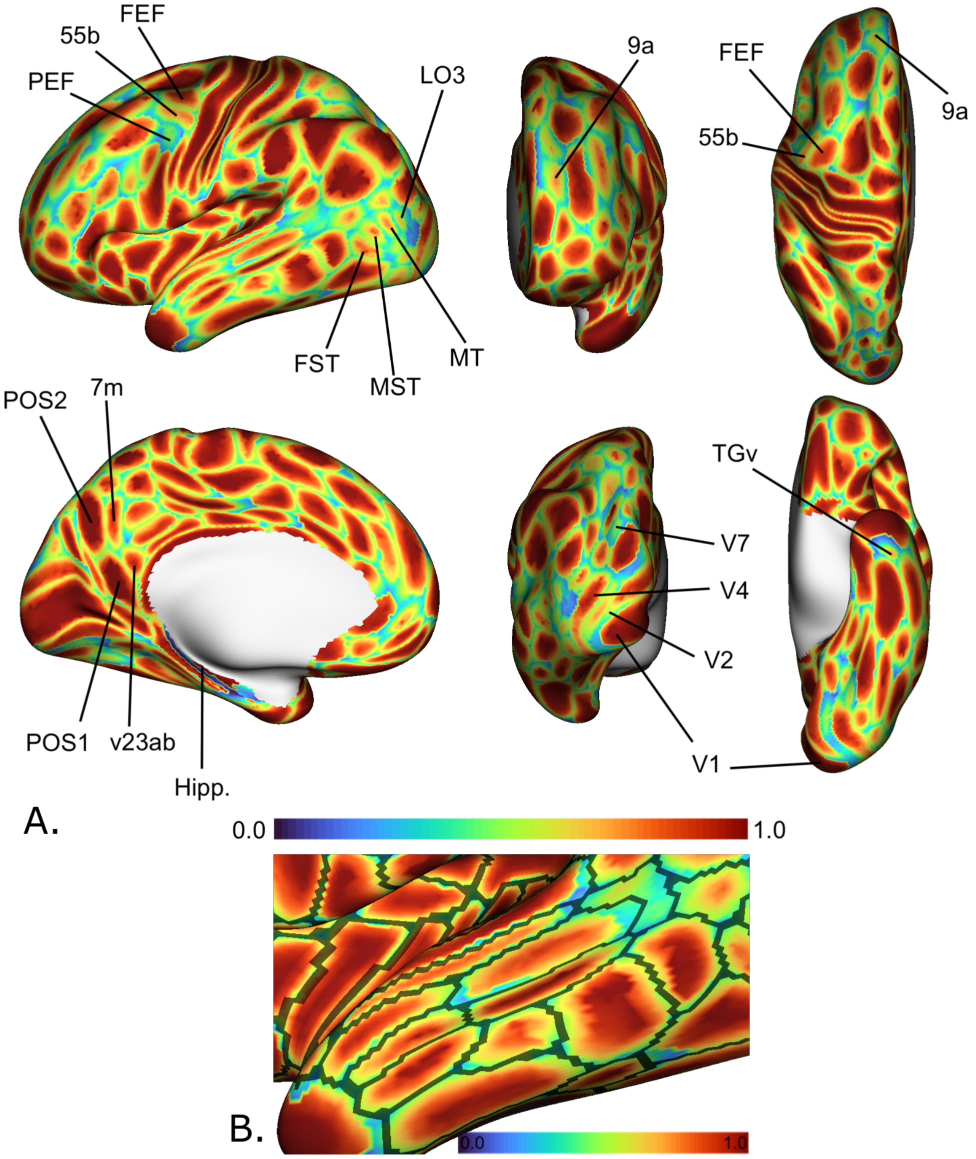


Figure 4.5: Average accuracy maps for the HCP test set using the optimal model for whole hemisphere (**A**), and zoomed in on the temporal lobe (**B**). Blue (0.0) = vertex incorrectly classified in all test subjects; Red (1.0) = vertex correctly classified in all test subjects. Areas in the lateral prefrontal and ventral / dorsal occipital areas showed the highest error rates. Across the cortex, errors mostly occurred at the boundaries of cortical regions. Black lines represent areal boundaries of the consensus prediction parcellation (bottom).

well in the primary auditory cortex and in the lateral higher-order visual areas related to object recognition (LO1, LO2, LO3). In the lateral prefrontal area, we found that the premotor eye field (PEF) shows decreased accuracy estimates relative to adjacent regions (55b and frontal eye field, FEF). [9] identified three unique topologies (typical, shifted, and split) for area 55b that varied across subjects, which might, to some degree, explain the lower accuracy rates in area PEF. Overall classification accuracy rates are still high, with a mean and median regional accuracy rates of 76.71% and 78.03%, and maximum, and minimum regional classification accuracy in areas V1 and VMV2 (ventromedial visual area 2) of 93.5%, 58.1%, respectively.

We quantified the relationship between the spatial distribution of errors and their distance to cortical areal boundaries. We computed the fraction of misclassified vertices that occurred at a geodesic distance of k edges (geodesic hops) from any cortical areal boundary. Using the default model parameters and regionalized features, we examined this distribution of errors as function of distance (Supplementary Material). Over 50% of misclassified vertices occurred at the region boundaries i.e. those vertices in the ground-truth parcellations that are directly adjacent to different regions, and roughly 30% and 12% of misclassified vertices were 1 and 2 edges away from areal boundaries, respectively. The simple feed-forward network misclassified vertices further away from region boundaries, while the three graph neural networks tended to misclassify only vertices close to the boundary.

Although the MSC subjects do not have corresponding ground truth maps, the data is in spatial correspondence with the fsaverage_LR32 map. We computed the correspondence of maps predicted on the MSC subjects with the HCP-MMP atlas in order to gain insight into the accuracy of these predictions. Mean correspondence of predictions computed on the MSC and HCP datasets with the HCP-MMP atlas was 70.04% and 84.35%, respectively (Supplementary Material).

Mean model probabilities computed by the optimal model for a set of cortical areas are

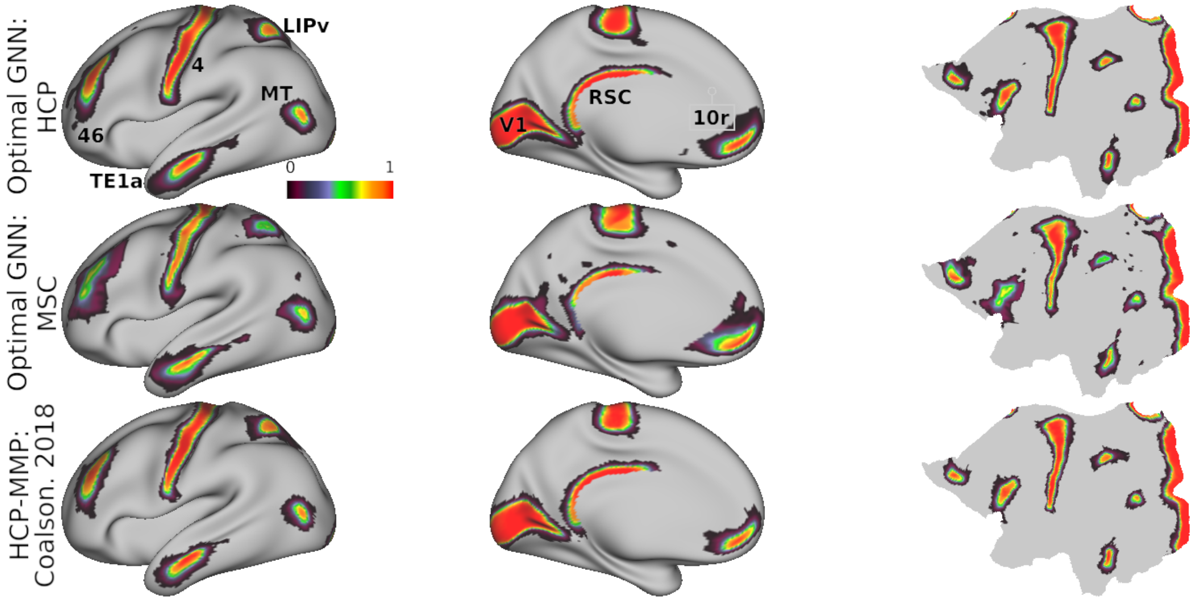


Figure 4.6: Mean model probabilities for a subset of cortical areas for the HCP (top) and MSC (middle) datasets computed using the optimal model, and the MMP binary class probabilities from [9, 13] (bottom). Probabilistic maps are illustrated for areas V1, 46, TE1a, LIPv, MT, RSC, and 10r. These maps are thresholded at a minimum probability value of 0.005, the probability of randomly assigning a vertex to one of the 180 cortical areas. Here, we excluded the spatial prior that masks the probability maps without changing the actual probability values.

illustrated in Figure 4.6, showing that areal probabilities are local in nature and restricted to precise anatomical locations. Individual areal probabilities computed by [9, 13] using their binary classifier are shown in the bottom row. Probability estimates in the HCP dataset mirror those estimated by the original HCP classifier [9], indicating that our model faithfully learns the proper spatial extent of each cortical areal. Estimates in the MSC dataset were slightly more diffuse and less confident (see areas V1 and 46), such that probability mass was assigned to more disparate areas of the cortex, relative to probabilities estimated in the HCP dataset.

4.5.2 Model predictions are reproducible across scanning sessions

The HCP acquired four 15-minute resting-state acquisitions per subject, while the MSC acquired ten 30-minute resting-state acquisitions per subject. We examined how reliable predictions generated from each resting-state session were within subjects, and how this reliability related to the scanning duration. For a given subject, we estimated session-specific reproducibility using datasets of the same scan duration. We defined reproducibility using the Dice coefficient, which measures the similarity of two images. The Dice coefficient between sets J and K is defined as

$$Dice(J, K) = \frac{2 * |J \cap K|}{|J| + |K|} \tag{4.10}$$

Figure 4.7 shows the mean areal Dice coefficients for each dataset from predictions computed using the optimal model. Predictions made on the HCP dataset were more reproducible across the entire cortex than predictions on the MSC dataset. In both datasets, sensory/motor and areas near the angular and supramarginal gyri were most reproducible. The visual cortex showed high reproducibility in area V1, while areas V2-V4 were less reproducible.

Figure 4.8.**A**, shows mean reproducibility estimates computed on the HCP and MSC datasets. Predictions for both datasets were highly reproducible across repeated scanning sessions, and reproducibility increased with increasing scan duration. Mean Dice coefficient estimates in the HCP dataset were 0.81 and 0.86 for the 15- and 30-minute durations. In the MSC dataset, the mean Dice coefficients were 0.69, 0.76, and 0.82 for the 30-, 60-, and 150-minute durations. When fixing scan duration (e.g. 30-minute durations), HCP data were more reproducible than the MSC data. One feature that we could not evaluate ourselves was reproducibility of the ground truth maps. [9] reported maximum and median Dice coefficient estimates of 0.75 and 0.72 for repeated scans

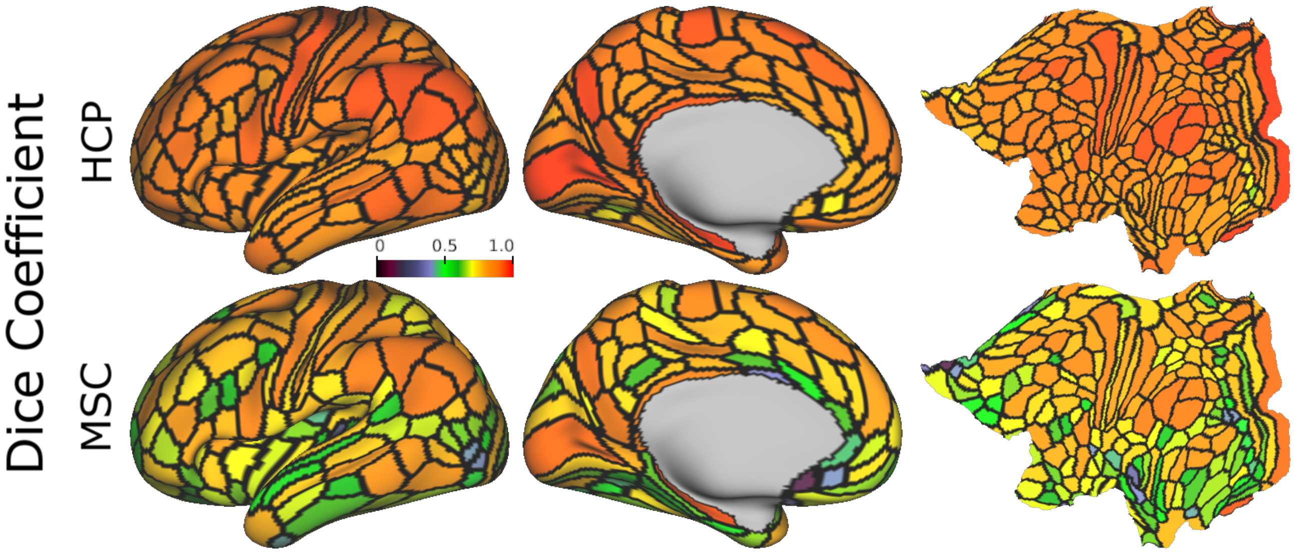


Figure 4.7: Mean areal Dice coefficient estimates, computed using the optimal model on 15-minute HCP data (4 repeated sessions) and 30-minute MSC data (10 repeated sessions) normalized with the same color map.

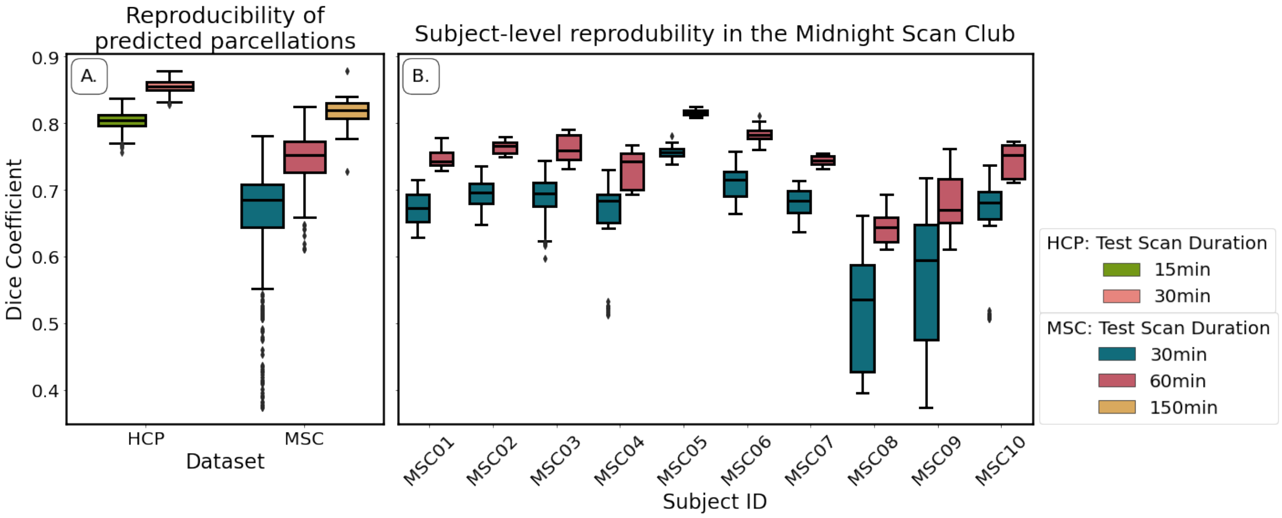


Figure 4.8: Reproducibility of predicted maps generated by the optimal model. We show mean Dice coefficient estimates for each dataset (**A**), and subject-level estimates in the Midnight Scan Club (**B**). We do not show estimates for 60min (HCP) and 300min (MSC) durations in panel (**A**) because there is only one image per subject for these durations. We exclude the 150min estimates in panel (**B**) because there is only a single scalar estimate per subject.

on HCP participants, indicating that our classifier learned parcellations that were more reproducible than those generated by the binary classifier.

Figure 4.8.**B** illustrates subject-level reproducibility estimates in the MSC dataset. Predictions for subject MSC08 were significantly less reliable, relative to the other subjects. [50] also identified MSC08 as having low reproducibility with respect to various graph theoretical metrics computed from the functional connectivity matrices. They noted that subject MSC08 reported restlessness, displayed considerable head motion, and repeatedly fell asleep during the scanning sessions.

Area-level topologies were also reproducible across scanning sessions. [9] identified three unique topologies of area 55b, corresponding to a "typical", "shifted", or "split" organization pattern, relative to the group-average cortical map. We were able to identify these same unique topologies in individual subjects, indicating that graph neural networks are identifying the unique connectivity fingerprints of each cortical area, and not simply learning where the parcel is (Figure 4.9). When we examined the predictions generated by the optimal model on the four independent 15-minute scanning sessions, we found that, within a given subject, the topological organization of area 55b was reproducible. Figure 4.10 illustrates topological reproducibilty for 6 independent subjects for each of the four resting-state sessions for area 55b. Subject 133928 (bottom row) is especially notable in that this subject's 55b topology deviates considerably from the group consensus areal boundaries, but this deviation is highly consistent across each of the four resting-sate sessions. Similarly, subject 114924 is unique because of it's "split" topology. However, we see that, across sessions, this "split" characteristic is remarkably conserved. Allowing for some variability in prediction boundaries and location due to resampling of the connectivity data and partial voluming effects, these results indicate that the graph neural networks are learning subject-specific topological layouts that incorporate their unique connectivity and histology patterns, even when those topological layouts deviate considerably from the group average map.

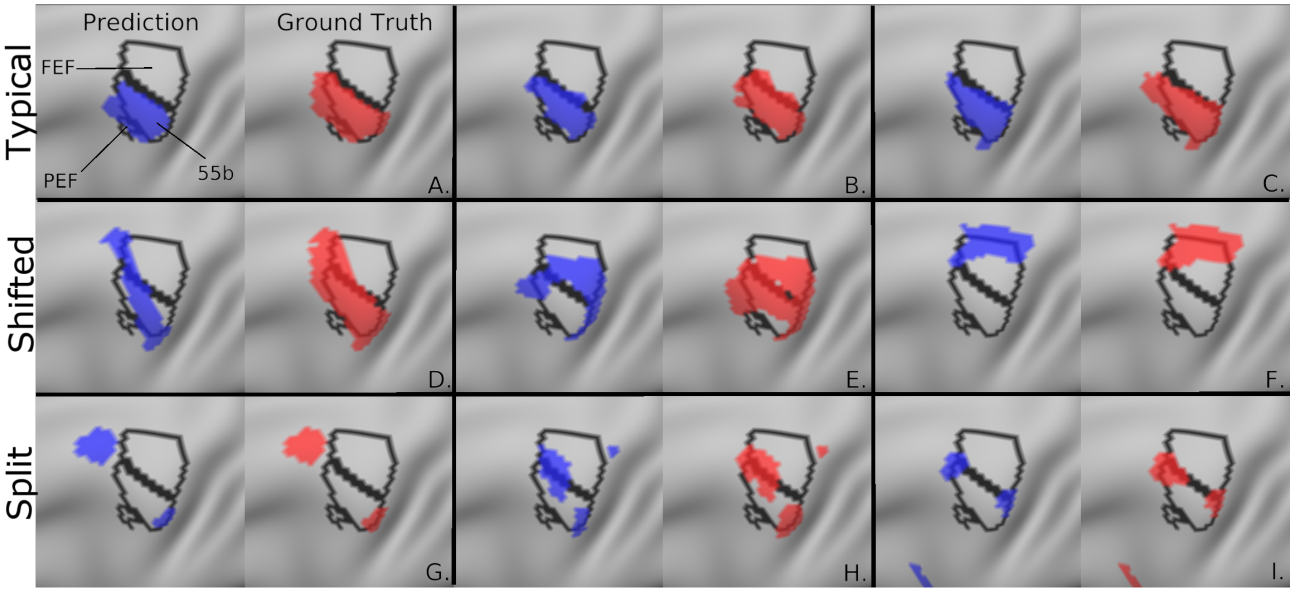


Figure 4.9: Exemplar topologies for area 55b. [9] identified "typical" (top), "shifted" (middle), and "split" (bottom) topologies. The group-level areal boundaries are shown in black for area FEF (superior), 55b (middle), and PEF (inferior).

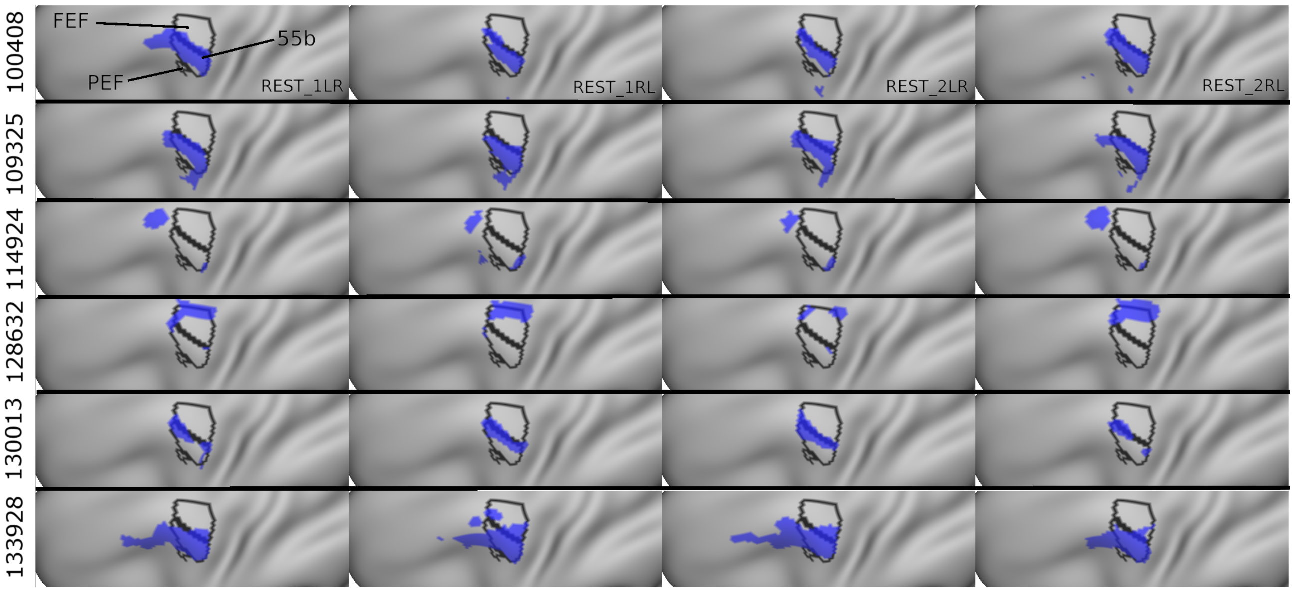


Figure 4.10: Reproducibility of subject-level topologies for area 55b for 6 exemplar subjects across each of the 4 independent resting-state sessions. The group-level areal boundaries are shown in black for area FEF (superior), 55b (middle), and PEF (inferior).

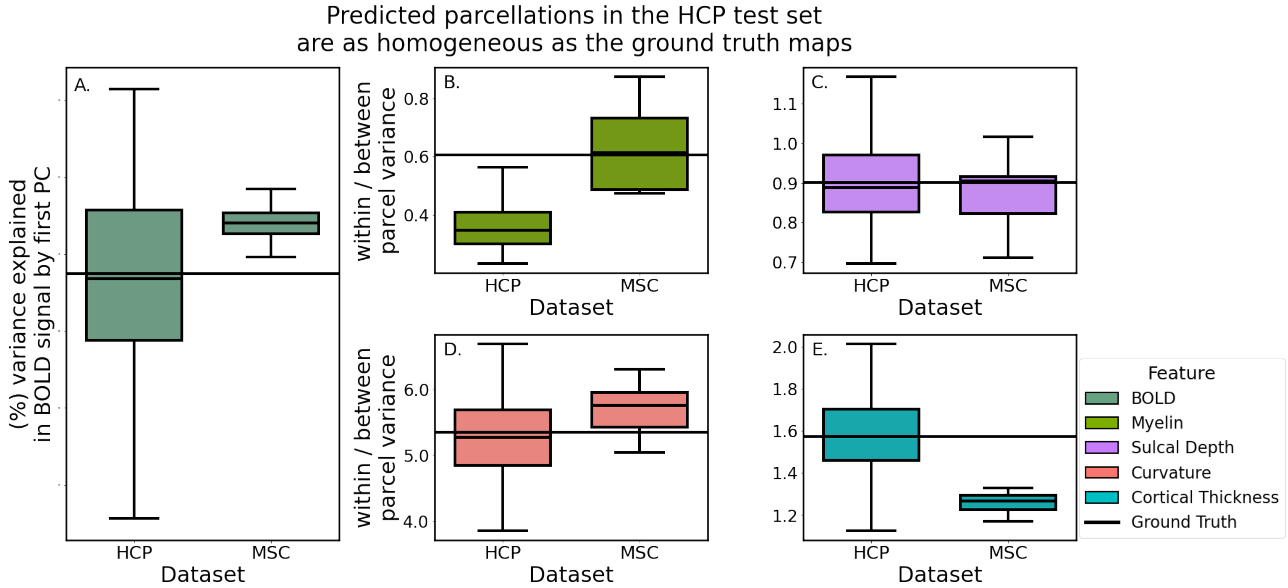


Figure 4.11: Homogeneity of predicted parcellations in the HCP and MSC datasets using the optimal model. (**A**) Predicted parcels in the HCP test set explained as much variability in the functional connectivity as the ground truth parcels. (**B**-**E**) Predictions in the MSC had more variable myelin content and less variable cortical thickness estimates, relative to the HCP predictions.

4.5.3 Parcellations learned by GNNs are homogeneous in their scalar and connectivity measures

If a model is in fact learning unique, discrete areas, the distribution of biological features in these areas should be relatively homogeneous. Unsupervised learning clustering algorithms designed to parcellate the cortex often incorporate objective functions that attempt to maximize within-parcel similarity and minimize between-parcel similarity. On the other hand, gradient-based approaches, like those proposed in [6, 26, 91], do not directly maximize an objective function in this manner, but rather identify putative areal boundaries by identifying where biological properties change dramatically in a small local neighborhood. It is assumed that this biological gradient captures differences in homogeneity between adjacent cortical areas. In order to group cortical voxels together, these voxels must inherently share some physical or biological traits.

We computed homogeneity estimates as described in Section 4.4.4. In order to compare the homogeneity and variance estimates between predicted parcellations, we fixed the features used to compute these estimates. For a given subject, we computed functional homogeneity using that subject's 60-minute BOLD signal (HCP), or the 300-minute BOLD signal (MSC). In this way, the only variable that changed with respect to the homogeneity estimate is the cortical map itself. We could then make meaningful quantitative comparisons between estimates for different maps, with respect to a given dataset.

Cortical maps predicted in the HCP dataset explained on average 67.03% of the variance in functional connectivity while MSC predictions explained 72.90% (T: 3.13, p: 0.007). (Figure 4.11). Scalar metric homogeneity estimates were comparable between the two datasets (myelin (T: -0.909, p: 0.38), sulcal: (T: 1.04, p: 0.32), curvature: (T: -2.42, p: 0.028), thickness: (T: 11.56, p: 3.34e-11)). We hypothesized that parcellations predicted in the MSC dataset would be less homogeneous, relative to those learned in the HCP dataset, due to the fact that the MSC imaging data was acquired with lower spatial resolution than that acquired by the HCP and is therefore subject to greater partial voluming effects – however, the results in Figure 4.11 run counter to that hypothesis. This is likely a consequence of using a dimensionless representation of homogeneity, which is internally normalized for each dataset as a ratio of the within-to-between parcel variances. This metric allows for the direct comparison of homogeneity estimates across datasets, instead of representing the raw variance estimates.

Predicted parcellations in the HCP test set were as homogeneous as the corresponding ground truth maps generated by [9], both explaining roughly 67% variance in functional connectivity (Paired *t*-test: T=0.305, p=0.76). Scalar metric homogeneity estimates were comparable between the parcellations predicted by the optimal model and the corresponding ground truth maps (myelin (T: 0.176, p: 0.86), sulcal: (T: -4.44, p: 1.7e-6), curvature: (T: -1.74, p: 0.082), thickness: (T: -2.55, p: 0.012)).

Table 4.1: Model classification accuracy as a function of network architecture and parameterization. Models were trained on 400 15-minutes datasets, and tested on 60-minute test data. Boxed values indicate the default parameter values: 3 layers, 32 hidden channels, a dropout rate of 0.1, 4 attention heads (for the GAT and JKGAT networks), and an LSTM aggregation function (for the JKGAT network). We varied one parameter at a time, keeping the other parameters fixed. The best performing model was the GAT network with 6 layers, achieving a mean classification accuracy of 67.60%.

		Model			
Parameter	Value	Indep-NN	GCN	GAT	JKGAT
Network Depth	$\boxed{3}$	62.64%	64.93%	67.02%	66.71%
	6	61.13%	65.14%	**67.60%**	67.33%
	9	57.72%	64.76%	67.36%	67.42%
Hidden Channels	16	60.54%	62.60%	66.37%	66.12%
	$\boxed{32}$	62.64%	64.93%	67.02%	66.71%
	64	63.84%	66.24%	67.15%	67.15%
Dropout Rate	$\boxed{0.1}$	62.64%	64.93%	67.02%	66.71%
	0.3	60.74%	63.94%	66.72%	66.58%
	0.5	58.34%	63.10%	65.45%	65.39%
	0.7	55.63%	61.18%	62.70%	62.60%
Attention Heads	$\boxed{4}$			67.02%	66.71%
	8			67.39%	67.30%
	12			67.56%	67.29%
Aggregation Function	concat				66.85%
	$\boxed{\text{LSTM}}$				66.71%

4.5.4 Network architecture impacts model performance

As noted in Section 4.5, we first optimized over network algorithms and architectures using the S500 dataset, and then utilized the S1200 dataset for model evaluation. We fixed the features used for network optimization to the regionalized connectivity features. We examined how varying each network parameter impacted model classification accuracy

(Table 4.1). As mentioned in Section 4.5.1, the best performing model was the GAT network with 6 layers with a classification accuracy of 67.60% on the S500 dataset. We found that optimal performance for the GAT and GCN networks was achieved with 6 layers, 9 layers for the JKGAT, and 3 layers for the Indep-NN model. In general, classification accuracy increased with the number of attention heads, and number of hidden channels, while classification accuracy decreased with increasing feature dropout rates. Using an LSTM aggregation function rather than a simple concatenation marginally decreased classification accuracy for the jumping-knowledge networks. In contrast to our predictions, we found that the GAT networks slightly outperformed the more flexible JKGAT networks for most parameterizations.

We used a fixed validation dataset of 20 subjects to determine when to stop model training and evaluated the performance of our models using a fixed test dataset of 148 subjects. In order to determine the reliability of our accuracy estimates, we computed the standard error of classification accuracy for each model using a bootstrapped approach (Supplementary Material). We randomly sampled 100 test subjects, with replacement, out of the 148, and computed the mean accuracy for each sample, for each model. We repeated this process 1000 times, and computed the variability of these bootstrapped estimates. Standard error estimates were less than 0.5%, indicating that test set accuracy estimates are robust with respect to resampling of the test dataset.

We examined how classification accuracy in the HCP dataset was related to the scanning duration of training and testing datasets using the default model parameters (as defined in Section 4.5). When fixing test scan duration, classification accuracy improved as the training dataset size increased for all model types, with maximum accuracy achieved by graph attention network models trained on 400 15-minute duration datasets (Supplementary Material). When training dataset size and training scan duration were fixed, longer test image duration yielded more accurate predictions across the board. Predictions on 60-minute test data were more accurate than those computed on 30-minute images, which were subsequently were more accurate than those generated from 15-minute images (Supplementary Material). However, models trained on 15-minute data performed best when tested on 15-minute data, and models trained on 60-minute data performed best when tested on 60-minute data (Supplementary Material) indicating an interaction between training and testing scan duration. Similarly, when fixing training and testing scan duration, we found that including the spatial prior significantly improved classification accuracy in all architectures.

4.5.5 Incorporating functional connectivity improves model performance beyond spatial location and scalar metrics

After identifying the optimal network architecture, we examined how model performance varied as a function of which features the model was trained on. Briefly, we delineated three broad feature types: 1) scalar features corresponding to myelin, cortical thickness, sulcal depth, and cortical curvature 2) global location features corresponding to the spectral coordinates computed from the graph Laplacian and 3) connectivity features computed from the resting-state signal. In our primary analysis, we utilized connectivity features computed by regionalizing over the Destrieux atlas with 75 cortical areas identified based on cortical folding patterns. We compared these features against those computed using the Desikan-Killiany atlas (with 35 cortical areas based folding patterns) and the Yeo-17 resting-state network atlas [63]. The Yeo-17 atlas is a functional atlas of discretized resting-state networks, computed via independent component analysis. We identified the connected components of each of the 17 resting-state networks and excluded component regions with sizes smaller than 10 vertices, resulting in a map of 55 discrete functionally-derived subregions of the cortex. We also examined the performance of models trained on continuous, overlapping connectivity features representing resting-state networks computed using group-ICA and dual regression.

Table 4.2: Feature combinations tested by our optimal model. Features included in a model are marked by a "+". "Full" models include histological features, global position information, and functional connectivity signals.

Feature Sets							
	Full Feature Sets				Connectivity	Scalar	Location
	DK (F)	DX (F)	YEO (F)	ICA (F)	DX	Hist.	Spect.
thickness	+	+	+	+		+	
curvature	+	+	+	+		+	
myelin	+	+	+	+		+	
sulcal depth	+	+	+	+		+	
Laplacian	+	+	+	+			+
Desikan (DK)	+						
Destrieux (DX)		+			+		
Yeo-17 (YEO)			+				
ICA-RSN				+			

Computing connectivity features over the folding-based Destrieux atlas with 75 regions yielded increased classification accuracy over the folding-based Desikan-Killiany atlas (72.01% vs 70.08%; see models "Full-DX" and "Full-DK"). We hypothesized that computing connectivity features over a functionally-aware parcellation (Yeo-17) would yield a large improvement in classification accuracy, relative to using a folding-based atlas, but this was not the case (see "Full-DX" vs "Full-YEO"). Models trained on the Yeo-17 features had a mean classification accuracy of 71.58% (Figure 4.12.**B**). Training on spatial location or histological features alone yielded mean classification accuracies of 44.10% and 54.45%, respectively (Figure 4.12.A). However, training on features defined by resting-sate networks had clear performance benefits. Models trained on ICA dimensions of 15, 25, 50 and 100 generated mean classification accuracies of 75.34%, 77.79%, 79.68%, and 79.91% respectively (Figure 4.12.**C**). Similarly, incorporating the logit mask also improved model performance. However, the mask added diminishing returns – models that performed better without the prior improved less by including it. Models trained on higher-dimensional ICA resting-state networks (50 and 100 networks), performed almost as well without the spatial prior as they did with it.

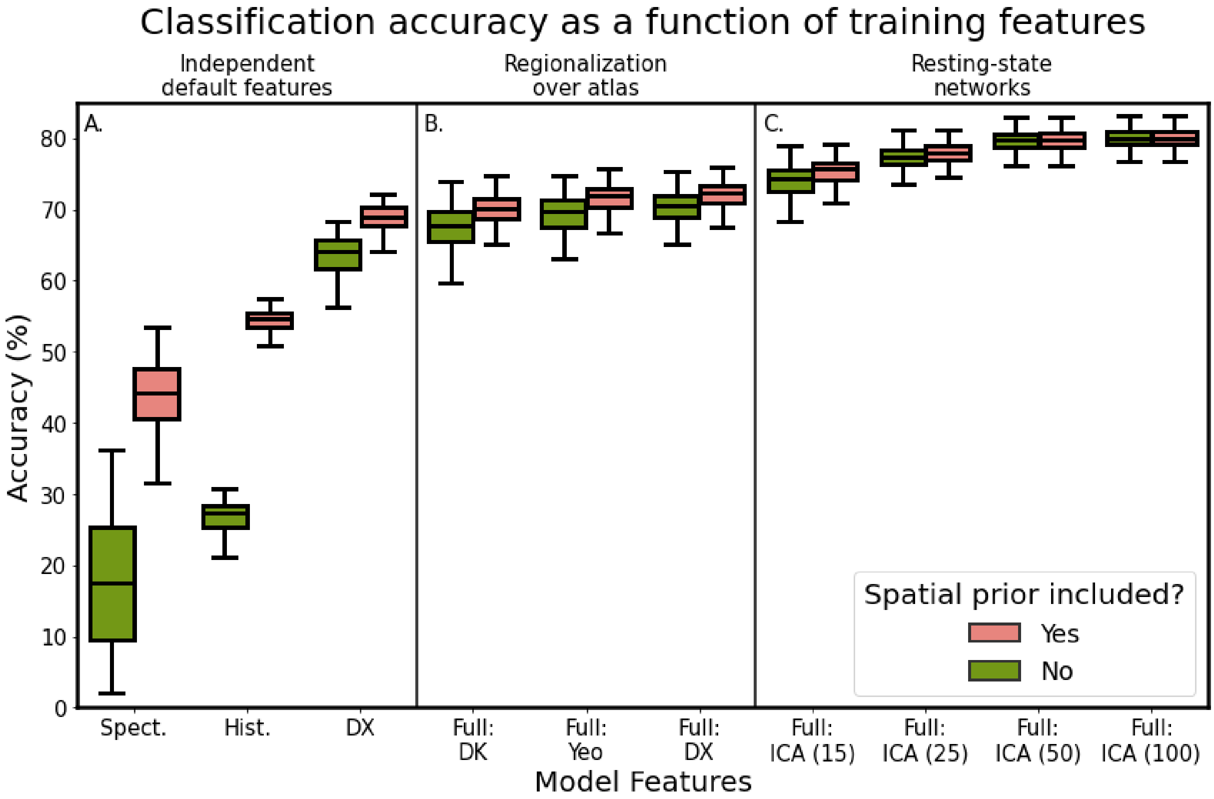


Figure 4.12: Classification accuracy as a function of model features, using the optimal model architecture for (**A**) single feature types, (**B**) regionalization over different cortical atlases, and (**C**) independent component analysis features. Refer to Table 4.2 for a description of each feature set.

Late into our analysis, we learned of differences in the preprocessing steps used to generate the minimally-preprocessed HCP resting-state data, and to generate the subject- and group-level HCP-MMP parcellations. Specifically, the S500 and S1200 data releases were preprocessed using different surface registration algorithms: MSMSulc and MSMAll [14, 15]. A consequence of these preprocessing differences is that data

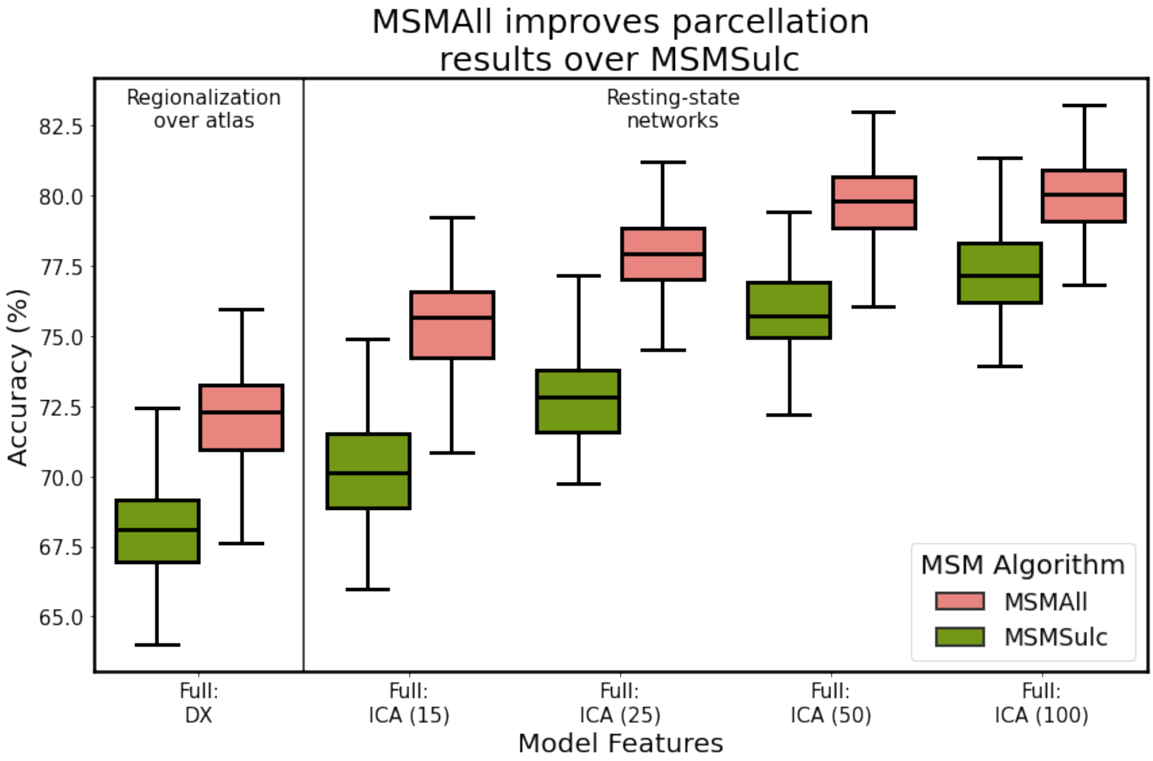


Figure 4.13: Classification accuracy as a function of HCP data release and corresponding multi-modal surface matching algorithm. S500: MSMSulc [14], S1200: MSMAll [15].

from the S1200 release is better aligned with the subject-level labels provided by MG. After performing network optimization using the S500 data, we evaluated final model performance on the S1200 dataset. Figure 4.13 illustrates model performance after training on each independent dataset. We found that utilizing the S1200 dataset showed significant improvements in mean classification accuracy by upwards 5%, relative to the S500 dataset. This indicates that the surface registration algorithm choice plays a critical role in cortical segmentation quality.

4.5.6 Resting-state algorithm choice impacts model performance

In Figure 4.14, we examined model performance when trained on subject-level ICA components computed using an *ad hoc* approach (dual regression) versus a more statistically-

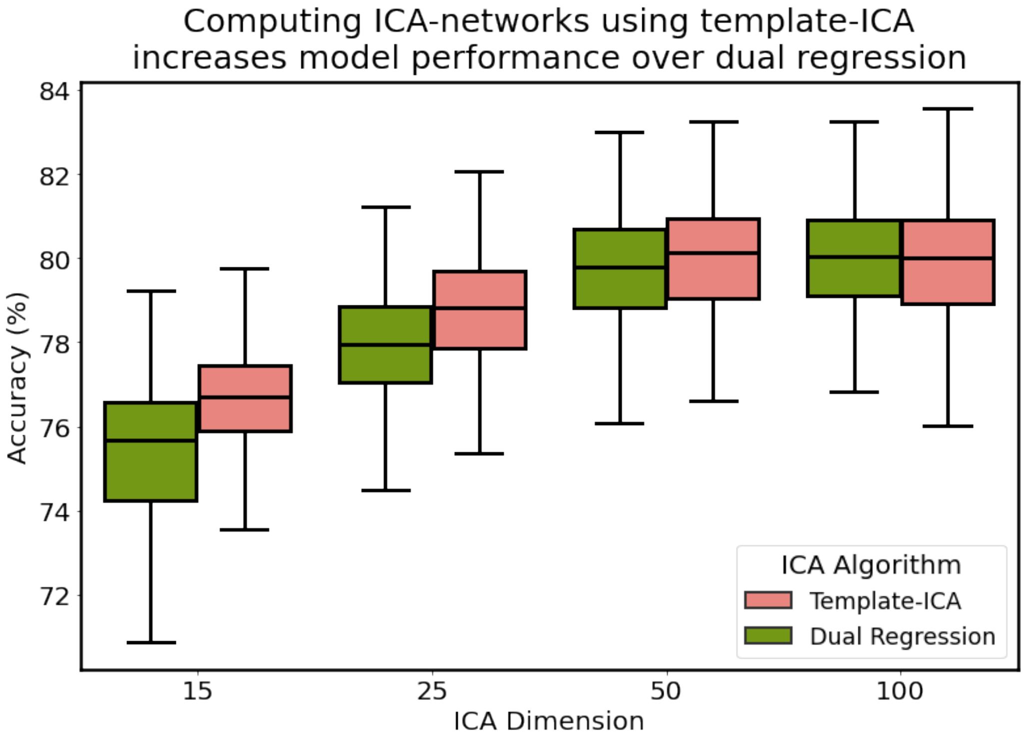


Figure 4.14: Classification accuracy as a function of ICA algorithm.

Table 4.3: Model performance as a function of ICA algorithm: dual regression (left) and template-ICA (right). We examine model performance as a function of ICA dimension, and test image duration.

		ICA Algorithm					
		Dual regression			Template-ICA		
		15min	30min	60min	15min	30min	60min
ICA dimension	15	72.63%	74.31%	75.33	75.67%	76.12%	76.56%
	25	75.21%	76.87%	77.79%	78.18%	78.46%	78.71%
	50	77.48%	78.95%	79.68%	79.27%	79.68%	**79.97%**
	100	78.05%	79.36%	79.91%	79.16%	79.63%	79.91%

driven approach (template-ICA). We found that, at lower ICA dimensions, models trained on template-ICA features outperformed those trained on dual regression features. However, at higher dimensions (e.g. 50 and 100 resting-state networks), performance is comparable.

We interrogated these performance differences further, by examining ICA algorithm performance as a function of test image duration. We found that classification differences between dual regression and template-ICA were larger for shorter-duration test images and at smaller ICA dimensions. At longer time-scales, resting-state networks computed using dual regression will be more salient, relevant to those computed at shorter time-scales. The benefits gained by using template-ICA, in terms of nuisance parameter identification and resting-state network denoising, are more relevant for shorter-duration images. As such, performance benefits are seen primarily for the 15- and 30-minute image durations at lower dimensions.

4.6 Discussion

In this analysis, we presented a general cortical segmentation approach that, given functional connectivity information and a set of corresponding training labels, can generate cortical parcellations for individual participants. This approach to segmenting the cortex requires accessible MRI acquisition sequences and standard morphological parcellations as inputs. We compared three different graph neural network variants to a baseline fully-connected network. We found that, in all cases, graph neural networks consistently and significantly outperformed a baseline neural network that excluded adjacency information. We identified the best performing model and explored its performance with respect to various metrics like segmentation accuracy, prediction reliability, and areal homogeneity in two independent datasets.

Predictions generated for both the HCP and MSC datasets were highly reproducible. However, we found that nearly twice as much resting-state data was required in MSC subjects to achieve the same reproducibility estimates as in the HCP data. Predictions generated on the HCP dataset were more reproducible than the ground truth maps themselves [9], while predictions in the MSC data were roughly as reproducible as the ground-truth parcellations. This may in part be due to the way we trained our models. Models were trained on repeated samples of BOLD images, such that for a given training subject, models were shown four BOLD datasets. This likely enabled the models to better learn the mapping between a given subject's unique BOLD signature, and its cortical map. Another possible explanation is that the ground truth maps were generated using a linear perceptron model, which does not take into account any spatial relationships between data points, while graph neural networks do take this spatial structure into account. It is likely the case that the perceptron model could not adapt to utilize spatial dependencies in the BOLD signal in local neighborhoods and thereby failed to fully learn unique subject-specific connectivity fingerprints, and consequently learned more variable parcellations.

The optimal model predicted parcellations that were roughly as homogeneous as the ground truth maps when considering multidimensional connectivity features and univariate scalar features. Though the models considered in this analysis are capable of learning parcels that capture inter-areal variation of functional brain connectivity and other cortical features, homogeneity as a measure of parcellation quality is an imperfect metric and should be used judiciously. For example, the primary sensory areas can be further divided into five somatotopic subregions corresponding to the upper and lower limbs, trunk, ocular, and face areas [11]. These subdivisions correspond well with task-based fMRI activity and gradients in myelin content, indicating that the parcels learned by GNNs in our analysis still incorporate significant variability due to the aggregation of signals from different somatosensory areas. While learning homogeneous regions is important in order to effectively capture spatial biological variation, maximizing homogeneity was not the training criterion for this analysis.

As noted in Section 4.3, the MSC study applied different preprocessing steps than the HCP. Specifically, the MSC did not perform FIX-ICA to remove noise components from the BOLD images and utilized the FreeSurfer spherical surface registration to bring surfaces into spatial correspondence with one another instead of the multi-modal surface matching algorithm [14, 15]. Given that the MSC dataset did not have "ground truth" labels against which we could compare predictions made on the MSC data, we compared predictions against the HCP-MMP atlas [9]. As expected, predictions generated on the HCP dataset more closely resembled the HCP-MMP atlas than predictions made from the MSC dataset (the HCP-MMP atlas was derived as a group-average of individual ground truth parcellations). Nevertheless, we found that correspondence of MSC predictions with the atlas followed similar trends with respect to testing image duration. We believe some discrepancy in results between the HCP and MSC datasets can be attributed to the differences in dataset-specific preprocessing choices noted above, although the relationship between methodological choices and parcellation outcome requires future analyses. Performance differences across the two datasets are also possibly a result of the models learning characteristics inherent to the training (HCP) dataset, and thereby performing better on hold out subjects from that same dataset.

Our optimal model was the 6-layer graph attention network, trained and tested on resting-state network components computed using a 50-dimensional ICA. This model performed as well with the spatial prior as it did without. However, models trained on regionalized connectivity features benefited from including the spatial prior. We believe it would be prudent for future studies to include a spatial prior of some form into their classification frameworks. Interestingly, predictions on HCP test subjects resembled the HCP-MMP atlas more closely than they resembled their ground truth counterparts, which might in part be driven by the specific form of the prior. We made the assumption that cortical map topology is relatively conserved across individuals. This assumption may be too conservative and may reduce model sensitivity to atypical cortical connectivity patterns. Nevertheless, there is evidence our GNN models learn subject-specific topologies of cortical areas, rather than simply learning *where* a cortical parcel usually is. Importantly, we found that the optimal GAT model could identify three unique topologies for area 55b (typical, shifted, and split) and that predictions generated by our model replicated, with high fidelity, the same spatial organization patterns as identified in [9]. This indicates that the model is capable of learning unique connectivity fingerprints of each cortical area on a subject-by-subject basis, rather than simply learning the group average fingerprint. As such, we do not believe that including the spatial prior in its current form inhibits the ability of the graph neural network models used in this analysis to identify atypical cortical topologies.

We compared three different graph neural networks: graph convolution networks, standard attention networks, and jumping-knowledge networks. We hypothesized that JKGAT networks would significantly outperform GAT networks due to the increased flexibility to learn optimized node-specific network depths. In their original formulation of the jumping-knowledge network architecture, [80] found that including the jumping-knowledge mechanism improved model performance relative to the GAT in almost all of their comparisons. However, we found this not to be the case. This may be a consequence of the increased number of estimated parameters in the JKGAT networks, relative to the GAT – the jumping-knowledge aggregation layer learns the parameters for the aggregation function cells in addition to the attention head and projection matrix weights learned in the GAT networks. The lower classification accuracy at test time is possibly the result of model over-fitting, necessitating a larger training dataset. It is possible that the jumping-knowledge mechanism is generally more useful in the case where graph topologies vary considerably across a network, as opposed to more regular graphs such as cortical surface data.

As expected, network performance was dependent on both the size and duration of the training set, and duration of the testing data. Classification accuracy increased when models were trained on larger datasets consisting of shorter-duration images.

Conversely, accuracy increased when models were deployed on longer-duration test data. We examined performance of our models on images of long scanning durations by concatenating multiple sessions together (30 / 60-minutes in the HCP, and 60 / 150 / 300-minutes in the MSC). It is unrealistic to expect study participants to be able to lay in an MRI scanner for single sessions of these lengths. However, it is useful to examine how model performance is impacted by tunable parameters like scan duration in order to best guide image acquisition in future studies. We found that utilizing repeated scans on individual subjects as independent training examples, rather than concatenating repeated scans together into single datasets, significantly improved our classification frameworks. This likely speaks to the ability of neural network models to generalize better to noise in the datasets. Training models on multiple samples of shorter-duration images more accurately captures the individual variability in the resting-state signal than fewer longer-duration images, thereby allowing the networks to more accurately learn a mapping between functional connectivity and cortical areal assignments.

Our methodology could be improved in a variety of ways. We chose not to perform intensive hyperparameter optimization, and instead focused our efforts on overall performance of the various network architectures as a function of network parameters and data parameters, and the applicability of trained models to new datasets. However, in the case where a classification model is meant to be distributed to the research community for open-source use, it would be prudent to perform a more extensive search over the best possible parameter choices.

The utility of functional connectivity has been shown in a variety of studies for delineating cortices [7, 26, 38, 43]. However, in recent years, using diffusion tractography for learning whole-brain cortical maps has been underutilized, relative to functional connectivity [40, 64, 92]. Given cortical maps defined independently by tractography and functional connectivity, it is difficult to "match" cortical areas across maps to compare biological properties, so heuristics are often applied. Few studies have simultaneously combined functional connectivity and tractography to better inform the prediction of cortical maps. Recent work has extended the idea of variational auto-encoders to the case of multi-modal data by training coupled auto-encoders to jointly learn embeddings of multiple data types. In [93], the authors apply this approach to jointly learn embeddings defined by transcriptomics and electrophysiology that allow them to identify cell clusters with both similar transcriptomic and electrophysiology properties. Future work could apply similar ideas to aggregate functional and diffusion-based connectivity signals.

In Chapter 3 and here in Chapter 4, we discussed methods that learn discrete partitions of cortical tissues, with partition boundaries that best conform to the specific functional and neurobiological properties of a given individual. As noted in Chapter 4, accounting for spatial information yields significant improvement in the fidelity of these learned cortical maps.

However, the question remains of how *useful* these approaches might be in a clinical setting. For example, can we glean any meaningful information from these methods with respect to disease diagnosis or prognosis? Various neurological disorders, such as Alzheimer’s or Parkinson’s disease, for example, are characterized as "disfunctional" disorders – that is, disparate areas of the cortex, which are strongly coupled or connected in healthy individuals, becomes increasingly disconnected as these diseases progress. One avenue of exploration might be to further interrogate the classification *probabilities* generated by a model for the relevant labels. We expect that, for a confident label assignment, that the probability of the "correct" label should be high, and low for the "incorrect" labels – can we relate these probabilities to disease status? From an information theoretic perspective, we might see the *entropy* of classification probabilities increase as a function of disease, or as a function of disrupted functional connectivity.

One limitation of our analysis concerns the use of different versions of the multi-modal surface matching for cortical surface alignment for the S500 HCP data release ([11, 14]) [11], the S1200 release ([15]), and for the subject-level HCP-MMP parcellations [9], which used a different regularization term. These differences between the three registration methods result in a slight spatial misalignment between the training labels and the cortical features. While the S500 data release utilized MSMSulc, a spherical surface registration driven by cortical folding patterns, the S1200 release utilized MSMAll, and incorporated functional connectivity into the spatial resampling step. [9] used a prototypical version of MSMAll in addition to MSMSulc, and thereby incorporated additional features derived from resting-state networks to drive the surface matching process. Importantly, this discrepancy between the training labels and training features is not a flaw in our methodology itself, and correcting for this difference in the registration approach would only improve the results of our analysis. As we showed in Figure 4.13, incorporating MSMAll-processed data from the S1200 dataset, instead of MSMSulc-processed data from the S500 dataset, improved model classification accuracy by nearly 5%. We hypothesize that this improvement would only increase if we had access to the data processed with the prototypical version of MSMAll. Based on the comparisons of subject-level predictions with the subject-level ground truth MMP maps, our models performed well in spite of these registration discrepancies. Our results lend evidence to the robustness of graph neural networks for learning cortical maps from functional connectivity.

Finally, participants in both the HCP and MSC studies were healthy young adults, and the datasets had been extensively quality controlled. Little to no work has been done on extending connectivity-based classifiers to atypical populations, such as to individuals with neurodegeneration. It is unknown how a model trained on connectivity properties from healthy individuals would perform in populations where connectivity is known to degrade. While our model (and that developed by [9]) predict maps based on healthy individuals, it is possible that some studies would need to train population-specific models.

Data Availability Statement

The group-level HCP-MMP parcellation can be found here. The Human Connectome Project dataset can be found on the ConnectomeDB here. The Midnight Scan Club dataset can be found here. Code developed for training models used in this analysis can be found here.

Chapter 5

Application: Identifying connectivity-derived stimulation sites in the inferior parietal lobe

5.1 Motivation

The overarching goal of this project is to test the hypothesis that differences in cortical field anatomy underlie individual differences in functional network characteristics, and explain differences in functional network vulnerability, when measured by cognitive task performance. The size of individual cortical fields (e.g. as defined by cytoarchitectonics) varies up to three-fold in contemporary post-mortem studies [8], but the implications of this variable anatomic endowment for system and cognitive function are largely unexplored. The results described in this chapter correspond to the first stage of this network vulnerability study: namely, identifying and characterizing the cortical fields corresponding to subregions of major resting state networks. This current work focuses on identifying components of the default mode network (DMN) and lateral frontoparietal network (LFP). However, the methods developed here are network-agnostic.

In the inferior parietal lobe (IPL), previous parcellation studies have distinguished two angular gyrus fields: PGa and PGp [37, 41] that correspond spatially with histologically-defined areal boundaries [5]. Activation of LFP and deactivation of DMN are observed during engaging attentional tasks (e.g., Eriksen flanker task) and connectivity within both networks is also observed during the resting state fMRI and EEG. A single cortical field preferentially interacts with the network of regions it is associated with. PGa is thought to act as node of the LFP whereas PGp likely corresponds to a node of the DMN [94]. Similarly, PGa is characterized by increased structural connectivity to nodes of the LFP, while PGp shows increased structural connectivity with the ventromedial prefrontal cortex, posterior cingulate, and parrahippocampal regions (e.g. nodes of the DMN) [94]. The full proposed analysis will follow a network-based approach to functional anatomy, to examine the effects of cortical field size and virtual lesions of a cortical field on the coordinated activities of the LFP and the DMN. As prerequisite, we need to be able to accurately and reliably identify cortical fields of interest on the basis of various imaging modalities.

Our fundamental hypothesis is that smaller surface area of a cortical field results in: 1) lower extent-related measures of activation and functional connectivity; and 2) less resilience to TMS-induced disruption. Differential endowment of specific cortical fields may help to predict individual differences in impairment when node function is compromised by aging, damage, or disease. At the same time, the coordinated network activity in the LFP and DMN are known to be associated with different task performance. For example, failure to engage the LFP and failures in coordinating LFP and DMN are associated with lower task performance in a variety of experimental paradigms. We will investigate individual sensitivity to interrupting network function with repetitive transcranial magnetic stimulation (rTMS). We hypothesize that low frequency deactivation of LFP nodes using rTMS will impair task performance, and more so with individuals with smaller LFPN surface areas. Analogously, rTMS over PGp should facilitate task performance, and more so for individuals with smaller PGp surface area. The proposed research will investigate how individual differences in cortical anatomy (size of a cortical node) affect network-characteristics, and how these two variables are related to resilience (as measured by behavioral tasks) to virtual TMS lesions.

We will evaluate connectivity-based parcellation to isolate the LFP and DMN components in the IPL, cross-validating the distinction of PGp from PGp against resting state as well as task-evoked MRI and EEG measures. To this end, we developed a turn-key pipeline to identify candidate rTMS stimulation sites from structural T1 and multi-echo fMRI data. Briefly, given an input structural MPRAGE and resting state fMRI image, the processing pipeline includes the following steps:

1. Motion and distortion correction
2. Multi-echo image denoising using tedana [95, 96, 97]
3. Spatial normalization between BOLD and T1 images
4. Volume-to-surface mapping of BOLD signal
5. Resting state network estimation
6. Field modeling of resting state networks to identify network centroids [22]

Steps **1-4** are related to image preprocessing choices – while parameter optimization, algorithm and software choice, and ordering of various preprocessing steps warrants examination, these steps are not discussed here further. We discuss steps **5** and **6** of this pipeline in more detail, and present preliminary results associated with the first phase of this analysis. More information on the full details of this pipeline can be found in the project tutorial.

Table 5.1: Parameters of five image acquisition protocols. The boxed protocol indicates the optimal protocol.

Image Acquisition Parameters

Protocol	Multi-band	SENSE	Voxel size (mm^3)	TR (ms)	TE (ms)
ABCD-like	6	1	2.25 x 2.25 x 2.40mm^3	850	30
3mm SE	4	1	3mm^3	800	30
3.5mm ME	3	2	3.5mm^3	1000	20/40/60
3.0mm ME	2	2.5	3.0mm^3	2000	20/41.1/62.2
$\boxed{\text{2.5mm ME}}$	2	3	2.5mm^3	2200	20/23.1/68.2

5.2 Data

We collected MRI data on two participants over two sessions. Table 5.1 shows each of the five candidate image protocols. After extensive parameter testing, we identified multi-echo acquisition schema **2.5mm ME** as the optimal acquisition protocol for the BOLD images. The preprocessing pipeline was applied to each subject, for each of the two sessions.

5.3 Pipeline

5.3.1 Computing resting state networks

We estimated group-level resting state networks using a subset of 268 subjects from the Human Connectome Project dataset [11], using MELODIC's Incremental Group-ICA (MIGP) algorithm [88]. Briefly, MIGP iteratively combines subject-level resting state data to estimate group-level resting state components. The algorithm is initialized with a subset of subject-level resting state data to estimate an initial spatial embedding using the singular value decomposition (SVD), retaining the first k spatial singular vectors. At every proceeding iteration of the algorithm, MIGP appends the resting state matrix of a new subject to the previously estimated singular vectors, computes the SVD of this concatenated matrix, and retains the updated first k spatial singular vectors. After iterating through all subjects, the FastICA algorithm is applied to the group-level spatial singular vectors to estimate statistically independent group-level resting state networks. Group-level components were dual-regressed onto each individual subject's BOLD data to estimate subject-level resting state networks. The resting state network maps are represented by β coefficients from the dual-regression linear model. A large β coefficient value for resting state network at a given vertex indicates that the corresponding network time-course is strong in the BOLD time-series of that vertex.

We identified three networks of interest in the left hemisphere and two networks of interest in the right hemisphere (Figure 5.1). In the left hemisphere, we identified a lateral-frontal parietal component (LFP), and an anterior and posterior default mode network component (aDMN and pDMN, respectively). In the right hemisphere, we identified a single LFP component and a single DMN component.

5.3.2 Identifying network centroids

We were interested in the parietal components of the LFP and DMN networks as candidate target sites for virtual lesioning via rTMS stimulation. As such, we sought to identify the "centroid" of each network component. We restricted our search space to consider only those areas of each functional network component that intersected with the angular and supramarginal gyri, as defined in by the Desikan-Killiany folding-based cortical parcellation. We identified network centroids using a connective field model approach applied to the resting state network coefficient maps (described in more detail in [22] and in Section 10). Briefly, this method fits an isotropic Gaussian to scalar data distributed over a cortical mesh, by estimating a mean μ, (distribution center of mass, and spatial location on the cortical manifold) and a sigma, σ, (distribution spread). For each resting state network p, we estimated a μ_p and σ_p parameter, and defined the network centroid as μ_p.

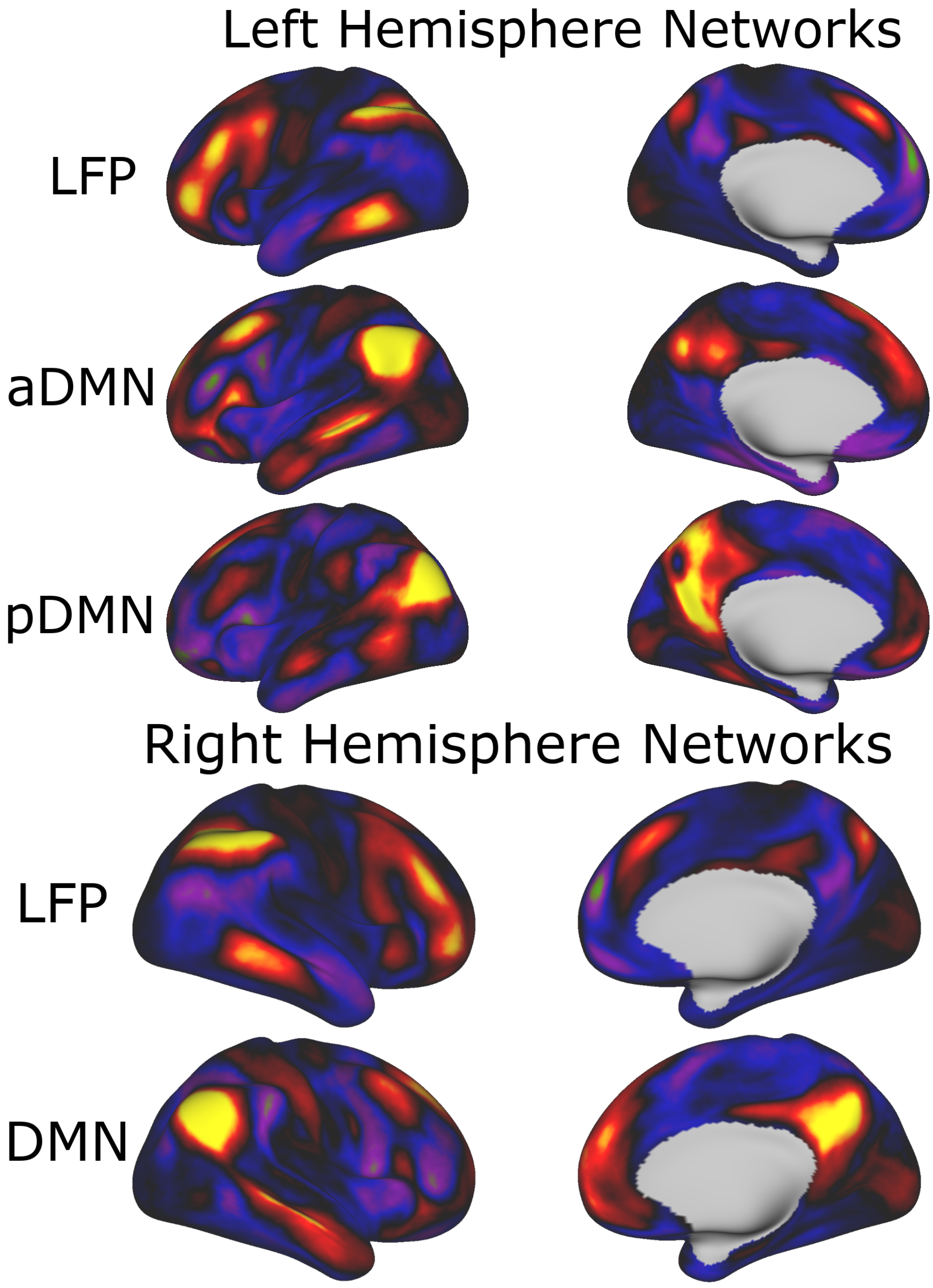


Figure 5.1: Estimated group-level resting state networks in the left and right hemisphere.

5.4 Results

5.4.1 Centroids correspond well with functional, but not multi-modal atlases

We examined whether centroid coordinates identified using dual-regression coefficient maps corresponded to areal boundaries defined by functional network atlases and a multi-modal parcellation atlas. Figure 5.2 shows the centroid locations of each group-ICA resting state network overlaid on the functional Yeo-7 and Yeo-17 atlases [63], and multi-modal HCP-MMP atlas [9]. The centroid locations corresponded well with the resting state network regions of interest in both the 7 and 17-network resolutions, in both the left and right hemispheres. However, coordinates did not align with the areal boundaries of the multi-modal HCP-MMP atlas. Specifically, in the left hemisphere, the LFP centroid was identified near the boundary of parietal area PFm and intraparietal sulcus region IPS2. The aDMN component was identified at the border of PGa and PFm, while the pDMN component was found near the posterior boundary of PGs. While the area of pDMN network activation did overlap with area PGp, the areal boundaries of PGp were more posterior to this area of activation. Similarly, in the right hemisphere, the DMN components was located in area PGi (anterior to PGp), while the LFP component was located at the boundary of area PFm and IPS2.

Of course, cortical areal boundary definitions are subject to the algorithm used to compute the parcels. [9] used a gradient-based approach to identify sharp transitions in cortical features (including resting state networks), and showed that these areal boundaries correspond closely with task-based activations. However, the assignment of each parcel to a name (i.e. posterior parietal area to "PGp", posterior intraparietal sulcal to "IPS1" and "IPS2", etc.) is to some degree dictated by the neuroanatomist(s) involved. It is possible that the areal boundaries of PGi, PGs, and PGp defined in the HCP-MMP atlas, and computed using *in vivo* features, do not correspond to those areal boundaries

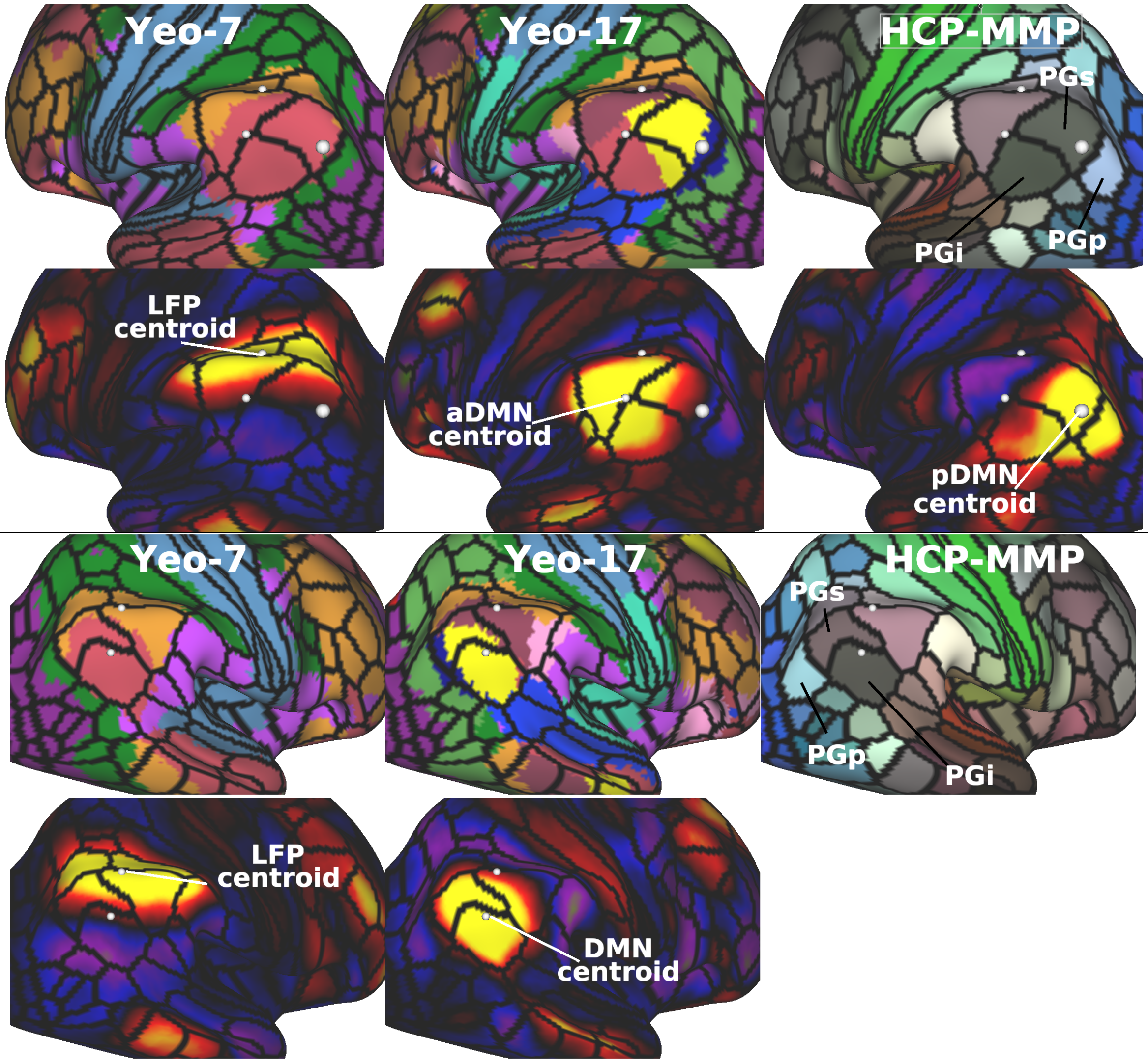


Figure 5.2: Centroid coordinates of the group-ICA resting state networks corresponding to the identified LFP, aDMN, and pDMN components in left hemisphere (top), and LFP and DMN components in right hemisphere (bottom). The left/right subfigures are organized as follows: top left: Yeo-7 functional parcellation, top middle: Yeo-17 functional parcellation, and top right: HCP-MMP parcellation. Black lines indicate areal boundaries of the HCP-MMP atlas.

identified by [5, 32] using *in vitro* features, or to the areas of LFP and DMN activation identified by [94].

5.4.2 Subject-level parcellations using GNNs

Our ultimate goal for developing the methodology described in Chapter 4 was to map the cortices of clinical-quality data with application to healthy and neurodegenerative patient populations. We applied the optimal model identified in Section 4.5.1. As a reminder, the optimal model was a graph attention network with 6 layers, trained on resting state features computed using independent component analysis. Figure 5.3 shows predictions computed for each of our test subjects, for each test-retest scan.

Relative to the predictions generated on HCP and MSC data, these test-retest maps were considerably more noisy, especially near areal boundaries. However, evaluating prediction "accuracy" was not possible, since these subjects did not have ground truth maps against which to compare predictions. When we examined the feature maps used to compute these predictions (resting state networks, myelin, cortical thickness, sulcal depth), and when comparing these feature maps to those of HCP and MSC subjects, we found the computed features are also more noisy than those in the HCP and MSC datasets. In fact, the subject-level resting-state networks displayed systematic signal increases near the somatosensory and temporal cortices, specifically along cortical gyri (not shown). We hypothesize that these signal differences reflect errors related to either 1) distortion correction of the BOLD image, and/or 2) registration of the BOLD image to the structural T1. We hypothesize that poor classification performance is related to these signal errors.

As a reminder, predicted parcellations in the Midnight Scan Club were empirically quite similar to those predicted in the HCP dataset. As discussed in Chapters 2 and 4, the Midnight Scan Club acquired clinical-quality data using a similar image protocol to the optimal protocol discussed here. The MSC acquired single gradient echo BOLD images with a voxel size of $4mm^3$, TR=2200ms, and TE=27ms. Our optimal acquisi-

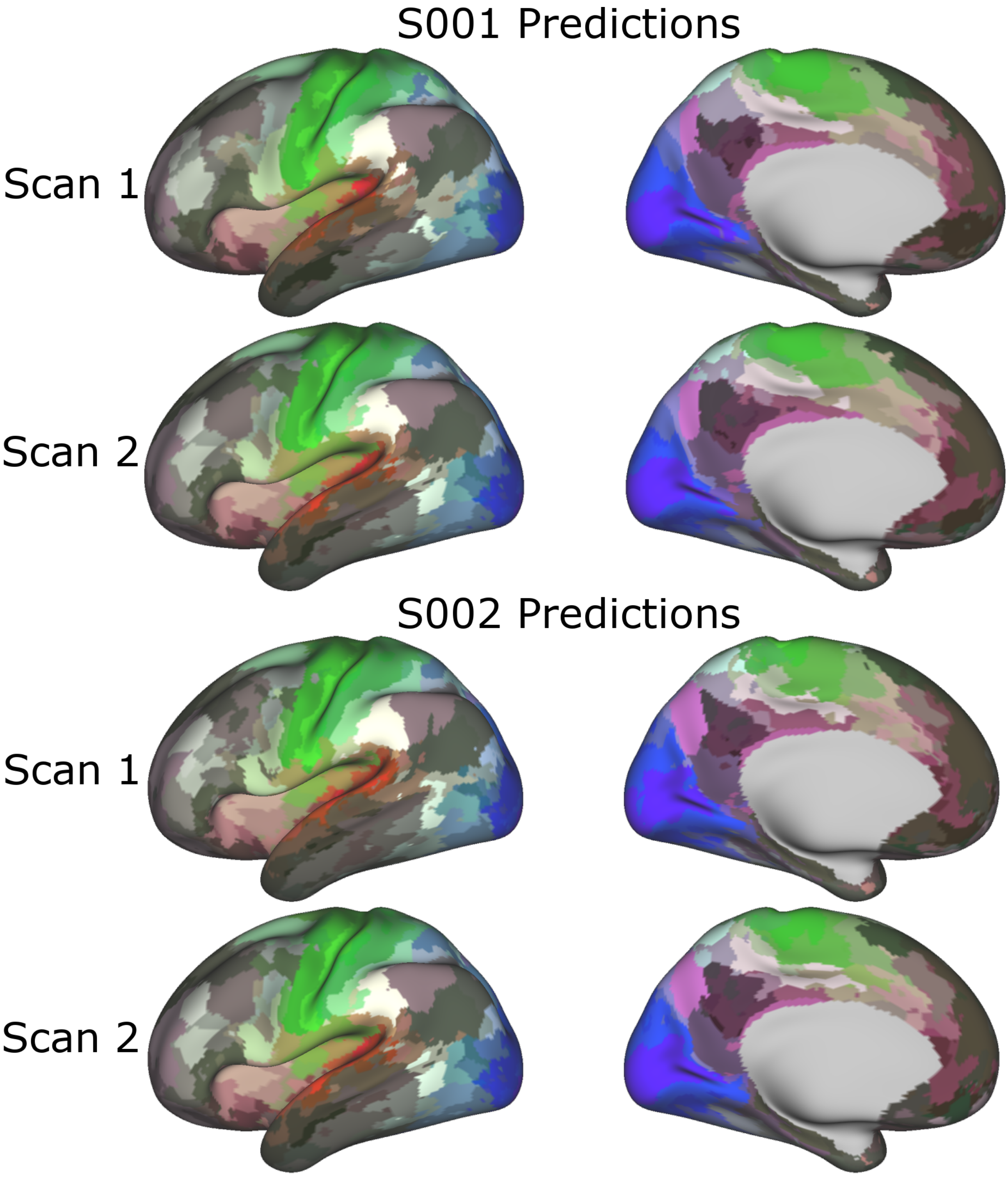


Figure 5.3: Subject-level connectivity-derived cortical maps generated from clinical-quality data, computed using a graph attention network model trained on resting state connectivity and cytoarchitectural features.

tion scheme acquired multi-echo fMRI with a voxel size of 2.5mm^3, TR=2200ms, and TE=20,23.1,68.2ms. Previous work has shown that multi-echo fMRI improves BOLD signal sensitivity and more accurately enables the removal of non-BOLD artifacts, relative to single-echo fMRI [97, 98]. As such, we believe the *quality* of data collected using our protocol (2.5mm ME) to exceed that of the MSC dataset. To this end, if distortion correction of the BOLD images and/or alignment with the T1 images can be improved, we expect that estimated subject-level resting state networks, and consequently the parcellation predictions and estimated network centroids, should improve accordingly.

5.4.3 Test-retest reliability of centroid statistics

For each resting state network centroid, we examined the distance between pairs of centroid coordinates, test-retest reliability of centroid coordinates, and the distance from each centroid to the skull. In order to properly assess the resilience of a resting state network to virtual lesioning with rTMS, we need to be able to independently isolate each resting state network from one another to minimize simultaneous co-stimulation of both networks. For example, if we are interested in performance during a working memory task with and without disruption via TMS, we must be able to stimulate the parietal centroid of the LFP network, while also simultaneously minimizing the stimulation of other functional networks. Otherwise, we risk co-stimulation of both networks, and may thereby confound cognitive interpretations.

With respect to estimating centroid coordinates, we sought to maximize the distance between centroids from different functional networks, and minimize the distance between centroid coordinates of the same network. Due to the 3-dimensional nature of the TMS magnetic field, we also minimized the distance from centroid coordinates to the skull. For each of these measures, we compared the statistics computed for each subject, with those computed for subjects from the HCP and MSC datasets. As a reminder, the HCP acquired data for each subject from four independent sessions, while the MSC acquired data for each subject from ten independent sessions. Results for the following analyses are averaged across 100 HCP subjects (4 sessions per subject) and 10 MSC subjects (10 sessions per subject).

Table 5.2 shows both the Euclidean distance and distance along the cortical manifold between test-retest centroid locations for each acquired subject. Test-retest centroids were on the order of $1cm$ apart, consistent with results from the HCP and MSC datasets (Figure 9.11).

Table 5.3 shows distances between network centroids for each acquired subject. In the left hemisphere, the LFP and pDMN were furthest apart, mirroring results in the MSC and HCP datasets (Figure 9.12). Additionally, within-subject standard errors were low – on the order of millimeters.

Table 5.4 and Figure 9.13 show distance between estimated network centroids and the skull. The LFP centroid was often hidden deep in the intraparietal sulcus (IPS), at least in the left hemisphere. However, for some subjects, the LFP centroid was located more closely to the periphery near the supramarginal gyrus, indicating that there might be considerable variability between subjects in the potential rTMS stimulation depth.

Table 5.2: Within-network centroid distances (mm). For each subject, for each network and test-retest scan, we estimated a network centroid. We computed the test-retest distance between centroids for the same network, and report distance estimates in Euclidean space, and along the cortical manifold.

(a)

Subject: S001

Hemisphere	Network		
L	**LFP**	**aDMN**	**pDMN**
euclidean (mm)	3.07	10.85	4.68
geodesic (mm)	3.07	12.06	4.89
R	**LFP**	**DMN**	
euclidean (mm)	10.03	5.80	
geodesic (mm)	11.69	6.25	

(b)

Subject: S002

Hemisphere	Network		
L	**LFP**	**aDMN**	**pDMN**
euclidean (mm)	5.68	3.93	3.12
geodesic (mm)	6.25	4.24	3.14
R	**LFP**	**DMN**	
euclidean (mm)	6.03	18.27	
geodesic (mm)	6.64	13.03	

Table 5.3: Between-network centroid distances (mm). For each subject, for each network and test-retest scan, we estimated a network centroid. Mean and standard deviation of distances estimated in Euclidean space and along the cortical manifold are shown for all network pairs.

(a)

Subject: S001

Hemisphere	Network Pair		
L	**LFP-aDMN**	**LFP-pDMN**	**aDMN-pDMN**
euclidean (mm)	17.11 (2.11)	24.85 (1.13)	20.79 (3.14)
geodesic (mm)	50.89 (5.09)	57.27 (0.95)	39.09 (1.99)
R	**LFP-DMN**		
euclidean (mm)	17.54 (3.49)		
geodesic (mm)	44.81 (4.25)		

(b)

Subject: S002

Hemisphere	Network Pair		
L	**LFP-aDMN**	**LFP-pDMN**	**aDMN-pDMN**
euclidean (mm)	9.49 (1.29)	38.98 (1.77)	31.58 (1.80)
geodesic (mm)	42.20 (2.57)	72.93 (2.09)	62.88 (1.12)
R	**LFP-DMN**		
euclidean (mm)	28.60 (0.92)		
geodesic (mm)	50.28 (5.96)		

Table 5.4: Centroid distances to skull (cm). For each subject, for each network and test-retest scan, we estimated a network centroid. We computed the test-retest nearest-neighbor distance from the centroid to the skull.

(a)

Subject: S001

Hemisphere	Network		
L	**LFP**	**aDMN**	**pDMN**
euclidean (cm)	3.05 (0.12)	2.79 (0.52)	2.11 (0.13)
R	**LFP**	**DMN**	
euclidean (cm)	2.59 (0.24)	2.73 (0.16)	

(b)

Subject: S002

Hemisphere	Network		
L	**LFP**	**aDMN**	**pDMN**
euclidean (cm)	3.19 (0.26)	3.04 (0.37)	1.63 (0.13)
R	**LFP**	**DMN**	
euclidean (cm)	1.89 (0.27)	2.05 (0.77)	

Chapter 6

Cortical gradients of localized functional connectivity reveal network-specific organizational properties

Eschenburg, K, Grabowski, T, Haynor, D. Cortical gradients of locally-defined functional connectivity reveal network-specific organization properties. In preparation.

Initiated by Kristian M. Eschenburg (KME) as a part of, and funded by, NSF grant BCS-1734430 for Principal Investigator Dr. Thomas Grabowski (TG): "Relationship of Cortical Field Anatomy to Network Vulnerability and Behavior". KME developed the software, performed all analyses, and wrote the manuscript, with mentorship from DH. Barring revisions or edits suggested by committee members, the following text, combined with work in Chapter 7, will serve as the draft of a journal submission.

6.1 Introduction

In the previous three chapters, we discussed methods for mapping the human cortex in order to delineate discrete or localized cortical subregions. We discussed two classification approaches to learn cortical maps, and analyzed the application of these methods to data collected in-house. As motivated in Chapter , these delineations are related to the "mesoscale" organization of the brain. However, we can dig deeper into this mesoscale organization, by examining the *intrinsic* organizational properties of these cortical regions themselves.

The human cortex displays an interesting organizational property where biological features, like myelin content, cell types, and gene expression patterns, vary smoothly in space within the gray matter. These unique spatial patterns show considerable consistency across individuals, but also capture unique individual variability. Recent studies have developed algorithmic approaches to examine this topographic organization of the cortex that generate cortical maps capturing large-scale trends in spatial variation of these biological signals. To date, various studies have examined these spatial trends applied to the whole brain and to exemplar cortical regions. However, to our knowledge, little work has been done on analyzing the overall consistency and replicability of these spatial profiles across and within subjects, or systematically examined regional variation in this topographic organization across the whole cortex. Likewise, local gradients defined on individual areas have not been decomposed into their constituent parts. Here, we used data from the Human Connectome Project to shed light on these questions using resting-state functional magnetic resonance imaging. In this analysis, our goal is to analyze cortical gradients from the perspective of whole-brain and regional connection patterns and assess the overall complexity of these recently-developed topographic biomarkers.

6.2 Background

The human brain exists as a complex hierarchical network, through which information flows from sensory organs, to the deep cortical nuclei, to sensory cortices, and further downstream to higher order areas. This information flow is mirrored by synaptic connection profiles that constrain, guide, and integrate the flow of these signals throughout the brain [20, 36].

Integration of signaling can be represented via cortical receptive fields. Receptive fields are areas of the cortex more proximal to the raw sensory signal, whose synapses all converge onto similar cortical areas. Receptive fields are a feature of the location of convergence – for example, if region A receives synaptic input from B, C, and D, then B, C, and D represent the receptive field of region A [20, 60, 99]. Receptive fields capture the notion of continuous topographic organization in the primary sensory areas via sensory stimuli, with increasing complexity in these fields as one moves into higher-order associative areas. As an example, the visual system is well known for its retinotopically organized receptive fields [100, 101]. Retinal cells in the visual field tuned to similar eccentricity and polar angles map onto spatially proximal areas in the striate cortex that vary smoothly in space. Similarly, the auditory cortex, is tonotopically organized, where smoothly varying cortical regions in the primary auditory cortex are tuned to smoothly varying transitions in pitch and tone [102].

Recently, work by [18] identified whole-brain topographic organizational patterns (deemed "gradients") using resting-state functional MRI (rs-fMRI). These patterns are represented by intrinsic axes that map onto the brain surface, that capture non-linear directions of spatial variation of biological features. The first primary axis is anchored at one end by primary sensory areas and at the other end by transmodal areas, notably identified as areas corresponding to the Default Mode Network. [24] identified similar spatially varying axes of cytoarchitectonic profiles computed from myelin density – areas in the sensorimotor cortices show low myelin density, with increasing density as when transitioning to the transmodal regions. Recent work has also identified similar profiles using microstructural patterns, including cell density and soma size, across the neocortex. Here, primary sensory areas are relatively homogeneous in their microstructural properties, with increasing disassociation and variability in the transition to associative and transmodal cortices [103]. Evidence increasingly points to the notion that the brain as a whole can be represented by intrinsic coordinates axes. Extrinsically, these axes represent directions of spatial variation of various biological signals – implicitly, however, these axes capture notions of increasing levels of information integration within the brain [104, 105].

The analysis of gradients has been extended to individual cortical areas, along with larger functional networks. [3] developed an approach using Laplacian eigenmaps to analyze the gradients of rs-fMRI signals and applied their method to the primary motor and visual areas. They show that the low-dimensional non-linear Laplacian embeddings of resting-state functional connectivity recapitulate the somatotopic organization of M1, along with the retinotopic eccentricity and polar angle maps in V1. Similar analyses by other groups have studied functional gradients in the cerebellum [106], structural connectivity gradients in the temporal lobe [92, 107], functional diversity curves and myeloarchitecture in the insular cortex [108, 109], iso-to-allocortex variations in the mesiotemporal lobe [110], and idiosyncratic variations in the Default Mode Network [111].

While there currently exist a variety of approaches to generate intrinsic gradients from brain connectivity profiles and from various other neurobiological signals, assessing the true complexity of these signals is still an open-ended question. It has been shown that high-dimensional neurobiological signals can be decomposed into a significantly smaller number of variables. Most familiar, perhaps, is the idea of using independent component analysis to decompose high-dimensional temporally-varying signals, including resting-state fMRI, EEG, ECoG, and MEG time series data, into a discrete number of spatial components, that accurately capture the spatial and temporal dynamic variations of brain function [112] [113] [114, 115, 116]. Analogously, high-angular resolution diffusion-imaging data can be represented by a much smaller number of spherical basis functions describing diffusion of water throughout the gray and white matter [10, 55, 56]. However, many of these methods suffer from the same problem: how many underlying parameters are there? Herein, we framed our study of cortical gradients from the perspective of better-understanding their complexity and functionally characterizing this complexity.

In this analysis, we were interested in three unique themes:

1. *Complexity of cortical regions*: how many "informative" gradients are there per region? How many parameters are needed to accurately characterize a single cortical area?

2. *Functional network specificity*: are gradients induced on a cortical area related to which other regions that area is mutually connected to?

3. *Local vs. global axes*: are gradients induced by local connections just perturbations of an average gradient or do they capture new spatial variation?

We present results characterizing the complexity of cortical regions by estimating the number of informative gradients in a region. We found that, for all cortical areas, only a small subset of gradients are needed to explain much of the variation in functional connectivity. In accordance with previous work on whole-brain gradients, our results show that each region is on average best characterized by three dimensions. We then discuss results showing the spatial organization is determined to some extent by functional network. And finally, we show that strength of functional connectivity between regions pairs influences how similar their spatial organization patterns are.

6.3 Data

In this analysis, we used data from 154 randomly chosen individuals from the Human Connectome Project (HCP) [11]. Structural MPRAGE data were pre-processed using a custom pipeline developed by the HCP using FreeSurfer [65] to generate surface meshes at the white/gray and pial/CSF interfaces. The surface meshes were spatially normalized to Montreal Neurological Institute (MNI) space and resampled to have 32k vertices.

For each subject, the HCP acquired 4 resting-state functional MRI (rs-fMRI) images: TR=0.720ms, TE=33ms, multi-band factor=8, FA=52 degrees, FOV=208mm x 180mm, Matrix=104 x 90 x 72. These images were minimally pre-processed using a custom pipeline developed by the HCP [11]. Briefly, this pre-processing included correction for motion and spatial distortions. Voxel-based time series were mapped onto the 32k surface mesh using a barycentric averaging algorithm, and then smoothed at FWHM=2mm to avoid mixing of signals across gyri. Surface-mapped data were spatially normalized across subjects using a multi-modal surface matching algorithm [15] and vectorized to CIFTI format, mapping each surface vertex to an index in a vector. These vector indices, referred to as "grayordinates" by the HCP, are in spatial correspondence across subjects (i.e. index i in subject j corresponds to index i in subject k).

The HCP consortium developed a pipeline to compute high-resolution multi-modal cortical parcellations (MMP) with 180 cortical areas using a spatial derivative based algorithm [9]. A classifier was trained, using the boundary-based cortical areas as response variables, and rs-fMRI / task-fmri, thickness, and myelin maps as independent variables. Here, we used the **Q1-Q6 210 RelatedParcellation** MMP, which can be downloaded here: https://balsa.wustl.edu/DLabel/show/nn6K.

We related our results to a functional resting-state atlas that assigns each HCP area to a functional network [16] (https://github.com/murraylab/hbnm). Briefly, [16] generated a region-by-region correlation matrix characterizing the resting-state functional connectivity between pairs of cortical regions. They then applied the Louvain algorithm

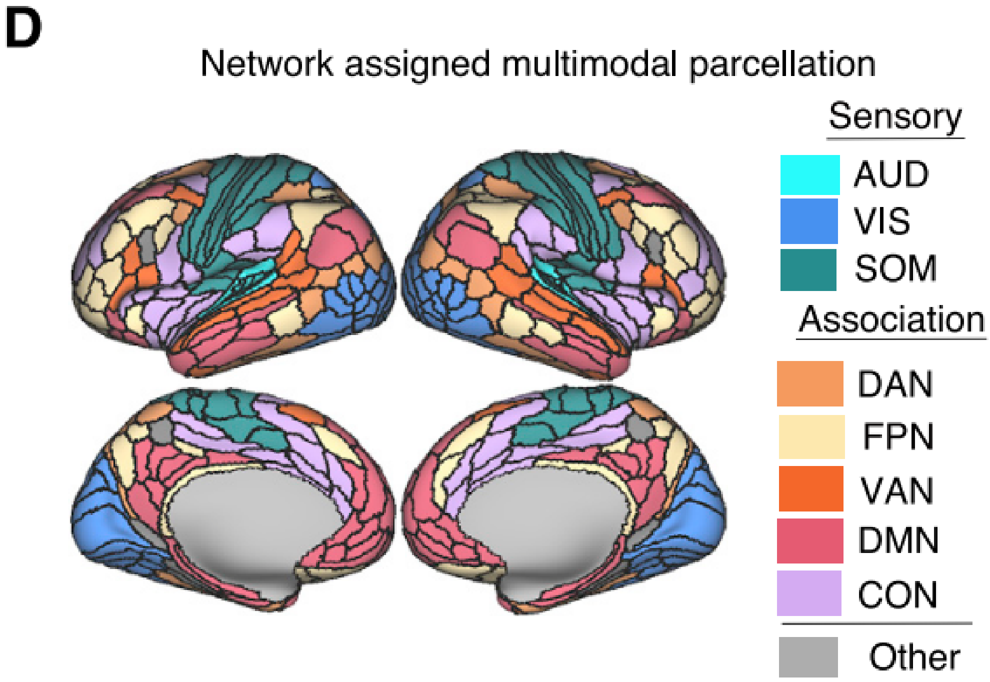


Figure 6.1: Resting-state network assignments for all 180 HCP cortical areas computed used the Louvain algorithm to assign network labels to cortical areas [16]. We used these network assignments to aggregate statistics across functional networks.

to maximize the modularity score of identified region clusters, that were then manually labeled as part of individual functional networks (see Figure 6.1).

6.4 Methods

6.4.1 Computing correlation features

We describe the steps to compute an affinity matrix from a matrix of time series measurements. Without loss of generality, assume we have matrix $T \in \mathbb{R}^{n \times t}$ with n samples, and t time points. Each time point t_i corresponds to the BOLD signal measured at a given cortical voxel at time $t = i$. We generated the Pearson correlation matrix $C \in \mathbb{R}^{n \times n}$ by cross-correlating T with itself. Each index (j, k) in C represents the Pearson correlation between the time-series for voxels j and k. We treated each row C_v for $v = 1 \ldots n$ as the feature vector for voxel v characterizing the correlation of voxel v's BOLD signal with every other voxel's BOLD signal in our set.

In the present work, we analyzed what we henceforth call "global" and "target-specific" connectivity gradients. For a given cortical area, we examined the connectivity between this area and the rest of the brain (global), and to other individual cortical regions (target-specific) (Figure 6.3). Figure 6.2 describes our general workflow, and is based loosely off of the pipeline described in [3].

For a cortical area s with n vertices, we defined the matrix of time-series signals $T_s \in \mathbb{R}^{n \times T}$. We defined a target area t with m vertices as either 1) the rest of the cortex excluding area s, or 2) a single cortical area, with time-series matrix t as $T_t \in \mathbb{R}^{m \times T}$. We computed the Pearson cross-correlation matrix between T_s and T_t as described above to generate the rectangular matrix $C \in \mathbb{R}^{n \times m}$. C is a matrix of correlation features, relating the n source vertices to the m target vertices.

For each subject q, where $q = 1 \ldots 154$, for each of the 180 regions $r \in R$ in the Q1-Q6 MMP cortical parcellation, we computed the cross-correlation matrix $C_{q,r}$. For each subject's cross-correlation matrix $C_{q,r}$, we computed the Fisher R-to-Z transform $Z_{q,r}$ to normalize the correlation values, and then took the mean

$$\bar{Z}_r = \frac{1}{|\mathrm{Q}|} \sum_{q \in Q} Z_{q,r} \tag{6.1}$$

of the Z matrices. We reapplied the Fisher Z-to-R to convert back to correlation values to generate $\bar{C}_r$. We repeated the above steps for the global time-series data, and the target-specific time-series data, to generate group-averaged global and target-specific rectangular cross-correlation matrices.

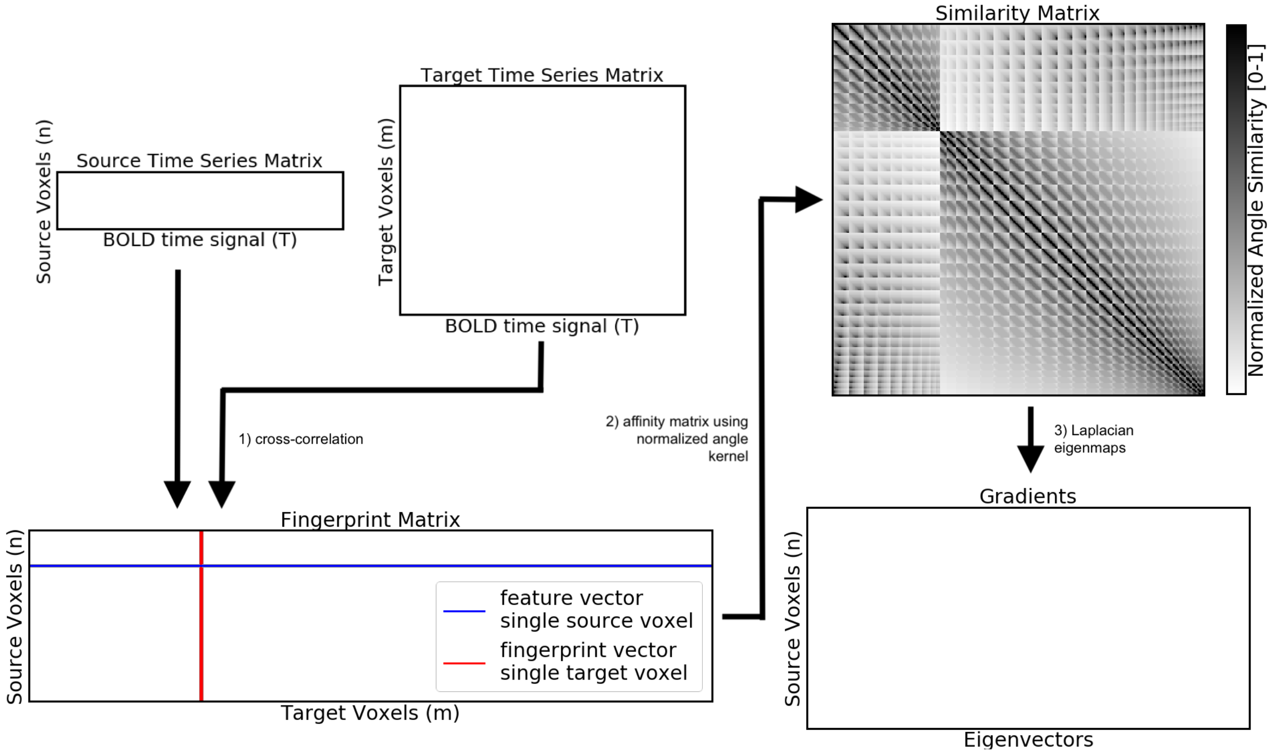


Figure 6.2: General workflow for computing cortical gradients from fMRI data. We modeled our workflow off of [3]. For a generic source-target pair, we computed the cross-correlation of their individual time-series matrices to generate a functional connectivity "fingerprint" matrix - each vertex in the source area is characterized by a feature vector, where each column in the matrix is a "fingerprint". We computed the pairwise affinity matrix between all feature vector pairs using the BrainSpace toolbox [17]. We applied a dimensionality reduction algorithm to compute low-dimensional embeddings of source-area functional connectivity. The target area (whether the rest of the cortex, or a single other cortical area, *induces* these low-dimensional embeddings on the source area – the embeddings capture large-scale patterns of spatial variation in the functional time-courses in the source area, with respect to a specific target.

In addition to the region-level matrices, we applied the same methodology as above to generate whole-brain, region-by-region correlation matrices. For each of the 154 subjects, we computed the mean time-series for each cortical region by averaging the time-series across all vertices in a cortical area – this generated a single time-series vector per region. We generated a region-by-region cross-correlation matrix by computing the Pearson correlation between the time-series for all region pairs. We computed the

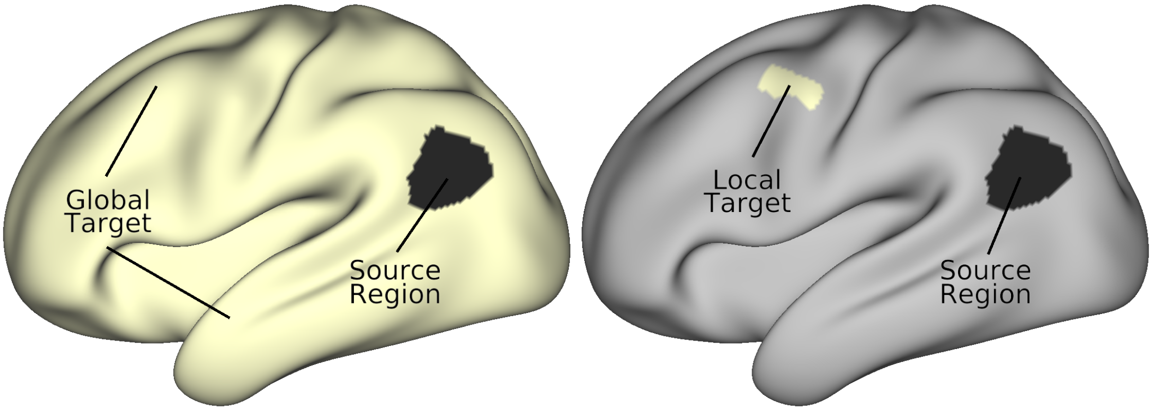


Figure 6.3: Example of a source region, with globally and locally-defined target areas. Here the source region is area PGi of the angular gyrus. On the left, the target area is defined as the rest of the brain, excluding PGi. On the right, the target is area 55b from the HCP-MMP parcellation.

Fisher R-to-Z transform to normalize the correlation values, computed the mean of all Z matrices across all subjects, and inverted the Fisher's Z-to-R transform the Z-values back to correlations. This generated a group-averaged region-by-region correlation matrix representing the similarity of BOLD time series between every pair of cortical areas.

6.4.2 Computing affinity matrices

We computed affinity matrices from the correlation matrices generated above. Without loss of generality, given a cross-correlation matrix C computed using the steps described above, we computed an affinity matrix, S, characterizing the pairwise similarities between all feature vectors C_v for $v = 1 \ldots n$ using the normalized angle kernel defined as

$$S_{\alpha,\beta} = 1 - cos^{-1}\left(\frac{C_\alpha C_\beta}{\|C_\alpha\| \, \|C_\beta\|}\right) \Big/ \pi \tag{6.2}$$

We chose this kernel because it bounds the similarity between 0 (entirely dissimilar) and 1 (entirely similar) [17]. Each index (i, j) in S describes the pairwise similarity between the correlation feature vectors C_i and C_j. The affinity matrix is a matrix induced on a source area, and captures the pairwise relationships between the functional connectivity patterns of pairs of locations in a source area. These functional connectivity patterns can describe the connectivity of the source area to the whole brain (global) or to a different cortical area (target-specific).

6.4.3 Computing cortical gradients

To compute cortical gradients, we can apply any number of dimensionality reduction algorithms. However, as was shown in [3], only non-linear reduction methods like diffusion embedding or Laplacian eigennmaps were able to recapitulate known retinotopic patterns of spatial organization in the primary visual area or somatotopic organization in M1. Here, we applied Laplacian eigenmaps to the affinity matrices using the BrainSpace Toolbox. For a similarity matrix S, we computed

$$D_{j,j} = \sum_{i=1}^{n} S_{j,i} \tag{6.3}$$

$$L = I - D^{-1/2} S D^{-1/2} \tag{6.4}$$

where D is the diagonal degree matrix of S, and L the symmetric, normalized graph Laplacian. We then computed the generalized eigendecomposition

$$L\Lambda = DX\Lambda \tag{6.5}$$

where X and Λ are the generalized eigenvectors and eigenvalues, respectively, of L. The eigendecomposition of the Laplacian matrix generalizes the notion of a Fourier decomposition to the graph domain – eigenvectors represent standing waveforms (analogous to standing sine and cosine waves), and the eigenvalues present spatial frequencies (analogous to frequency components). The spatial eigenvectors are not guaranteed to be sinusoidal in nature, but they do reflect increasingly higher-frequency spatial trends. We sorted the eigenvalues and corresponding vectors increasingly, such that $\lambda_1 < \lambda_2 ... < \lambda_k$. Each eigenvector $x_i \in X$ is a gradient of region r that represents a non-linear embedding from the space of high-dimensional connectivity fingerprints onto a one-dimensional subspace. Each eigenvector captures some unique spatially-varying properties of the connectivity profiles in a cortical area.

As a reminder, given an eigendecomposition of a symmetric matrix, the multiplicity of eigenvalue 0 is the number of connected components in the graph. The corresponding eigenvector for an eigenvalue of 0 represents the DC component of the graph Laplacian, and is constant over the graph domain. The first non-constant eigenvector, known as the Fiedler vector, attempts to assign constant embedding values to areas of similar connectivity profiles – i.e. it attempts to "flatten" the embedding.

We repeated above process for the

1. whole-brain correlation data to compute whole-brain gradients like those described in [18]
2. regional *global* data to generate global gradients for each source area
3. regional *target-specific* data to generate target-specific gradients for each source area

6.4.4 Characterizing functional connectivity variability

We were interested in analyzing the *complexity* of cortical regions. Here, we defined complexity as the number of parameters needed to capture most of the functional information in a cortical region. We related this notion of complexity to the number of informative connectivity gradients within a cortical area, and quantify complexity by examining how well global gradients explain variation in connectivity fingerprints. Recently, the authors in [117] showed that, using multidimensional scaling (MDS), 3 whole-brain cortical gradients were enough to explain 98% of the variance in whole-brain similarity matrices. Given these results, and in conjunction with work by [118, 119], we hypothesized that only a few gradients are actually informative – we proposed that variation in the connectivity fingerprints will be explained largely by a small subset of gradients, and thereby provide another measure of this region's intrinsic dimension.

Estimating functional connectivity profiles using cortical gradients as spatial basis functions

As mentioned previously, linear matrix decomposition methods, such as SVD, are unable to identify gradients that recapitulate known patterns of brain organization, such as somatotopy in M1/S1 and retinotopy in the visual cortex. In contrast, non-linear decomposition methods, like laplacian eigenmaps, are able to identify these axes of organization. However, whereas the eigenvalues of linear approaches can be naturally extended to examine the variance explained by each spatial basis function, non-linear approaches do not have such an extension. One reason for this is that when computing gradients using a generalized eigendecomposition, the eigenvectors are not-guaranteed to be orthogonal, and as such, do not account for independent sources of variation in the input signal.

It has previously been shown that whole-brain gradients computed using the whole-brain graph Laplacian can be treated as spatial basis functions (also known as "functional harmonics") of rs-fMRI [119] and task-fMRI data [118]. Using nested linear models of whole-brain gradients, [119] show that linear combinations of these functional harmonics can faithfully reconstruct known resting-state networks using only a small subset of harmonics. We hypothesized that we could use the nested linear model approach to explore the number of "informative" global gradients and thereby better-inform our estimates of intrinsic dimension of cortical regions from a regression perspective.

We treated the global gradients as independent variables, and the functional connectivity matrices as dependent variables. We examined how much spatial variation in connectivity fingerprints can be explained by the first k global gradients, and show that only a few gradients are needed to explain most of this variation.

For a matrix of cross-correlations $C \in \mathbb{R}^{n \times p}$, with p fingerprints, and matrix of gradients $X \in \mathbb{R}^{n \times K}$ with k gradients, we fit p nested linear models $C = X_{1:k}\beta$ as:

$$\begin{bmatrix} c_1 & c_2 & \dots & c_p \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \dots & x_{n,k} \end{bmatrix} \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_p \end{bmatrix} \tag{6.6}$$

where $\beta \in \mathbb{R}^{(k+1) \times p}$ is a matrix of coefficients. Each column vector $c_i \in \mathbb{R}^n$ is a spatial map induced on the source region, containing the Pearson correlation between all source area vertices with a single target vertex (row 2 in Figure 6.4). Likewise, each column vector $x_i \in \mathbb{R}^n$ is a spatial map induced on the source region representing the i-th cortical gradient (rows 1 in Figure 6.4). Each column vector $\beta_i \in \mathbb{R}^{k+1}$ represents the coefficients for a single model. We fit p independent models and compute the coefficient of determination, R^2, for each model. We computed the average $\bar{R}^2_k = \frac{1}{p}\sum_{i=1}^{p} R^2_{i,k}$ for $k = 1 : K$ and visualize the variance explained as a function of k.

We proposed to estimate the intrinsic dimension of a cortical region using the "elbow" method, by applying the Kneedle Algorithm to identify the point of maximum curvature in a monotonic function [120]. Here, we modeled functional connectivity with series of nested linear combinations of global gradients – for each model, we computed the variance explained, which by definition must be monotonically increasing. We identified the point of maximum curvature in this variance-explained curve as the intrinsic dimension of the cortical region.

Assessing gradient information content

In addition to better-understanding how well linear combinations of global gradients can explain variations in functional connectivity, we were interested in identifying the "dominant" gradient of a given area. As a reminder, gradients capture information about large-scale spatial variations in connectivity. However, we wish to emphasize that non-linear *axes along which connectivity profiles vary*, do not necessarily best explain the functional connectivity profiles. As such, we were interested in exploring which axis of spatial variation can best explain functional connectivity profiles.

Because of how the generalized eigendecomposition is computed on the graph Laplacian, there is no direct interpretation of the "importance" of an eigenvector. Contrast this with, for example, Principal Component Analysis (PCA), where the eigenvalues of a corresponding eigenvector directly relate to the amount of covariance explained by a principal component vector. The principal component with the largest eigenvalue explains the most covariance in the data. Likewise, when applying dimensionality reduction via diffusion embedding (DE), the eigenvalues of the Markov transition matrix correspond to the variance explained by each orthogonal component. The Laplacian matrix does not have a direct analog to either PCA or DE in terms of variance explained. Instead, the Laplacian eigenvalues represent frequency components of these standing eigenvector waveforms. As such, we were interested in quantifying the contribution of each eigenvector individually to the functional connectivity signal. More specifically, we wished to determine whether the first gradient always explains the most spatial variation in functional connectivity, or if higher-order gradients can better-explain functional connectivity. To this end, we regressed the fingerprint matrices C on each individual gradient X_i independently, and compare the variance explained by each single gradient.

6.4.5 Characterizing target-specific gradients

Thus far, we have described a methodology for characterizing the functional complexity of cortical areas using *global* gradients – i.e. those that describe the spatial organization of a cortical region with respect to its functional connectivity to the rest of the brain. However, we wondered whether global-scale spatial variation could be recapitulated on the local scale, or if these two properties diverge. That is, we wondered if the *global* axes of spatial variation in a region are the same as the *local* axes of spatial variation, *and* whether local axes of spatial variation were dependent on the network identify of the specific cortical target area.

Without loss of generality, for two arbitrary gradients g and h induced on a source region, we defined the similarity between this gradient pair as

$$sim(g,h) = \left| \frac{cov(r_g, r_h)}{\sigma_{r_g}, \sigma_{r_h}} \right| \tag{6.7}$$

the absolute value of the Spearman rank correlation, ρ, between the two induced gradients, where r is the rank-ordering of a given gradient, and σ_r is the rank-ordered standard deviation. We took the absolute value of ρ to factor in the arbitrariness of eigenvector sign.

For a given source region, for all target regions in the HCP MMP atlas, we computed the symmetric pairwise similarities $\in \mathbb{R}^{180+1 \times 180+1}$ of all induced gradients on the source area. This generates 180 symmetric matrices, one for each source area, where each index in the matrix contains the similarity between the gradient induced by target 1 and the gradient induced by target 2.

We hypothesized that the similarity of any two target-specific gradients will be dependent on the functional network to which both the source and the target region belong, and on the functional connectivity strength between these two regions. We addressed two questions:

1. For a given source area and its global gradient, is the similarity of the global gradient with a target-specific gradient induced on the source by a target, related to how strongly the source and target are connected?
2. Given two target-specific gradients induced by different targets on a source area, is the similarity between these two gradients related to how strongly the two targets are connected?

In both cases, we hypothesized that gradient similarity will be positively correlated with functional connectivity strength. However, we hypothesized that the strength of this relationship will be modulated by the functional networks of the source and target areas.

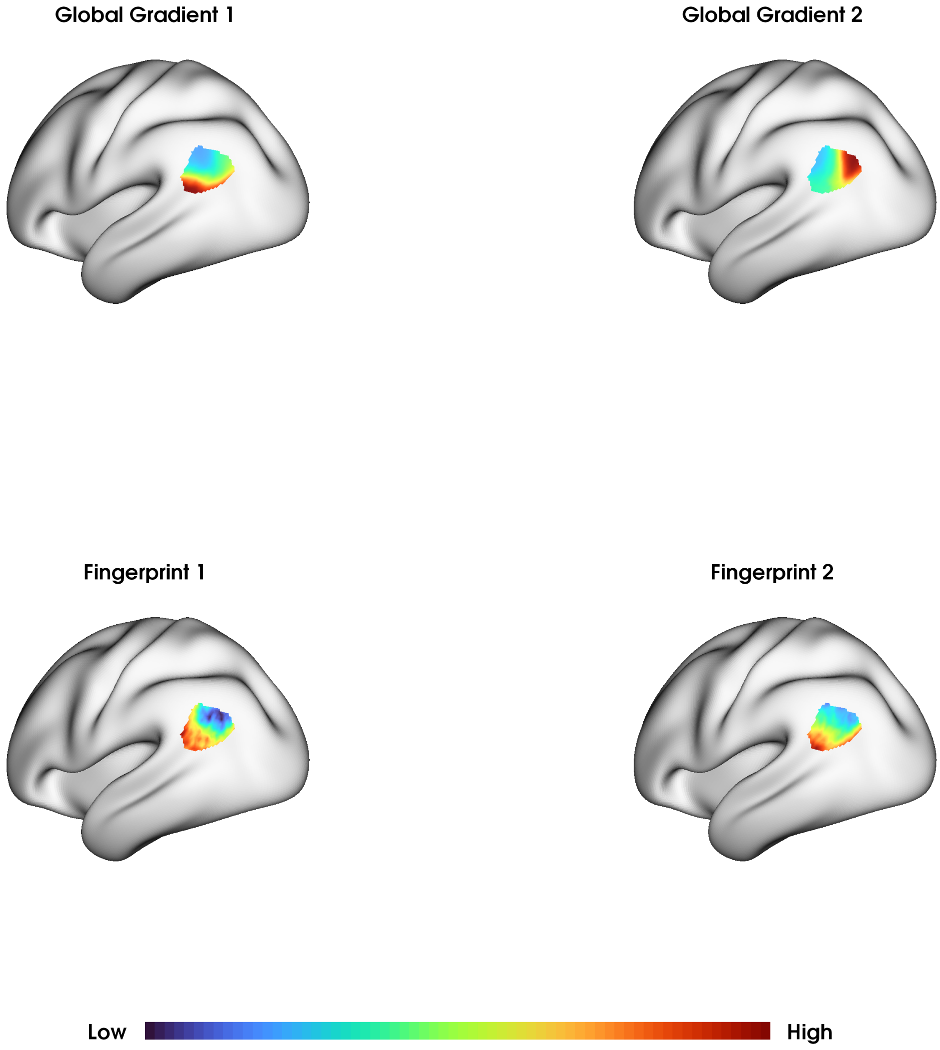


Figure 6.4: Examples of the first two global cortical gradient vectors (first column) for area PGi and two random fingerprint vectors (columns 2-5). The global gradients capture spatial trends in the functional connectivity profiles of PGi in relation to the rest of the brain. The fingerprint vectors represent the Pearson correlation of every vertex in PGi with a single vertex in the target area, which for global gradients, is the rest of the brain, excluding area PGi. We treated the fingerprint vectors as dependent variables, and regress them on the gradients, to analyze the variance explained by these gradients.

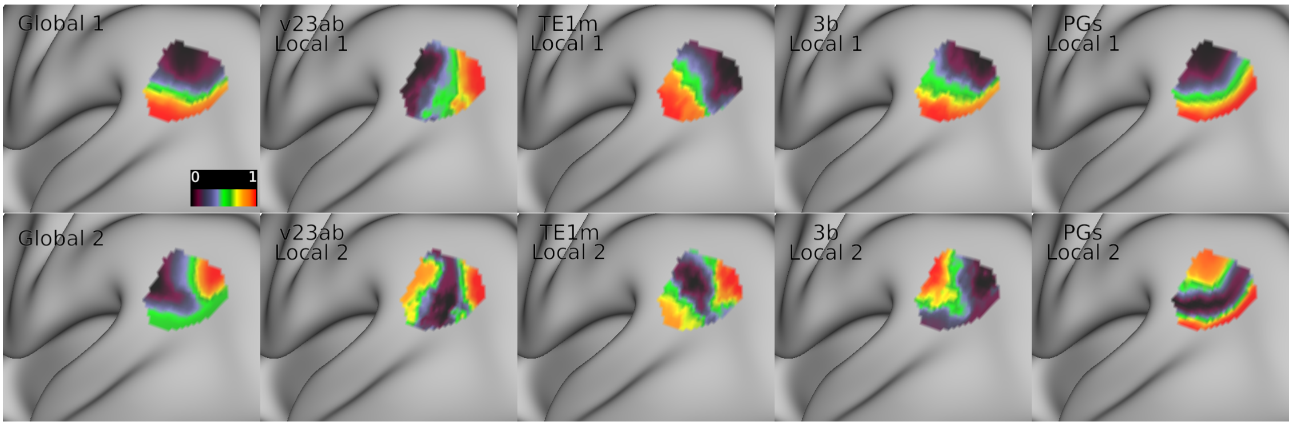


Figure 6.5: Examples of the first target-specific gradients of a source region (PGi) induced by four unique target areas (v23ab, TE1m, 3b, and PGs). Row 1 = gradient 1. Row 2 = gradient 2. Column 1 = global gradients. Columns 2-5: target-specific gradients. Interestingly, we found that, rather than the second target-specific gradients looking similar to the second global gradient, that they look more like higher-spatial frequency components of the first global gradient.

6.5 Results

6.5.1 Whole-brain gradients at the voxel and regional levels

For comparison purposes, we computed whole-brain gradients analogous to those computed by [18] and [104]. However, rather than computing voxel-level embeddings, we computed region-level embeddings, that are simply lower-resolution versions of the original work in [18]. The first gradient captures the transition in connectivity from the primary sensory cortices to the transmodal cortices, notably anchored at the Default Mode Network. The second gradient is primarily an axis characterizing the sensory areas, anchored at one end by the visual cortex and terminating at the sensorimotor cortex.

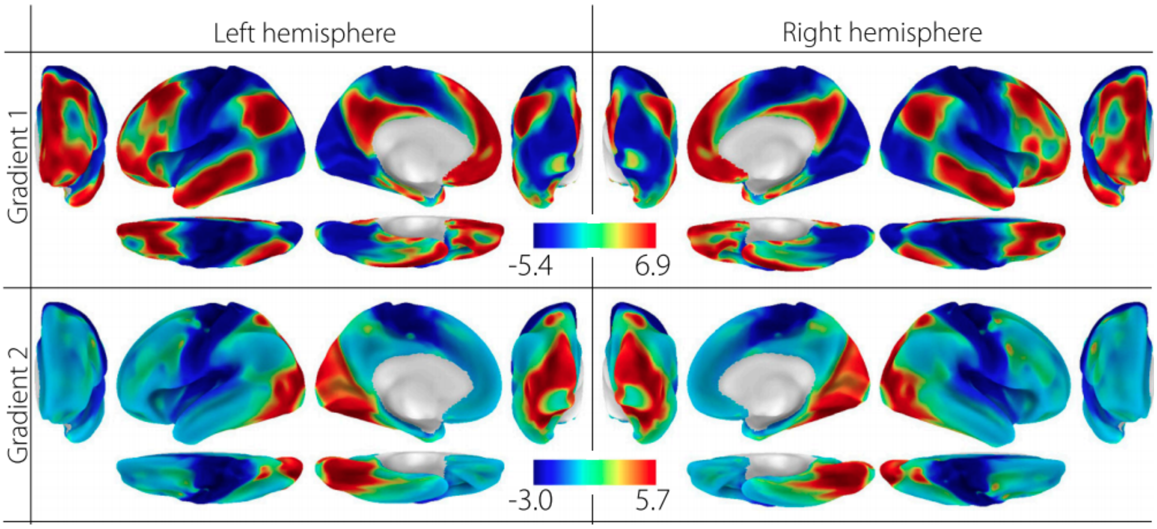


Figure 6.6: The first two whole-brain gradients computed from resting-state fMRI [18]. The first gradient captures a transition from primary sensory areas to transmodal cortices.

Clearly, the empirical region-level gradients computed on the HCP-MMP parcellation in Figure 6.7 mirror the voxel-level gradients in Figure 6.6. We related these region-level whole-brain embeddings to region-level statistics in downstream analyses.

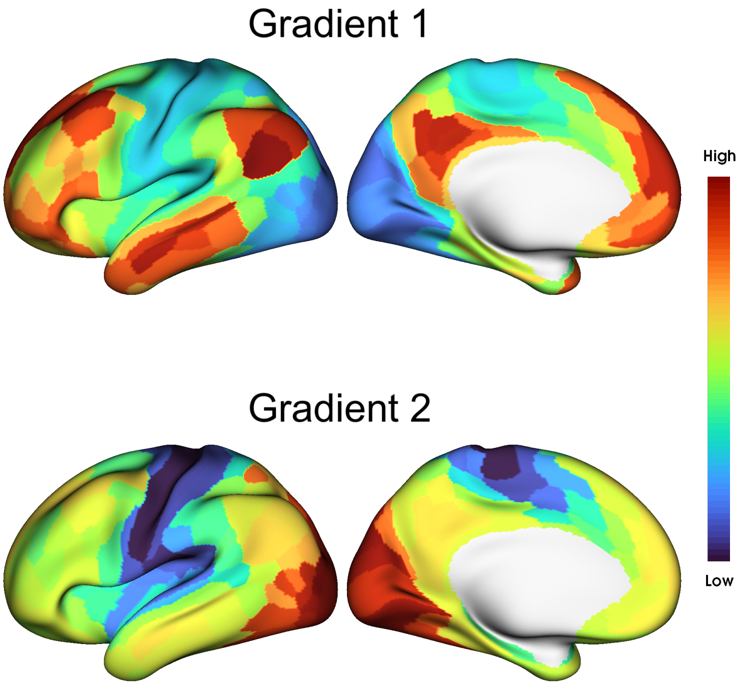


Figure 6.7: Empirical examples of the first two HCP-MMP region-level global gradients computed using resting-state fMRI. The first gradient captures a transition from primary sensory areas to transmodal cortices.

6.5.2 Regional connectivity profiles can be explained by three informative dimensions

We examined how well global gradients explain variation in functional connectivity. First, we'll discuss results on the group-averaged data and follow with examples for subject-level results. In almost all cases, we found that the first two global gradients of a region explain more than half of the variation in connectivity profiles; the first 3-5 gradients explain nearly 80% of the variation (Figure 6.8). Similarly, after averaging curves across all regions within a functional network (i.e. each subplot of Figure 6.8), the average variance explained curves (indicated by the dashed black line) are similar across networks. These results are in line with what [118] and [119] found in their whole-brain functional harmonic analyses – namely that only a few lower-order gradients are needed to explain the majority of variation in resting-state and task-based fMRI signals. Likewise, recent work by [117] estimated that resting-state functional connectivity can be decomposed into three statistically-dependent dimensions using a multidimensional scaling (MDS) approach. The authors identify three primary axes, though only the first axis recapitulates the axes from recent resting-state fMRI and microstructural analyses. As expected, when extending this same analysis to single-subject data, we found that subject-level gradients explain less variation in the connectivity relative to the group-averaged data (Figure 9.5). This is likely due to higher variability from subject to subject – the group-averaged correlation matrices will tend to smooth out variability in functional connectivity, resulting in smoother gradient profiles less likely to fit noise.

For comparison purposes, we also showed the variation explained by the singular vectors of the correlation profiles using the same linear model framework. Analogous to the Laplacian eigenvectors, we could have used the singular value decomposition (SVD) to compute low-dimension embeddings of the functional connectivity profile of a region. In contrast to the Laplacian eigendecomposition approach, SVD is a *linear* decomposition algorithm that returns the best rank-K approximation of a matrix, in terms of the Frobenius norm. Because of this, we expected that the singular vectors would explain more variance in functional connectivity than the global Laplacian embeddings. We were interested in how much more variance could be explained, and whether the estimated intrinsic dimension was dependent on decomposition algorithm choice. Figure 9.4 shows an analogous plot to Figure 6.8 – however, we found that the first k singular vectors explained more variation in functional connectivity than the first k Laplacian embeddings.

For each matrix decomposition approach (Laplacian eigenmaps and SVD) We computed the intrinsic dimension of each region from the variance explained curves. The mean intrinsic dimension estimate computed using the Laplacian gradients was 2.8, and 2.6 using the singular vectors. We compared the paired intrinsic dimension estimates for

Dominance Proportions***

Gradient 1	Gradient 2	Gradient 3
75.5%	21.1%	2.8%

Table 6.1: Estimates of the proportion of times a given gradient explains the most variation in functional connectivity, averaged across all cortical areas in the HCP-MMP atlas. ***In one case, for cortical area **OP2-3** in the posterior opercular cortex, the 6th gradient was dominant – we excluded that region here.

each region using the Wilcoxon Rank-Sum test (W=2326, p=0.042, r=0.14). The two approaches showed a significant difference, though the effect size ($r = 0.14$) was small. Our results support the notion of 3 gradients, with some variability from one region to the next.

We next sought to determine the importance of each global gradient. Based on the intrinsic dimension estimates, we considered only the first three gradients as candidates. Because there is no direct interpretation of Laplacian eigenvalues as explaining some form of variation in the affinity matrix, we treated the variance explained in the functional connectivity by each gradient independently as a proxy for its information content. We refer to the gradient explaining the *most* variation as the "dominant" gradient. As shown in Figure 6.9, the first gradient is dominant in 75% of cortical areas. However, this is not always the case – the second gradient is dominant in 21% of cases, and the third in 3.8% of cases. From a spatial frequency perspective, the first gradient represents low-frequency transitions in connectivity profiles, and as such captures large-scale, slow variations in the spatial signal (see Figure 9.6. One cortical region, area **OP2-3** in the posterior opercular cortex, did not follow suit – the sixth gradient was most dominant here, explaining 15.6% of the variation in its functional connectivity.

We binned cortical areas based on which gradient was dominant and examine how much variance is explained by the first, second and third gradient independently (Figure 6.10). By definition, for the k-th bin, the k-gradient will explain the most variance. In cortical areas where the first gradient is dominant, it explains on average 40% of the variation in functional connectivity. For the second and third bins, the corresponding dominant gradient explains increasingly less variation (28% and 24%, respectively). We interpreted this to mean that, for those areas where the first gradient is not dominant, the principle global gradients are less reflective of the functional connectivity profiles in that area, possibly due to higher variability in the measured BOLD signal.

In Figure 6.11, we examined the variance explained by the dominant gradient for each region, and plot this variance against the whole-brain gradient embedding value (e.g. analogous to the [18] gradients).

6.5.3 Functional network assignment determines intrinsic spatial organization

We next examined large-scale similarities of target-specific gradients. As a reminder, we used the functional assignments of the HCP MMP labels generated by [16] in Figure 6.1 – we related our results to these functional network assignments.

In order to glean insight into the spatial axes of functional variation within regions, we examined how gradient similarities relate to the functional networks of the source and target areas (Figure 6.12). We emphasize again that a gradient is induced on the domain of a source region *by another region* (either the rest of the cortex, or a single region), This gradient captures information about the spatial organization of the source, with respect to this other region. Because the first gradient was dominant in 75% of cortical regions, we restricted our analysis to this embedding for the remainder of this work.

For each source, for all target region pairs, we computed how similar the induced embeddings on the source are, by these two targets. We subseted the data into 9 unique groups, based on the 9 functional networks in [16] and show our results as symmetric similarity matrices. Each similarity matrix in Figure 6.12 aggregates data for across source regions in the same functional network. Each index in a similarity matrix contains the mean similarity between gradients induced by targets from two functional networks. For example, the plot in the first row and first column corresponds to results for source areas in the Auditory Network. Index $(1, 1)$ in this matrix shows the similarity of gradients induced by an Auditory-Auditory target pair, whereas index $(1, 2)$ shows the similarity of gradients for an Auditory-Somatomotor target pair.

For source regions belonging to the sensory networks (Figure 6.12, row 1), we found that induced gradients are generally more disassociated than source areas in the associative, control and transmodal source regions. We saw this more clearly in Figure 9.9 – overall mean similarity of gradients in the sensory areas is lower than in the other networks. For all networks, we identified a strong diagonal effect – when a target area belongs to the same network as the source area, the induced target-specific gradient tends to be more similar to each other (and to the global gradient (last row / column)).

In the case where the source and targets all belong to any combination of the attention (DAN, VAN), control (FPN, CIN), or Default Mode networks, the induced gradients on the source region are more similar to one another than otherwise. For example, given a source region in the Ventral Attention network, targets in the Cingulo-Opercular, Dorsal Attention, Fronto-Parietal, and Default Mode Networks tend to induce similar gradients on this source area. We identified a strong "block-diagonal" signal that differentiates more integrative areas from the primary sensory areas. This indicates that the spatial organization patterns induced by and on areas in these higher-order networks are highly consistent. We found that this conclusion does not apply to the sensory cortices.

In Figure 9.8, we aggregated the data from the similarity matrices in Figure 6.12. We split the similarities by whether a pair of targets areas are in the same or different network from the source region. In all cases, mean similarity is greater in the case where a source and pair of target areas all belong to the same network, as compared to when they belong to different networks. As identified in Figure 6.12, we identified network-specific characteristics: gradients induced on the primary sensory areas are on average more disassociated, while those induced on areas in the attention (DAN, VAN, CIN), control (FPN) and transmodal (DMN) networks are more similar. The Dorsal / Ventral Attention and Fronto-Parietal networks, specifically, show high similarity.

In aggregate, we showed that there is a strong network effect on the induced gradient similarities, that depends more on whether the source and targets all belong to the same network, as compared to when the networks are different. When all three areas belong to the same functional network, the target-specific gradients are highly similar to one another with a mean Spearman $\rho = 0.48$ for the first gradient, versus Spearman $\rho = 0.40$ when the source network is different. This similarity declines with gradient order – the second ($\rho = 0.26$ vs. $\rho = 0.21$) and third ($\rho = 0.19$ vs. $\rho = 0.17$) gradients become more disassociated.

6.5.4 Gradient similarity is modulated by functional connectivity strength

The results above show that cortical areas from the same network induce similar patterns of spatial organization on source regions (as exemplified by the diagonal signals in subplots shown in Figure 6.12). We asked if this observation was related to the strength of functional connectivity between regions (Figure 9.7). We hypothesized that, if two target areas were more strongly connected to a source region, that these two targets would induce more similar target-specific gradients to the global gradients, and to one another.

We computed the similarity of the global gradient with the target-specific gradients induced on the source, and regressed these similarities on the strength of functional connectivity between source and target regions. We found a significant positive relationship between gradient similarity and functional connectivity strength, (R^2: 0.026, p: 0.00) – however, the variance explained was small (Figure 6.13). We repeated this analysis after aggregating gradient similarity and functional connectivity strength within- and between-functional networks. At the aggregated level, functional connectivity explained more variance in functional connectivity (R^2: 0.42, p: 0.04), suggesting that the apparent connectivity strength effect is due to the fact that strongly connected parcels are more likely to belong to the same functional network (Figure 6.14).

While the above approach could plausibly be useful for identifying global-level principles of organization between gradients and functional connectivity, it is most likely the case that the relationship between gradient similarity and functional connectivity is region-specific. By averaging across similarity matrices, we risk losing this variability. For each source area, we regressed the pairwise similarity of target-specific gradients of the source area on the functional connectivity between target pairs – i.e. we fit 180 models, one for each source area – and examined the model fit and strength of association for each region. Here, all models were significant ($p < 0.05$).

As expected, we found a significant regional effect on the relationship between gradient similarity and functional connectivity strength (Figure 6.15). Of specific note was the improved model fit in the sensory areas, especially in the somatomotor areas (where functional connectivity explained upwards of 15% in gradient similarity) and the relatively weaker fits in the more integrative networks (Default Node Network), where variance explained is close to 0% (Figure 6.16). We identified what seems to be a graded transition in the relationship between gradient similarity and functional connectivity that spans the same axis as the whole-brain functional gradients, anchored at one end by the primary sensory areas, and at the other by the transmodal regions – and in fact, regional model fit is significantly and positively correlated with the whole-brain functional connectivity gradient in Figure 6.7 (Pearson r = 0.548, $p =< 0.0001$).

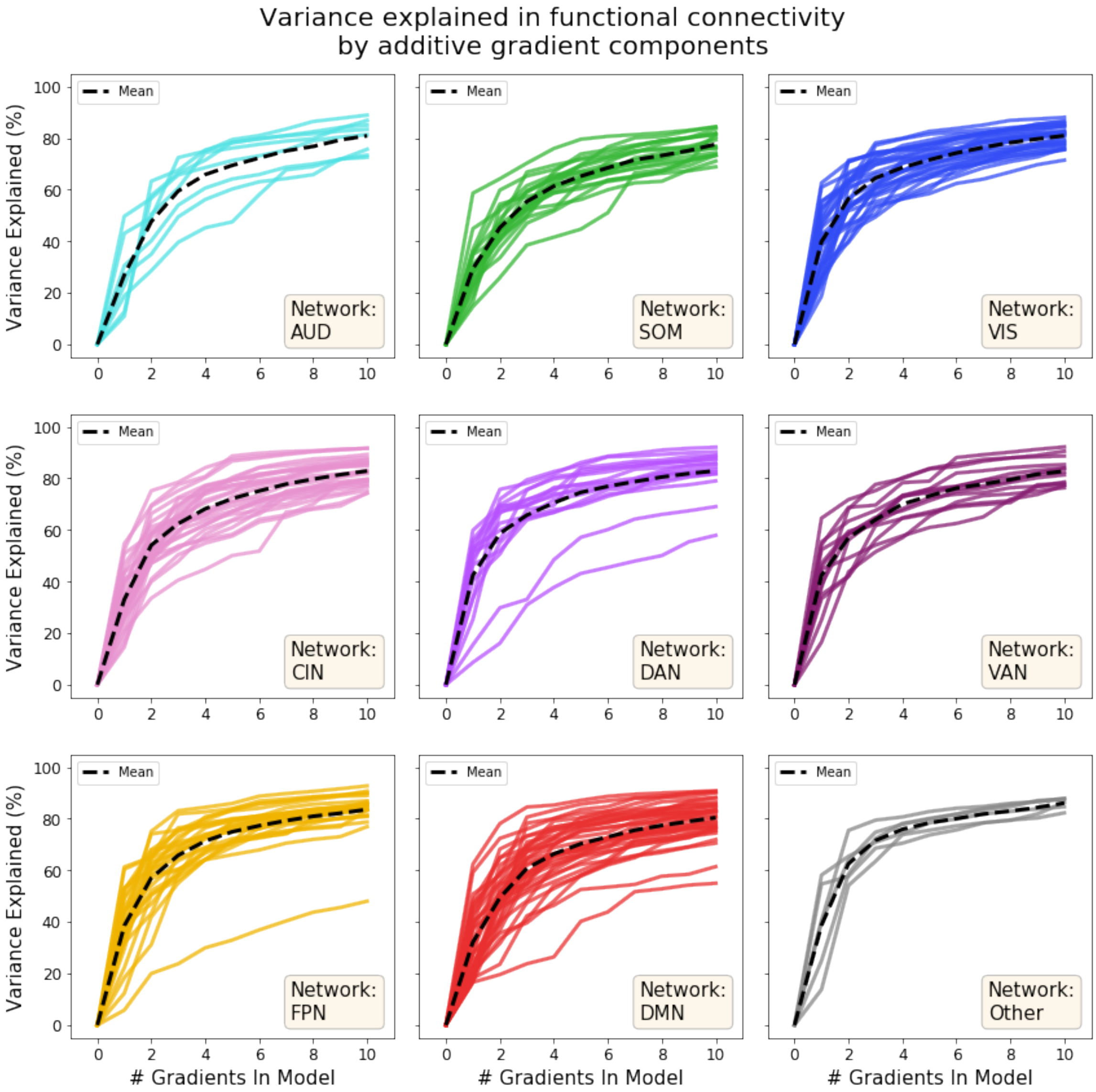


Figure 6.8: Variance explained by the first k global gradient components in group-averaged connectivity profiles. Each individual plot in the 3-by-3 grid shows the curves for all areas in a single functional network. Each curve corresponds to the variation explained by a single cortical area. The dashed black line represents the mean of all curves for a single functional network. Data points where the number of gradients is equal to 0 represent the variance explained by the intercept vector.

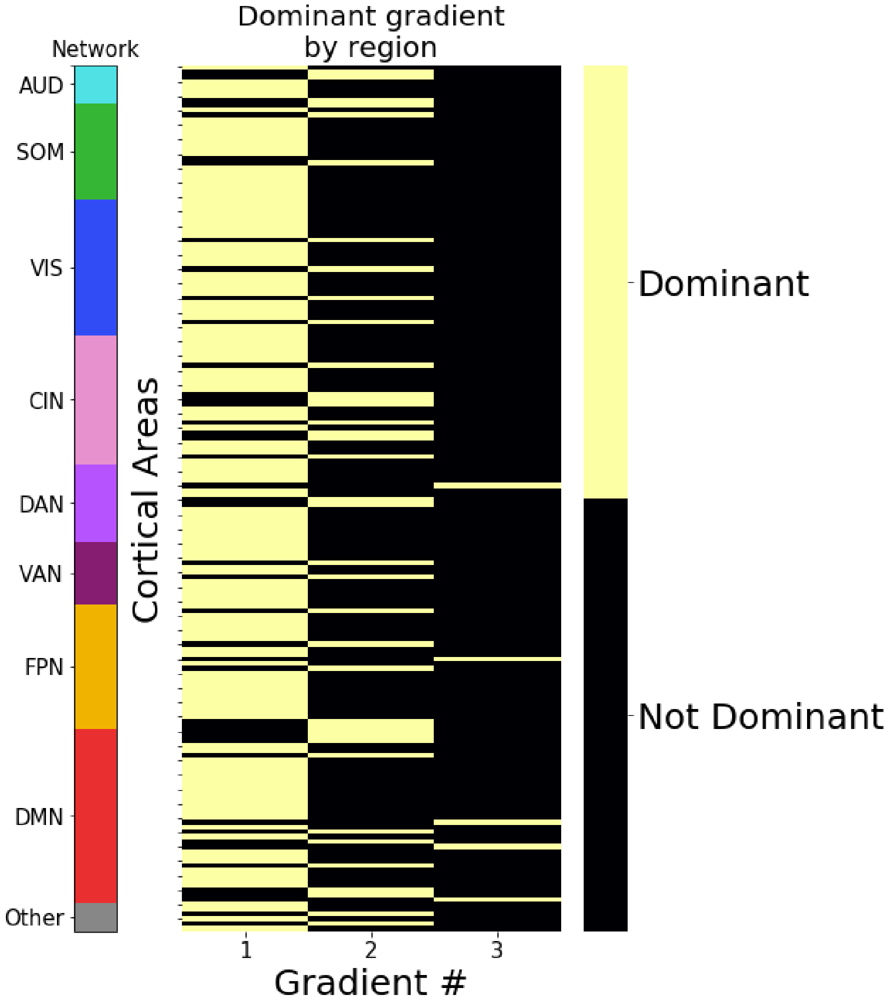


Figure 6.9: Dominant global gradients for each cortical area. The dominant gradient is the gradient that maximizes the explained variance in functional connectivity independently. Each column corresponds to a single gradient, and has a binary indicator value of 0 (not dominant) or 1 (dominant). Each row sums to 1 (i.e. only one maximum). Cortical areas are sorted by their functional network assignment.

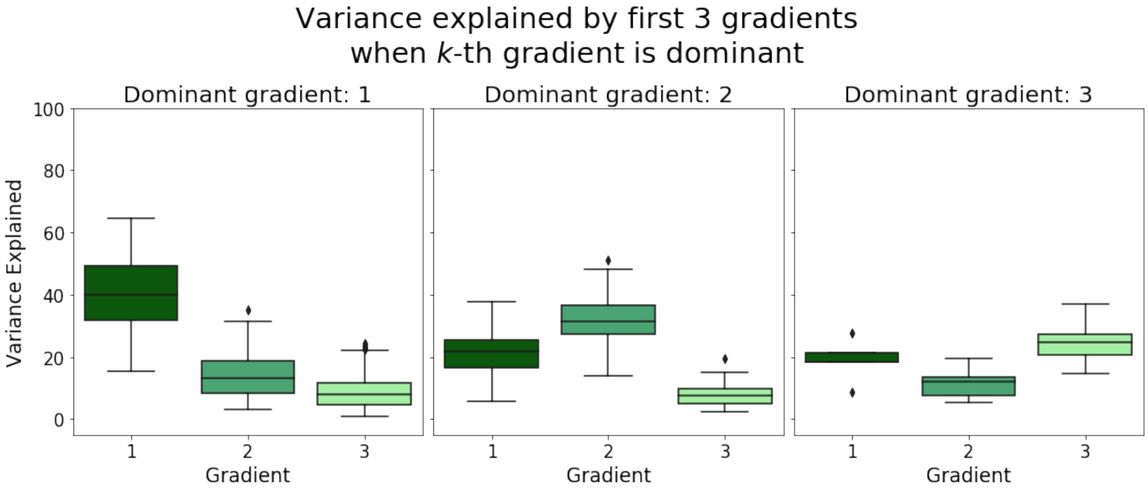
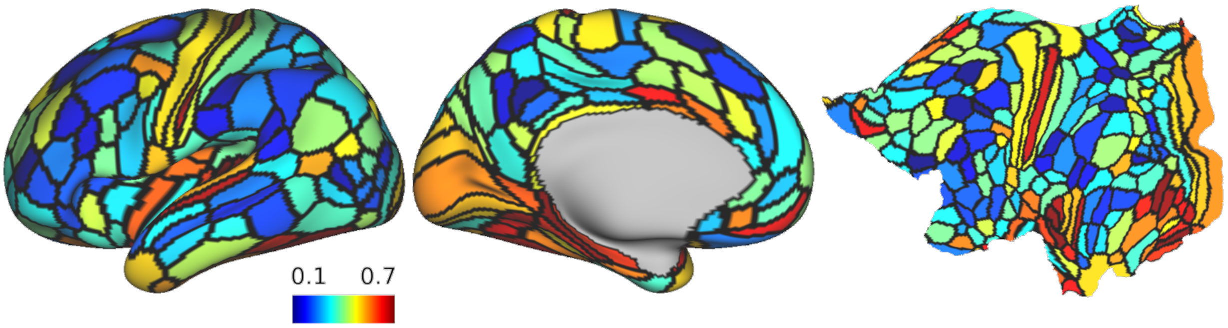
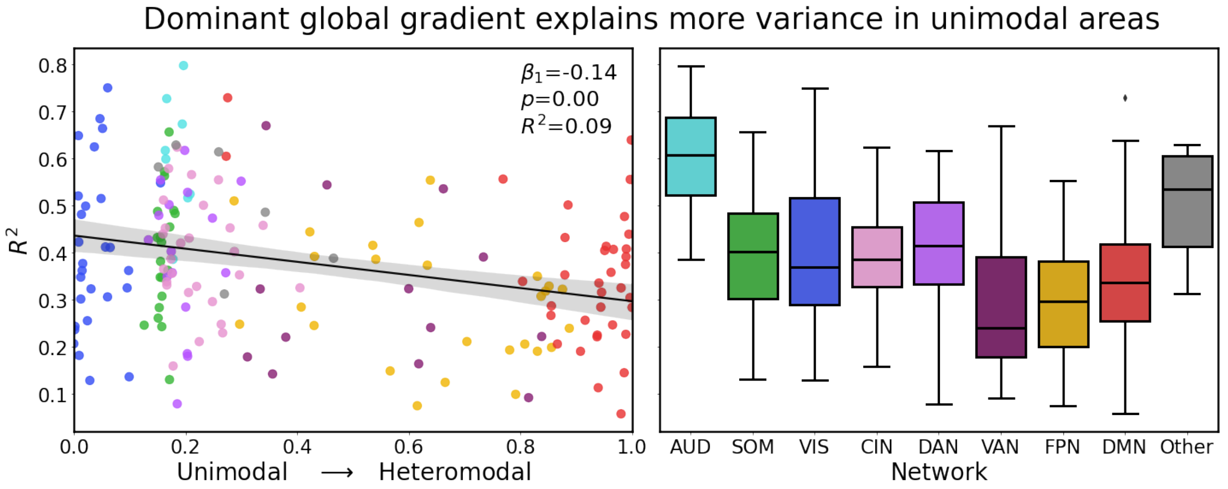


Figure 6.10: Variance explained by the k-th dominant gradient. Each subplot shows the variance explained by the 1st, 2nd, and 3rd gradients, in cases where the k-gradient is dominant. Each plot shows data for a subset of cortical areas



(a) Variance explained (R^2) by the dominant global gradient (first Laplacian eigenmap) for each cortical area, projected onto the cortical surface.



(b) Variance explained (R^2) by the dominant global gradient, plotted against the whole-brain cortical gradient embedding value, and colored by functional network assignments in the [16] atlas.

Figure 6.11: Variance explained by the dominant global gradient for each cortical area related to the whole-brain cortical gradient embeddings.

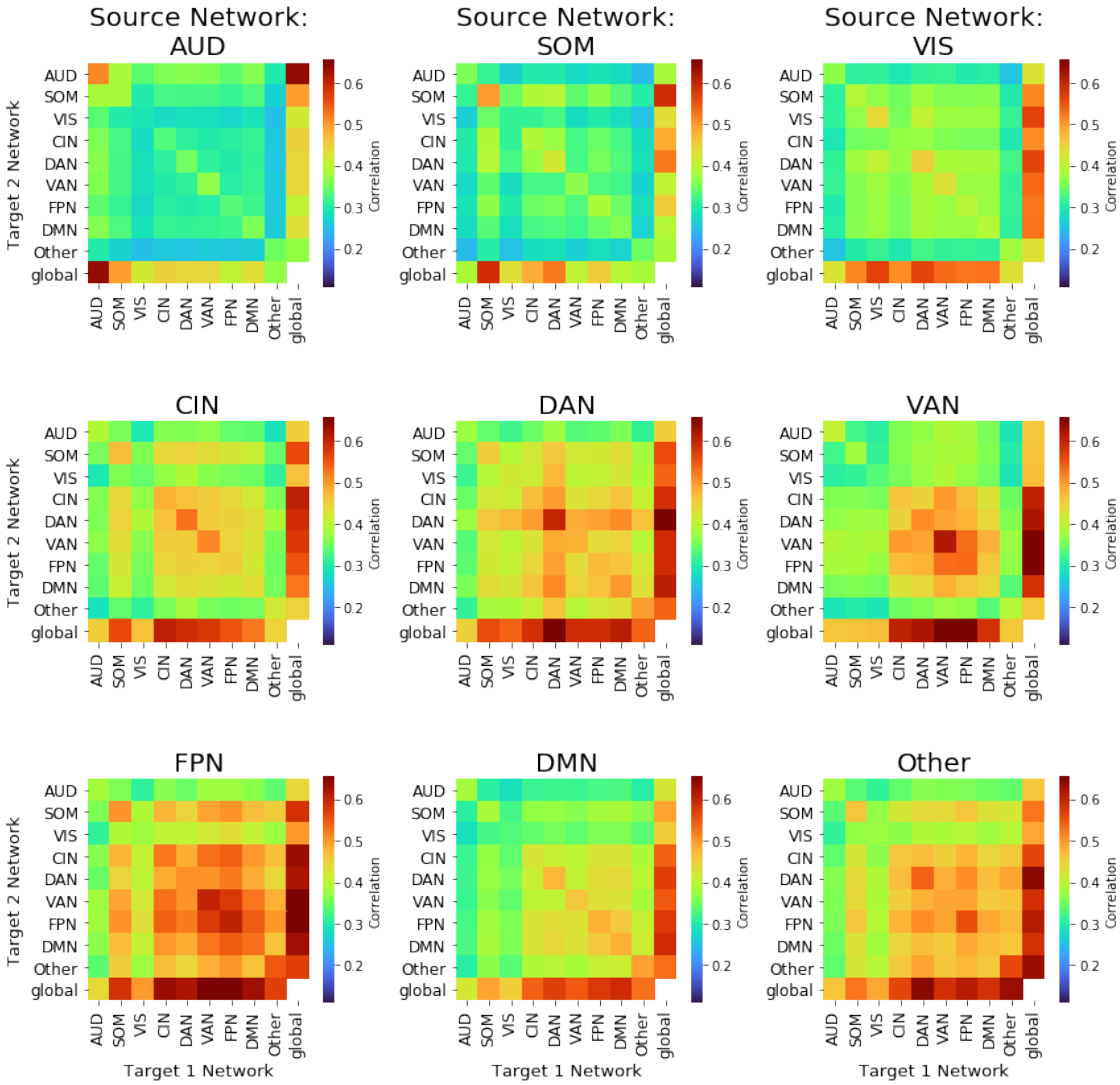


Figure 6.12: Similarity of induced gradients, delineated by the functional networks of the source and target areas.

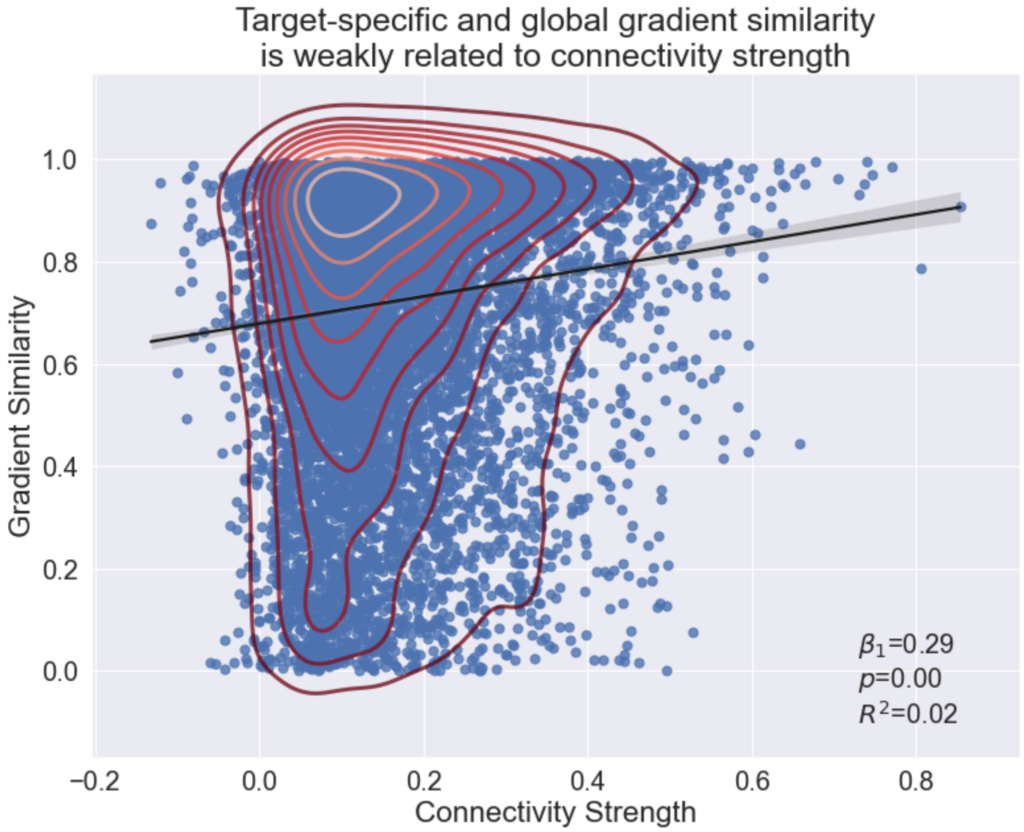


Figure 6.13: Regression of similarities between target-specific and global gradients, on the functional connectivity strength between the target and source area.

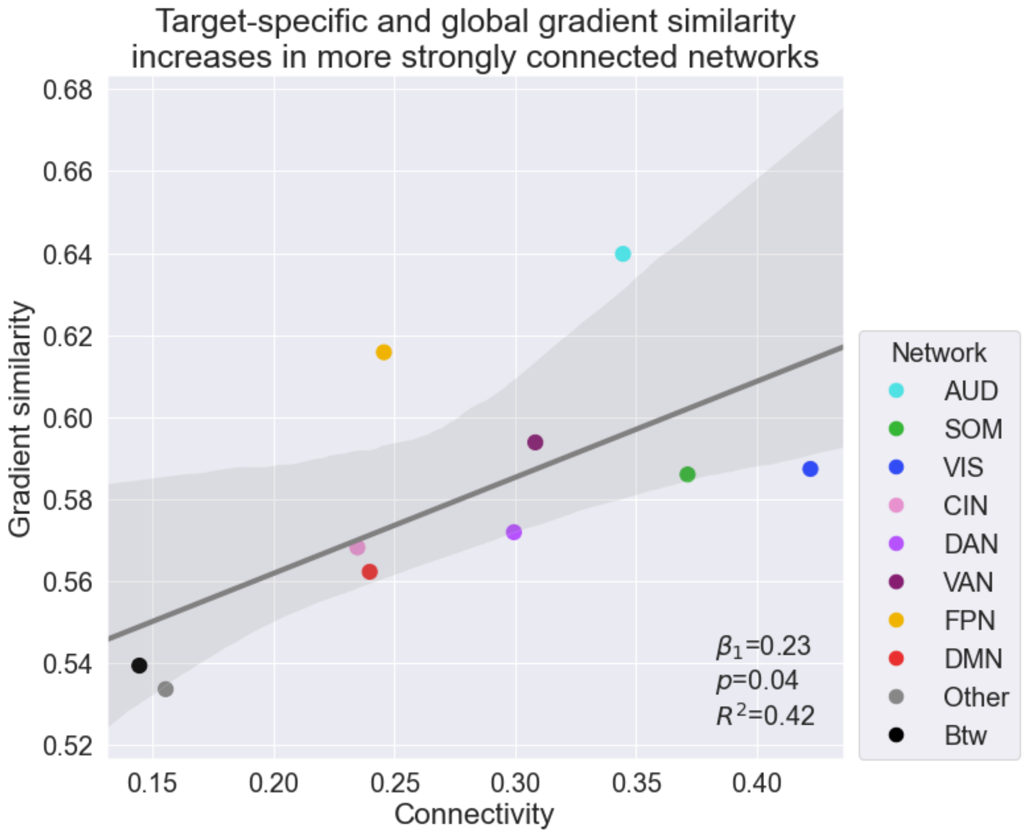


Figure 6.14: Regression of similarities between target-specific and global gradients, aggregated by within functional network (colored dots) and between functional networks (black dot). Colors refer to the network assignments of the Demirtas atlas [16].

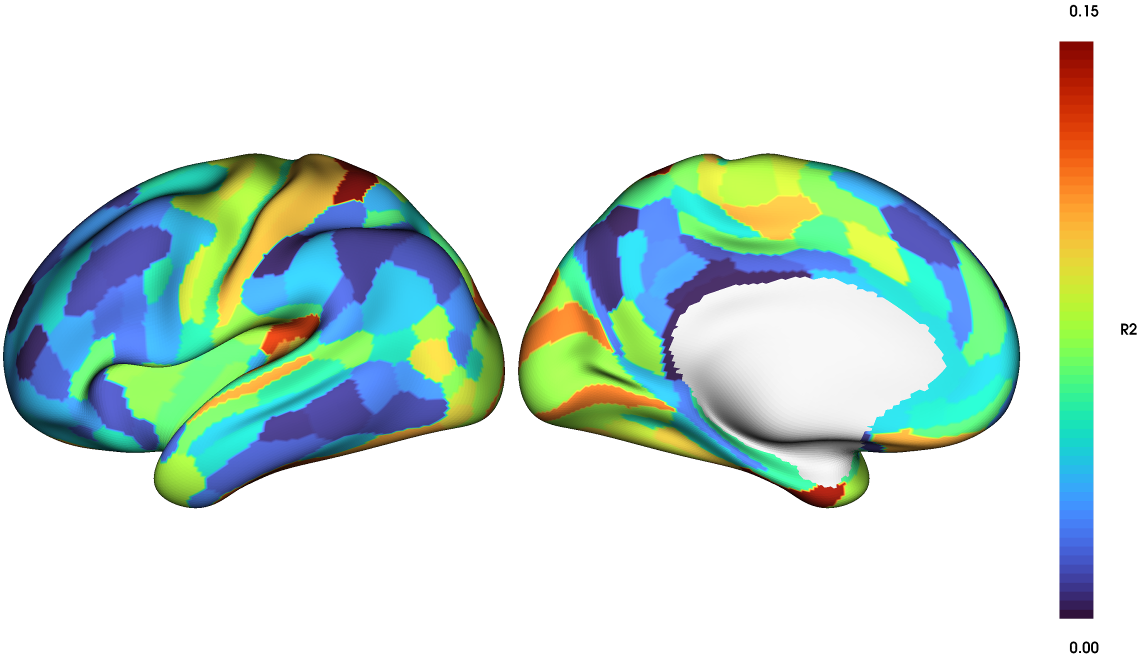


Figure 6.15: Variance explained (R^2) between gradient similarity and functional connectivity between target-specific gradients. We considered each source area independently, and mapped the model statistics onto the surface.

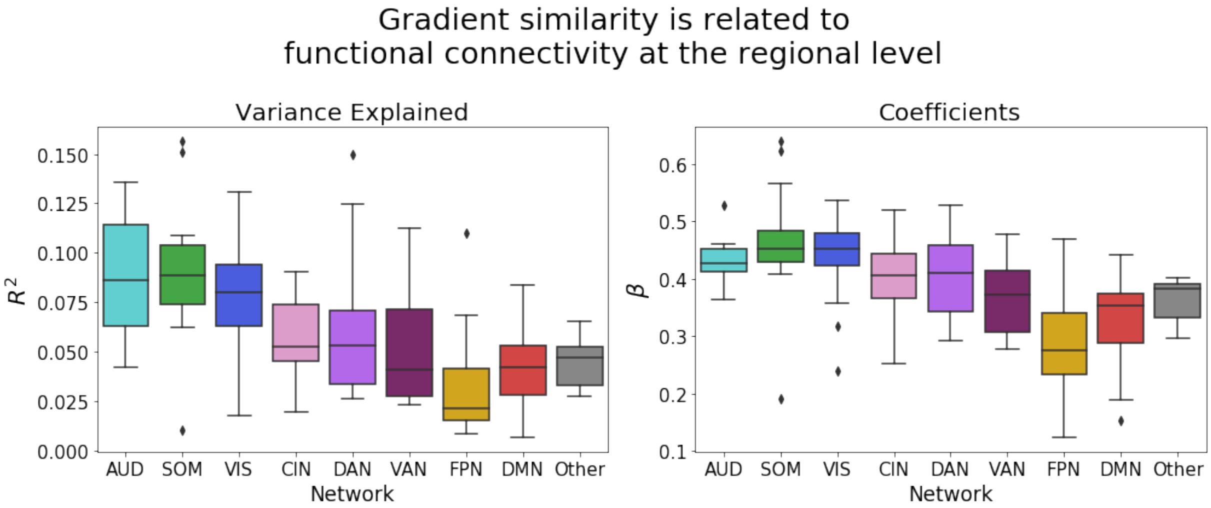


Figure 6.16: Regression of similarities between target-specific gradients on the functional connectivity strength between the two target areas. Statistics are aggregated across functional network.

6.6 Discussion

We developed a general framework to compute functional connectivity gradients for individual cortical regions that describe the intrinsic axes of spatial variation of connectivity profiles within a cortical area. We made the distinction between *global* gradients that capture the axes of variation within a region with respect to its connectivity to the whole brain, and *target-specific* specific gradients that capture the axes of variation within a region with respect to a single target. Our goal to was to compare these axes of variation, and assess the dependency of spatial organization on the specific target areas and functional network membership. Because these gradients capture the spatial variability in functional connectivity, we sought to examine how well any given axis could explain the functional connectivity profile vectors, with the goal being to deduce the importance, or contribution of, any given axis of variation. Likewise, in an effort to deduce the drivers of spatial organization of cortical areas, we also related the similarity of cortical gradients to the functional connectivity strength between pairs of regions.

Before we discuss the results, it is important to add a bit of exposition to the conceptualization of cortical gradients. Within a cortical area, at any given location, we can sample some multivariate biological information, such as the resting-state BOLD signal, microstructural properties like cytoarchitecture or soma size, or gene expression patterns. While we can easily examine a single variable within a cortical area simply by projecting it onto the cortical manifold, these multivariate features often display some hidden spatial gradations, which can be identified using dimensionality reductions, that reduce these high-dimensional features into a few meaningful embedding dimensions (gradients). These gradients capture global transitions in the multivariate features – often these are the directions along which the features show the most variation (in the case of singular vectors), or perhaps in which information flows most quickly (in the case of diffusion embedding vectors). Any two gradients can be compared to examine how closely these axes of spatial variation relate to one another (for example, are the axes parallel or orthogonal?) One of the goals in this analysis was to determine whether the axes of spatial variation within a given source region consistently varied in the same directions, or whether these axes were dependent on the area to which that source region was connected. If these axes show target-dependence, then we conclude that there are simultaneously spatially-overlapping patterns of organization within a single area. However, if the axes are consistent and do not vary from target to target, we interpret this as meaning that the source region is relatively uniform in its organizational pattern.

As has been well-studied, sensory information becomes more integrated as the synaptic level increases towards distal cortices, relative to the primary sensory organs. Inherent in this flow of information is a tendency towards more distributed communication between cortical regions – primary sensory organs, like V1 and A1, show few synaptic connections with other cortical area, while more integrative regions, such as those related to the executive networks, show more interconnected synaptic profiles, indicative of their central role in information processing and parallelized communication [20].

The idea of an integrative cortical "core" has been shown using both functional and diffusion connectivity analyses [121, 122, 123]. These core regions are shown to modulate and integrate the connectivity across networks and act as communication "hubs" through which much communication flows [124]. While many studies assign each cortical area to an individual functional network, other evidence shows that some regions can exist in multiple networks simultaneously, acting as cross-network links. For example, in [19], the authors discuss a "multiple-demand" (MD) network, primarily composed of regions in frontal and parietal cortices in the Fronto-Parietal, Attention, and Default Mode networks (Figure 6.17). These MD areas are co-active across a broad range of tasks, including working memory, math/language, and relational reasoning, and show preferentially stronger intra-MD functional connectivity. However, within this multiple-demand network, individual regions still show preference towards specific tasks. This indicates that while brain communication is transformed and rerouted by this functional core, there exist communities of regions within the core that are distinguished by their communication outside the core.

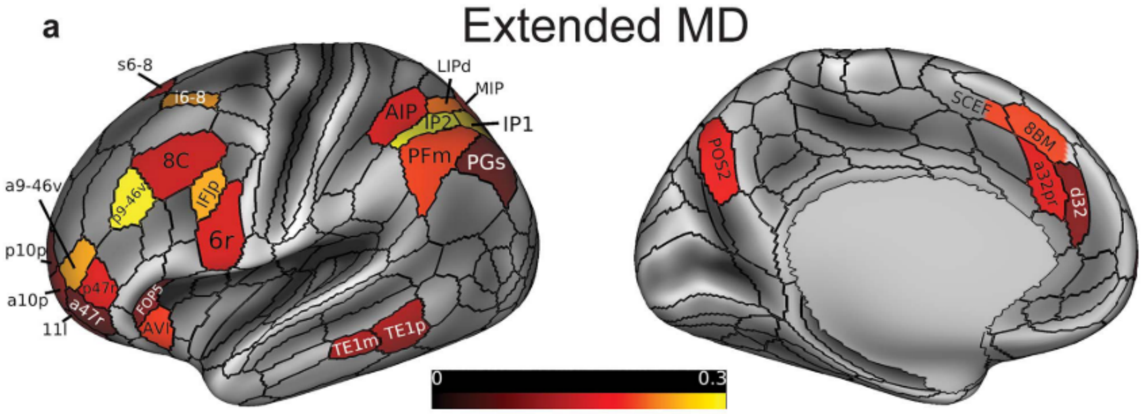


Figure 6.17: Multiple demand (MD) network, from [19]. The MD network areas are a core group of task-agnostic cortical regions from multiple control and attention networks that show preferential connectivity with one another and are active across most general tasks.

In our analysis, we compared the similarity of functional connectivity gradients in connectivity profiles across the cortex, and related these gradients to individual target areas. Our results show that gradients are significantly impacted by the functional network of the region of interest, and on the target area inducing this gradient (Figure 6.12). Of particular note is the fact that cortical areas corresponding to this multiple-demand network tend to co-induce similar gradients on cortical regions. Given a source area in the Fronto-Parietal, Dorsal / Ventral Attention, Cingulo-Opercular, or Default Mode network, and two targets also in any of these networks, the induced gradients are on average, highly similar to one another as compared to the somatosensory cortices (Figure 9.8). This implies that areas in this multiple demand network induce spatial axes of variation that co-vary strongly, and that the functional organization for an area in this integrative core, relative to other areas in the core, is consistent.

In an effort to determine if we could identify the main drivers of spatial organization of a cortical area, we related the similarity of the spatial axes to the strength of connectivity

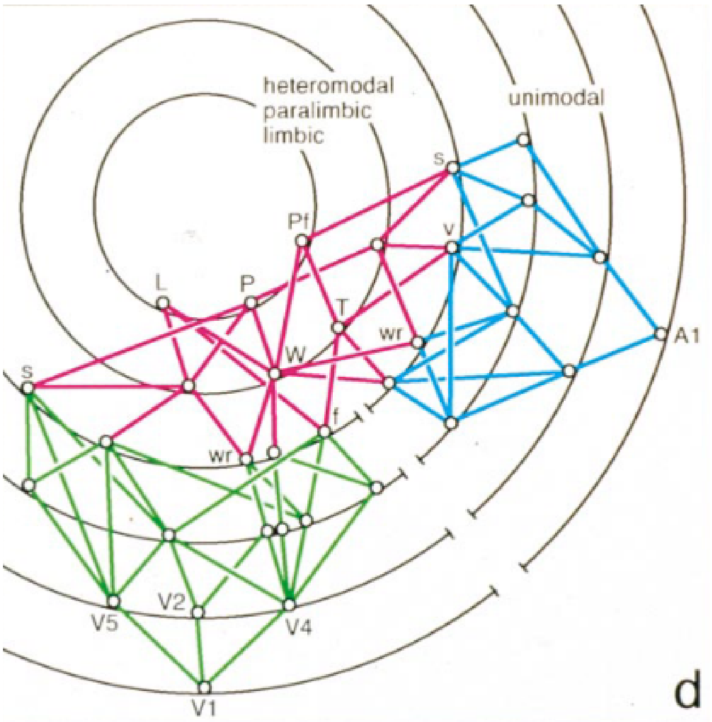


Figure 6.18: Hierarchy of information flow in the visual and auditory cortices, beginning at the primary sensory cortices and convalescing at more transmodal regions. Taken from [20]: "Each concentric ring represents a different synaptic level. Any two consecutive levels are separated by at least one unit of synaptic distance. Level 1 is occupied by the primary sensory cortex. Small empty circles represent macroscopic cortical areas or 'nodes', one to several centimeters in diameter. Nodes at the same synaptic level are reciprocally interconnected by the black arcs of the concentric rings. Colored lines represent reciprocal mono-synaptic connections from one synaptic level to another."

between cortical areas. We found that when averaging similarities of target-specific gradients across source regions, the relationship between similarity and functional connectivity was weak, and largely uninformative. However, this is likely due to the fact that averaging similarity matrices sacrifices the local signal. We examined source regions independently, and regressed similarity on functional connectivity for each source region. We found that, for cortical areas in the sensory cortices, gradient similarity could in large part be explained by functional connectivity strength, with R^2 values upwards 15%. In the more integrative cortices, however, this relationship was weakened, with less variation explained in the Fronto-Parietal and Default Mode networks, specifically. Interestingly, we saw that the variance explained, and the strength of relationship, varied smoothly along an axis that mirrored the whole-brain gradient axes computed by [18, 104].

Though seemingly contradictory with the results from Figure 6.12 and Figure 9.8, this can be explained by the fact that, for source regions in the Frontal-Parietal and Default Mode Network especially, gradients tend to be more consistent – i.e. there is less variability in the induced cortical gradients on regions in these transmodal areas. As has been discussed in previous work, the control networks and processing networks show high-between network connectivity as from of top-down information integration [125]. Consequentially, the relationship between similarity and functional connectivity is dampened simply due to decreased variation in the dependent and independent variables. Taken together, the results show that, the global gradient of somatosensory cortices is driven mostly by connection profiles to areas in the same somatosensory network – i.e. auditory areas drive auditory organization – while for more integrative areas, the global gradient is a composition of gradients induced by targets spanning multiple networks, though generally those associated with the Attention, Control and Default networks.

Chapter 7

Analyzing pairwise topographic cortical organization

7.1 Background

Analyzing connectivity between pairs of cortical regions is important for understanding how the brain is structured and how it transmits information. Various cortical areas show graded organization along intrinsic axes defined by functional connectivity. Visual areas are organized along two retinotopic axes, defined by the position of objects in the visual field. These two axes are characterized by radial distance from the center of the visual field (eccentricity), and angle of rotation around the visual field (polar angle). In the primary visual cortex, eccentricity maps onto V1 along a posterior-to-anterior axis, anchored at fovea, while polar angle maps orthogonally onto V1 in a superior-inferior manner (see Figure 3). Higher-order visual areas like V2 and V3 display similar retinotopic organization patterns to V1 (i.e. eccentricity and polar angle maps) albeit with sign-flipping (opposed direction of rotation) of retinotopic organizations, relative to V1. Prior research has shown that functional connectivity is stronger between V1-V2 and V2-V3 for areas corresponding to the same eccentricity values. Cortical areas that map onto the same (eccentricity, polar angle) retinotopic coordinates in the visual field are more-strongly functionally connected to one another. The strength of this functional relationship weakens as the distance between (eccentricity, polar angle) retinotopic coordinates increases [126]. That is, the more dissimilar two locations are in terms of their visual field coordinates, the weaker their functional connectivity. These functional relationships provide a basis for examining the "pairwise" organization between cortical areas e.g. how coordinates in two cortical areas are related to one another. In neural tracing studies in mice, prior research identified topographically organized projection patterns from V1 throughout the rest of the mouse visual cortex. These studies found that topographic organization patterns in V1 are mapped onto higher-order visual areas, such that termination of anterograde tracers retained the spatial organization of the injection sites (accounting for rotations, translations, and projection density differences) (Figure 7.1) [21]. Combined with the functional connectivity results, these analyses motivate the hypothesis that there is a degree of coherence in the spatial relationships of connectivity patterns between pairs of cortical areas in visual cortex, and possibly throughout the isocortex.

As described in Chapter 6, connectivity gradients – i.e. connectopy maps – capture the dominant axes of functional variation within a cortical region. Analogous to principal component analysis (PCA), approaches like laplacian eigenmaps or diffusion embedding compute low-dimensional embeddings of functional connectivity that capture some notion of spatial variation within a cortical area. However, in contrast to PCA, laplacian eigenmaps and diffusion embedding are non-linear decomposition methods that do not enforce strict orthogonal projections onto linear subspaces. As such, these approaches can identify low-dimensional manifolds embedded in a higher dimensional space along which data varies, as opposed to strictly on planar sub-spaces. Recent work has explored the spatial organization of whole brain [18] and local functional connectivity [3] using gradient-based approaches. However, these methods stop short of examining the topo-

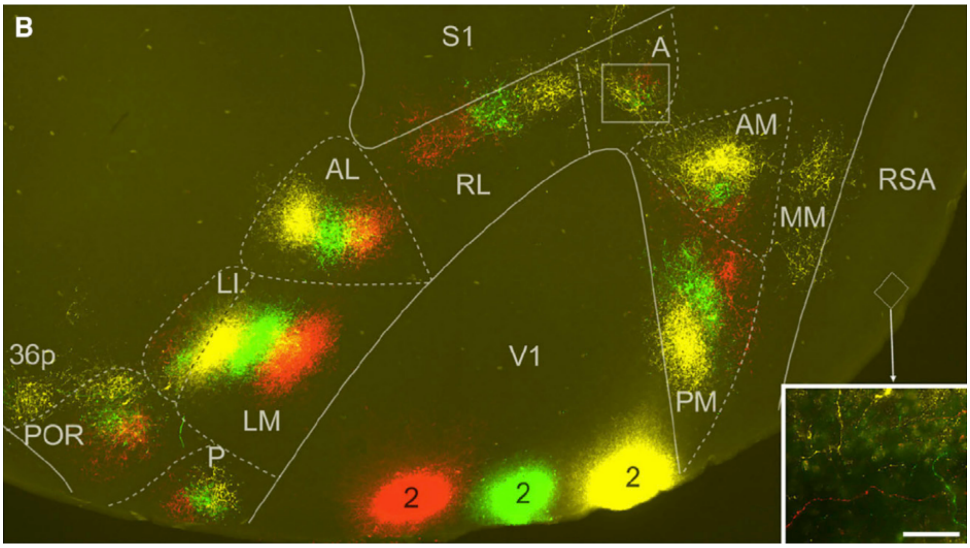


Figure 7.1: Streamline projection patterns in the mouse visual cortex identified via triple-color anterograde tracer injections into V1 [21]. These tracing results demonstrate topographically organized projection patterns (yellow, green, red) to higher-order visual areas, albeit with variable connection spreads, densities, and sign flips.

graphic organization *between* pairs of cortical areas. Recently, by expanding on the work of [22], [23] introduced the idea of pairwise regional linear maps between cortical areas to examine the topographic relationships between *pairs* of cortical areas. They found that, at the voxel-level, areas corresponding to the same functional networks shower higher pairwise topographic ordering, mirroring those results in [21] in the mouse visual cortex.

Assume we are given a whole-brain BOLD time-series matrix $T \in \mathbb{R}^{N \times t}$ with N vertices and t time-points. Each vertex v is characterized by a BOLD time-series vector $T_v \in \mathbb{R}^t$. We define the functional connectivity between two vertices i and j as $\rho(i,j) \mapsto \mathbb{R} \in [-1,1]$ as the Pearson correlation between their respective time-series vectors, T_i and T_j. We represent the pairwise vertex-level correlation matrix between all vertices in regions a and b as R. For vertex i in region a and vertex j in region b, we are interested in the relationship between $R_{i,j}$ and the coordinate pair $(f_a(i), f_b(j))$, where $f()$ is some arbitrary spatial function of a cortical vertex.

We asked: is the coordinate pair $(f_a(i), f_b(j))$ predictive of the strength of functional connectivity between this coordinate pair, $R_{i,j}$? We hypothesized that the connectivity organization between two pairs of regions would be related to the organization of those regions independently, and that we would be able to identify these relationships, at a minimum, in the primary sensory areas. We hypothesized that these relationships would be strongest between pairs of regions belonging to the same functional networks and weaker between regions corresponding to different functional networks.

7.2 Methods

7.2.1 Connective field models between coordinate pairs

Assume we are given two cortical areas a and b. We computed the pairwise functional connectivity matrix R between a and b that characterizes the correlation between every vertex in region a with every vertex in region b. We are interested in examining whether the functional connectivity between the two region pairs can be modeled as a function of *space*. That is, if we choose an arbitrary location in region a, can we predict where in region b this location will be *most strongly connected*.

We identified (source, target) coordinate pairs using connective field models, as described in Section 10.2 and in [22], where field models were fit using the geodesic distances between spatial cortical points (Equation 7.1). Field models identify the "center of mass" of a scalar field, and are consequently more robust than simply identifying the maximum / minimum values of a scalar field. Given a scalar field distributed over a set of vertices in a target region (e.g. functional correlation between a vertex and a cortical area), we fit an isotropic Gaussian density to the scalar field by estimating the mean, μ (centroid coordinate of the density, corresponding to spatial location on cortex) and variance σ (spread of isotropic Gaussian), such that the "similarity" between the density and the scalar field is maximized (see Figure 10.1). Here, we measure "similarity" using

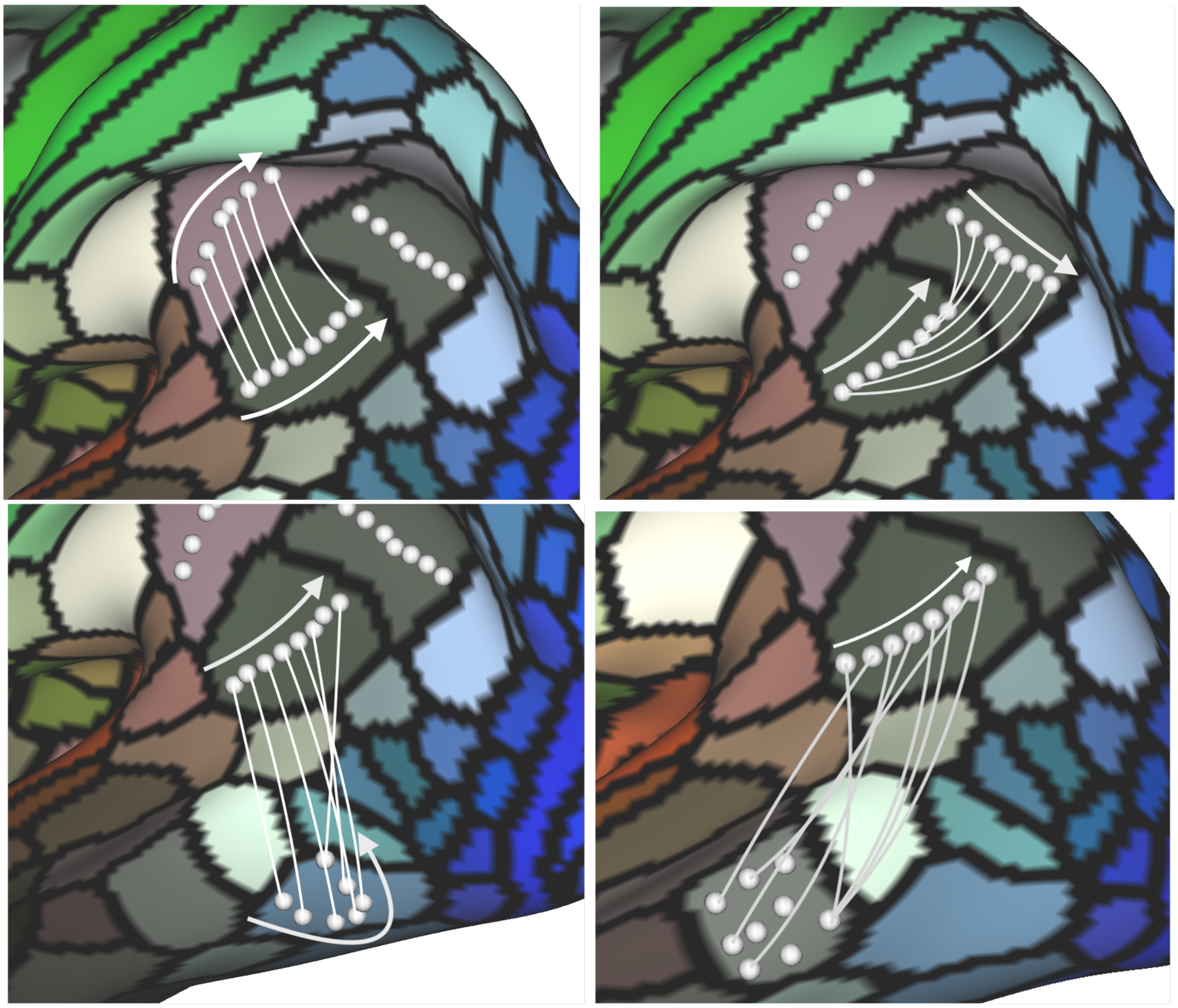


Figure 7.2: Example synthetic pairwise regional mappings for trajectories within source region PGi. While these examples show organized mappings, in that they are smooth, this is not guaranteed for actual estimated mappings computed using functional connectivity. Additionally, estimated mappings do not need to be invertible or symmetric, e.g. source-to-target mappings are not guaranteed to equal target-to-source mappings.

the Pearson correlation between the density and the scalar field. Field model "amplitude" (α) is estimated by optimizing a scalar factor, such that the least-squares distance between the scaled density and the scalar field is minimized.

$$g(v) = \alpha \times exp - \left[\frac{d(v, v_0)^2}{2\sigma^2}\right] \tag{7.1}$$

The field model centroid, μ, corresponds to a vertex in the target region. As such, the field model maps a spatial location in the source region to a spatial location in the target region. Each vertex in the source region maps onto a vertex in the target region – however, this mapping is not necessarily invertible. For source and target regions a and b, for each vertex $v_{a,i} \in a$, we fit a connective field model over the connectivity of $v_{a,i}$ with all vertices in b, where $\mu_{a,i}$ and $\sigma_{a,i}$ are the estimated mean and variance parameters for $v_{a,i}$, such that $\mu_{a,i} \in \{b\}$. We then have $X = [v_{a,1}, v_{a,2} \dots v_{a,j}]$ and $Y = [g(v_{a,1}), g(v_{a,2}), \dots g(v_{a,j})]$, where $X, Y \in \mathbb{R}^{j \times D}$, and are interested in the relationship between X and Y (Figure 7.2).

7.2.2 Computing pairwise topographic maps

We are interested in assessing whether knowing the location in a source region is predictive of functional connectivity strength in the target region, and further, how topographically organized the connectivity between two region pairs is. We examined these questions from two perspectives: 1) a linear transformation approach as described in [23], and an approach examining the variance in functional connectivity between pairs of cortical areas.

Given two cortical areas, fitting connective field models as above maps every source vertex to a single, possibly non-unique target vertex. We can examine how "organized" these spatial mappings are. By "organized" we mean: are the spatial relationships between points in the source region maintained in the mapped target coordinates? For example, as described above, the visual areas show organized maps when examined using anterograde tracing studies. Here, we computed a metric of "organization" using a least-squares approach. Given a set of source and target coordinates, $X, Y \in \mathbb{R}^{D \times N}$ (where N is the number of samples, and D the sample dimension), we treat this problem as an orthogonal Procrustes problem,

$$\hat{\Omega} = \underset{\Omega}{\arg\min} \, \|\Omega X - Y\|_F^2 \tag{7.2}$$

and solve for $\Omega \in \mathbb{R}^{D \times D}$, a linear, affine Procrustes transformation matrix with $\Omega^T \Omega = 1$, that maps X to Y, and accounts for rotations, translations, scaling, and reflections between the source and target coordinates. Analogous to computing R^2 in a linear model, we can estimate a goodness-of-fit statistics, ν, for the Procrustes transformation, that is based on the sum-of-squared errors between the transformed source coordinates, $\hat{\Omega} X$, and the target coordinates Y. A value of $\nu = 0$ indicates no organization at all (e.g. source-to-target mappings are random), while a value of $\nu = 1$ indicates maximal pairwise topographic organization, where spatial relationships between source and mapped target coordinates are maintained. Here, we represent ν as a percentage, [0%,100%]. Using the HCP-MMP atlas, we estimated ν for every (source-target) region pair. We consider the case where our spatial function $f_a(i) = (X_i, Y_i, Z_i)$, the Euclidean spatial coordinate of vertex i.

7.2.3 A non-parametric analysis of connectivity variance

We sought to examine the relationship of within-region organization and between-region organization by relating local connectivity gradients to pairwise topographic organization patterns. Given that connectopy maps capture the primary directions of functional variation within a cortical area, we sought to determine whether position along these axes

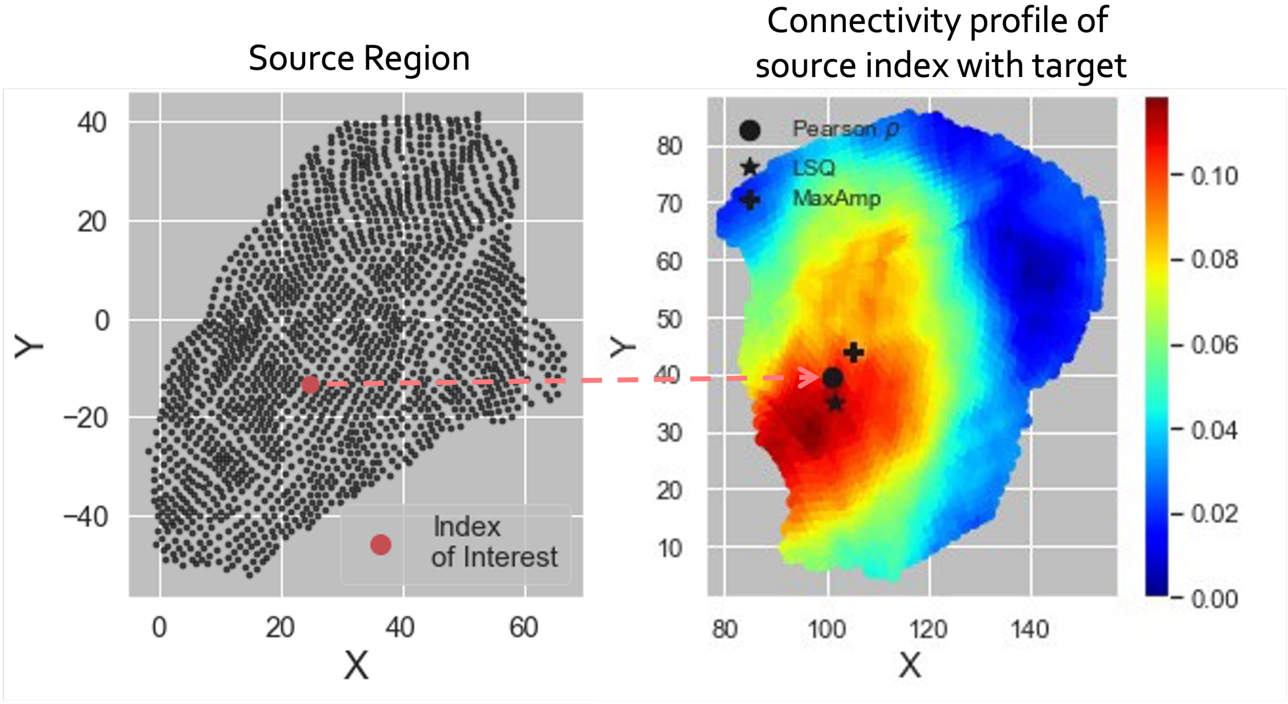


Figure 7.3: Estimating source-target coordinate maps using a connective field model (CFM) approach, as described in [22, 23]. The "+", "$\circ$", and "$\star$" symbols in the right subplot represent three candidate centroid coordinates for the corresponding shown scalar map.

was predictive of organization of connectivity *with another cortical area*. Assume region a has k vertices and corresponding connectopy map $C_a \in \mathbb{R}^k$. The connectopy value of vertex i in region a is characterized $C_a(i)$ and captures the position of vertex i along this specific axis of functional variation.

We considered a non-parametric approach to examine pairwise cortical organization, where the spatial function is defined using the connectopy of each cortical area, as $f_a(i) = C_a(i)$. Assume we have two regions, a and b, for which we have computed connectopies C_a and C_b. We bin each connectopy into N discrete regions, and examine functional connectivity within all bin pairs We demonstrate this problem in Figure 7.4. Fundamentally, our hypothesis is that the connectivity between two region pairs is *not* spatially random – that is, that there is some coherence to the pairwise connectivity

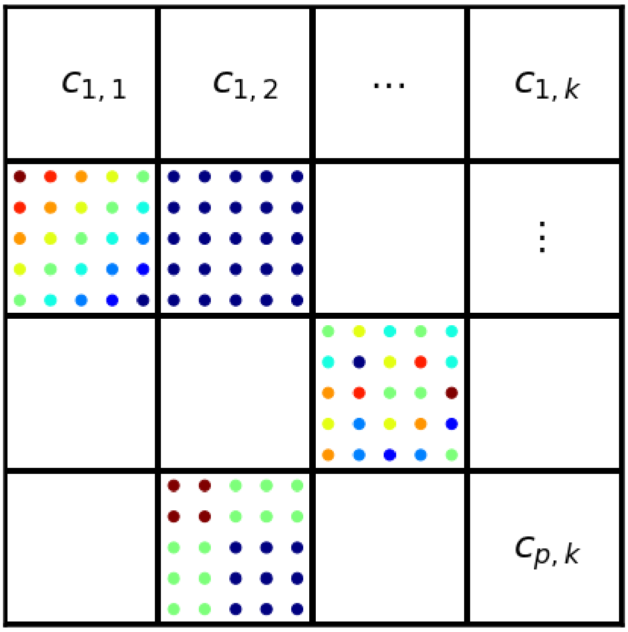


Figure 7.4: Example schematic demonstrating binning of the spatial domain ($C_a \otimes C_b$) with unique functional connectivity profiles per bin. Here, we utilize colors to reflect different functional correlations. For example, we see that bin $c_{2,2}$ is characterized by homogeneous temporal correlation between regions a and b. Similarly, bins $c_{2,1}$ and $c_{3,3}$ are characterized by random temporal correlations, while in $c_{4,2}$ is characterized by a more structure temporal correlations.

between two regions. The degree of coherence can be measured using an analysis of variance of the functional connectivity between pairs of regions.

First, we estimated the pairwise functional correlation matrix R for regions a and b, and binning connectopy C_a into p discrete bins and connectopy C_b into k discrete bins. We then computed the ratio, ν of aggregate variance and mean within-bin variance, as follows:

$$\nu = \frac{\sigma^2(R)}{\frac{1}{p \times k} \sum_{i \in p} \sum_{j \in k} \sigma^2(R_{i,j})} \tag{7.3}$$

This is roughly analogous to fitting an ANOVA model, where the dependent variable is the functional connectivity strength, and the independent variable is a discrete indicator variable corresponding to a pairwise connectopy bin. Irrespective of the source-target region pair, we expected within-bin variance to be considerably less than the aggregate variance. If our hypothesis regarding more *structured* functional organization is true, we hypothesized that the relationship between within- and aggregate-variance would be related to which pair of regions we considered. Specifically, the stronger the topographic organization between two regions pairs, the larger the ratio ν – this would indicate that within-bin variance is less "random" than the aggregate variance, and would represent some inherent structure in the functional organization patterns between two regions.

We hypothesized that the aggregate variance in functional connectivity between two cortical parcels would be significantly greater than the mean variance of connectivity within a set of discrete bins, and that the magnitude of ν, the ratio of these two measures, would be related to mean connectivity strength and mean distance between the two regions of interest. This statistic is analogous to an F-statistic comparing within-group and between-group variance in an ANOVA model, while not making any assumptions about the distribution of errors in the linear model, and without accounting for degrees of freedom. As such, η represents a "model free" analysis of the variation of functional connectivity between two regions. We hypothesized that the strength of these relationships would be moderated by functional network identity.

7.3 Results

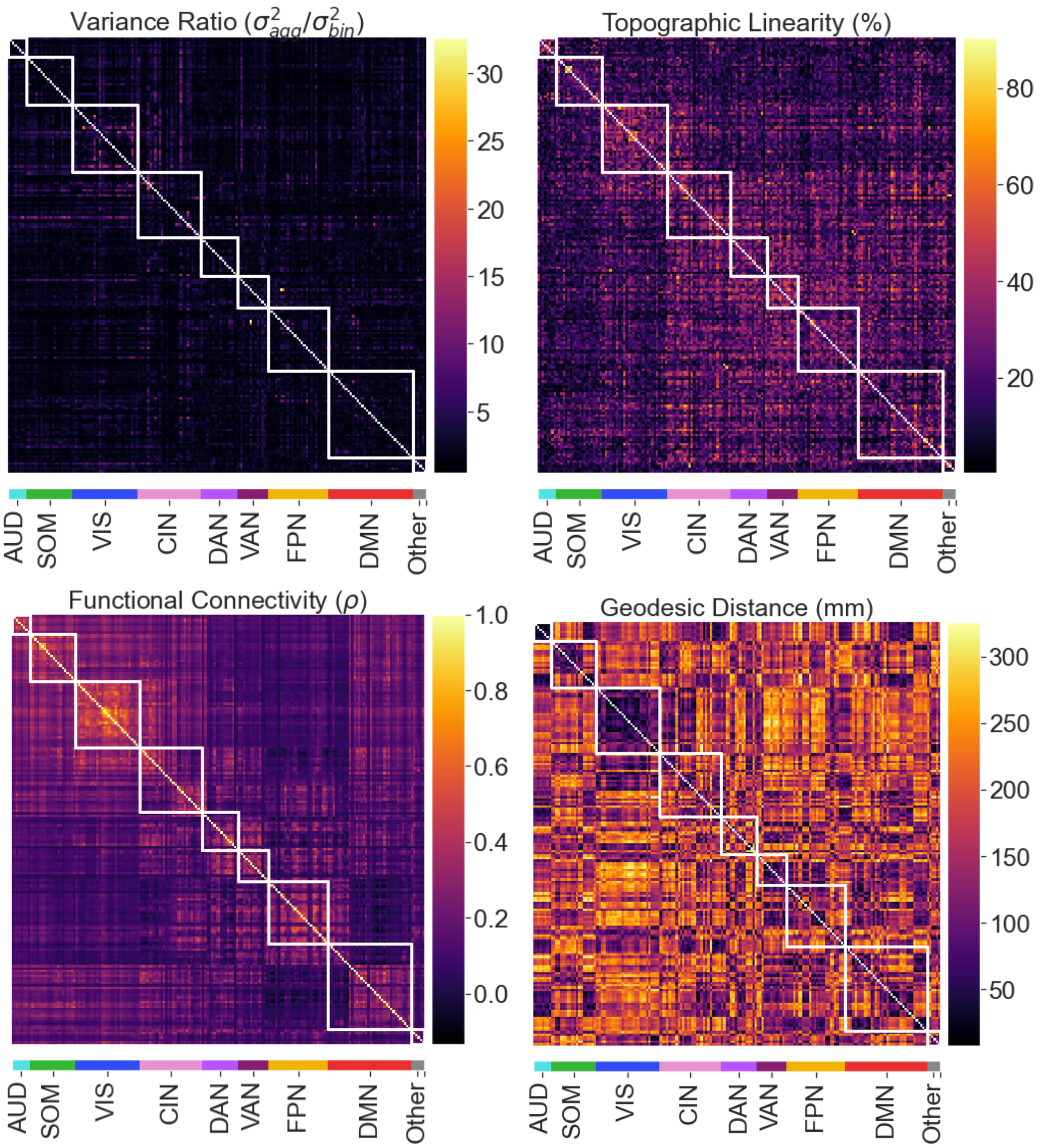
In Figure 7.2, we show the pairwise matrices for aggregate-vs-within-bin variance estimates, functional connectivity, geodesic distance, and linear topographic organization. Empirically, we see a strong within-network effect, and explored this further. Prior research has shown that functional connectivity strength decreases with increasing geodesic distance between cortical areas [127], and we find similar results (β=-0.005ρ/cm (+/-

1.23e-4); R^2=0.086; p=0.0) (Figure 9.3). We examined transformation linearity, ν, and variance-ratio, η, estimates from three perspectives: 1) simultaneously across all pairs of cortical regions, 2) averaged within functional networks, and 3) independently within functional networks.

7.3.1 Pairwise organization is related to functional connectivity strength *and* distance between region pairs

We find that, on average, topographic organization increases when regions are more strongly connected to one another, and decreases when regions are farther apart (Figure 7.6). When we aggregate across region pairs within a single functional network (colored points) and across region pairs from different functional networks (black point), we see that topographic organization increases between regions belonging to the same functional network, and is minimized for regions belonging to different networks. Contrary to our original hypothesis, that topographic organization would be greatest in the primary auditory, sensorimotor, and visual areas, we found that mean topographic organization was maximized in dorsal and ventral attention networks (DAN and VAN, respectively), and in the frontal parietal network (FPN).

However, when we examined each functional network independently, results were more aligned with this prior hypothesis. We found that functional connectivity strength explains a significant amount of variation in topographic organization for all but the ventral attention network (VAN), while geodesic distance was inversely weakly related to topographic organization. The strongest effects were found within the primary auditory, somatosensory, and visual areas. These results are inline with previous studies that identified coherent tonotopic and retinotopic functional organization patterns [35, 102, 126, 128].


Variance Ratio ($\sigma^2_{agg}/\sigma^2_{bin}$)

Topographic Linearity (%)

Functional Connectivity (ρ)

Geodesic Distance (mm)

Figure 7.5: Pairwise regional statistics. Each index represents the estimated statistic for a single region pair, sorted by the functional networks to which regions are assigned. Top left: ratio of aggregate-to-within-bin connectivity variance. Top right: topographic organization computed using Procrustes transformations. Bottom left: functional connectivity between region pairs. Bottom right: geodesic distance between region centroids.

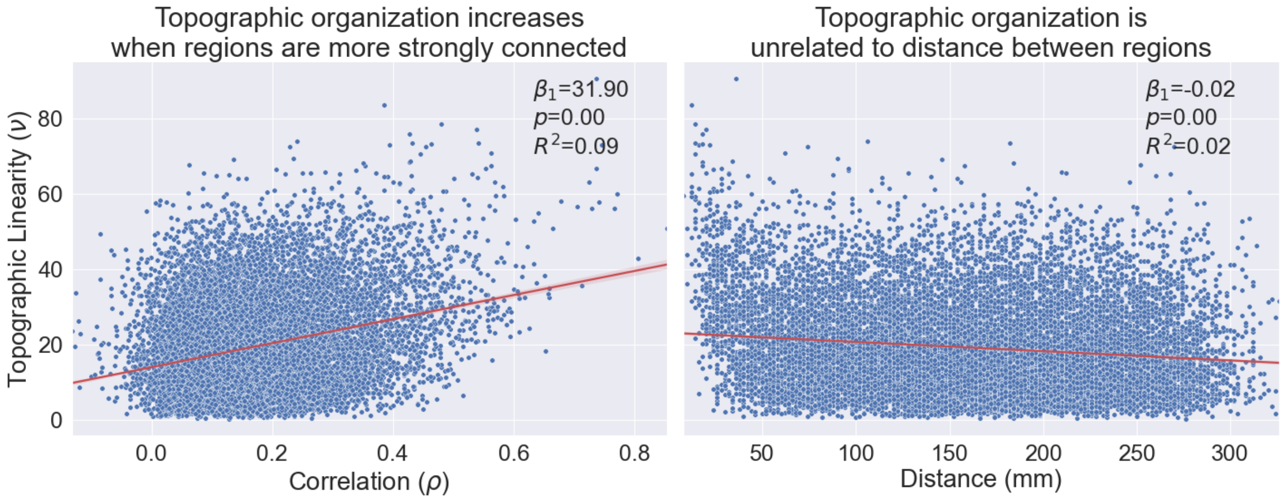


Figure 7.6: Topographic organization is related to functional connectivity strength and geodesic distance at the *regional* level.

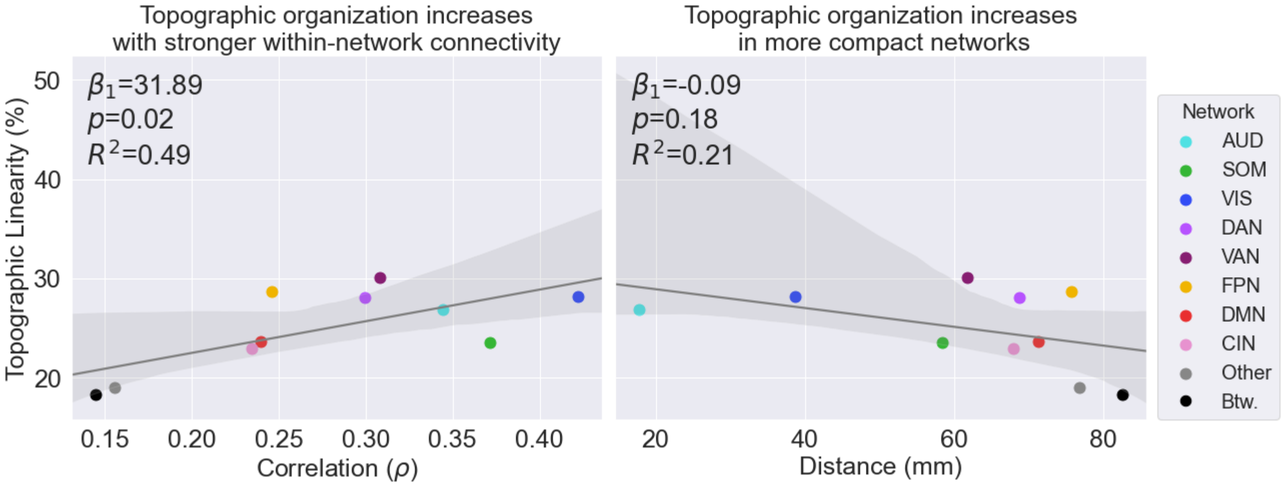


Figure 7.7: Topographic organization is related to functional connectivity strength and geodesic distance at the *resting-state network* level.

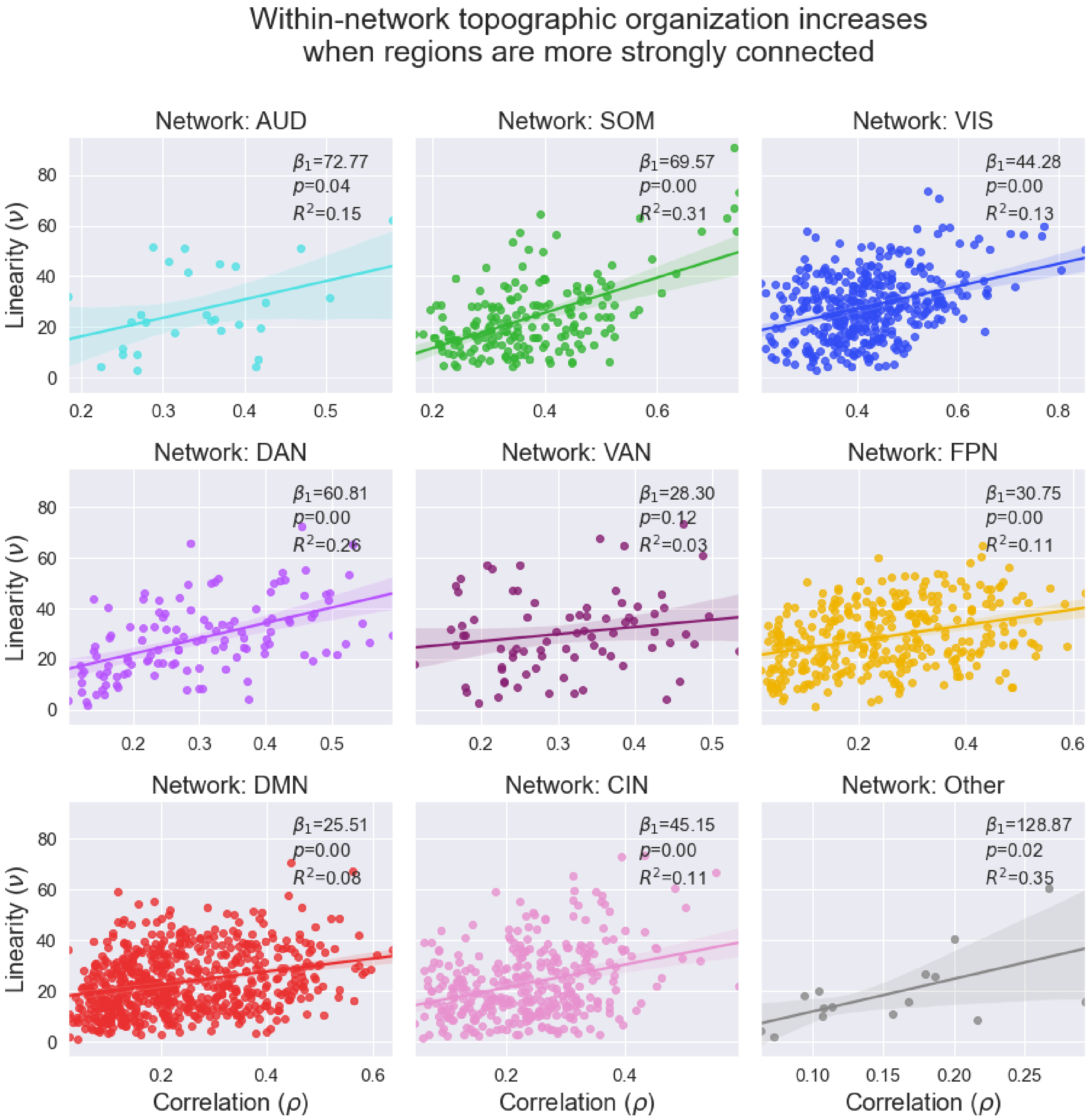


Figure 7.8: Topographic organization is differentially related to functional connectivity strength when considering each resting-state network independently.

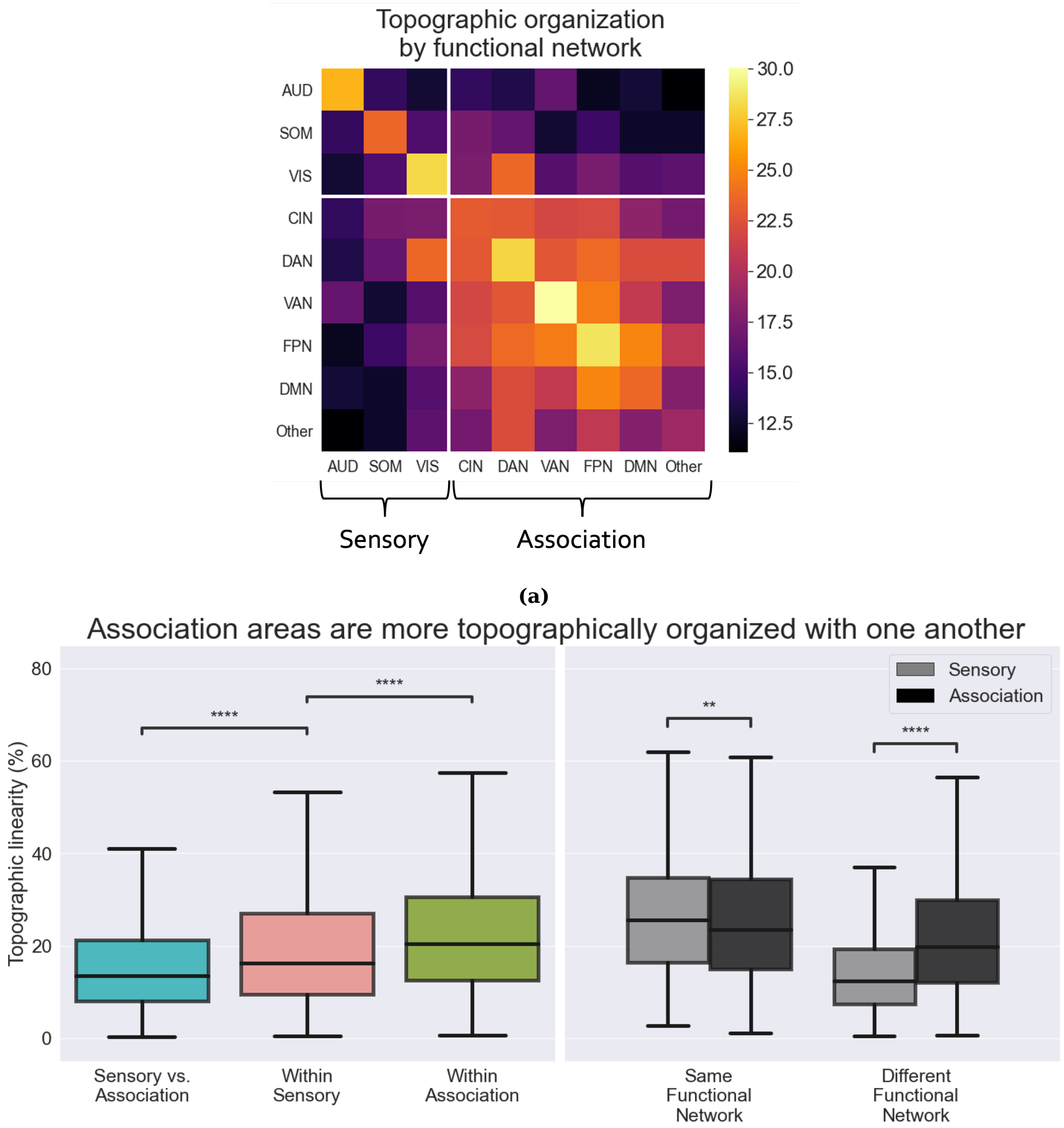
7.3.2 Topographic organization is mediated by source and target functional networks

We examined the strength of topographic organization as a function of the source and target region functional network. We hypothesized that primary sensory areas would be more organized than heteromodal networks (such as the attention networks, DAN, VAN, and FPN, or the default mode network, DMN) due to the high-degree of inter-network connectivity displayed by the higher-order networks (as discussed in [19]). We hypothesized that this stronger inter-network connectivity would preclude, or at least restrict, any *spatially-specific* topographic organization patterns.

In Figure 7.9(a), we show the mean ν estimates for within- and between-network maps. In support of our hypothesis, we see a strong-diagonal effect e.g. higher ν within-network estimates, as compared to between-network. This lends evidence to the claim that connectivity for pairs of regions belonging to the same network is more topographically organized, relative to pairs of regions corresponding to different functional networks. In contrast to our hypothesis above, we see that primary sensory areas show weak between-network topographic organization while the heteromodal networks show high between-network topographic organization. If we bin networks into "sensory" (AUD, SOM, VIS) and "association" (CIN, DAN, VAN, FPN, DMN, Other) groups, we see that association networks show significantly greater between-network organization, relative to sensory networks (Figure 7.9(b)).

7.3.3 Topographic organization follows a unimodal-to-heteromodal axis

We examined the whole-brain gradients induced by the pairwise topographic organization statistic ν and found that they closely resemble those gradients induced by functional connectivity (Figure 7.10(a)). We found that the topography gradients follow a similar



(a)

(b)

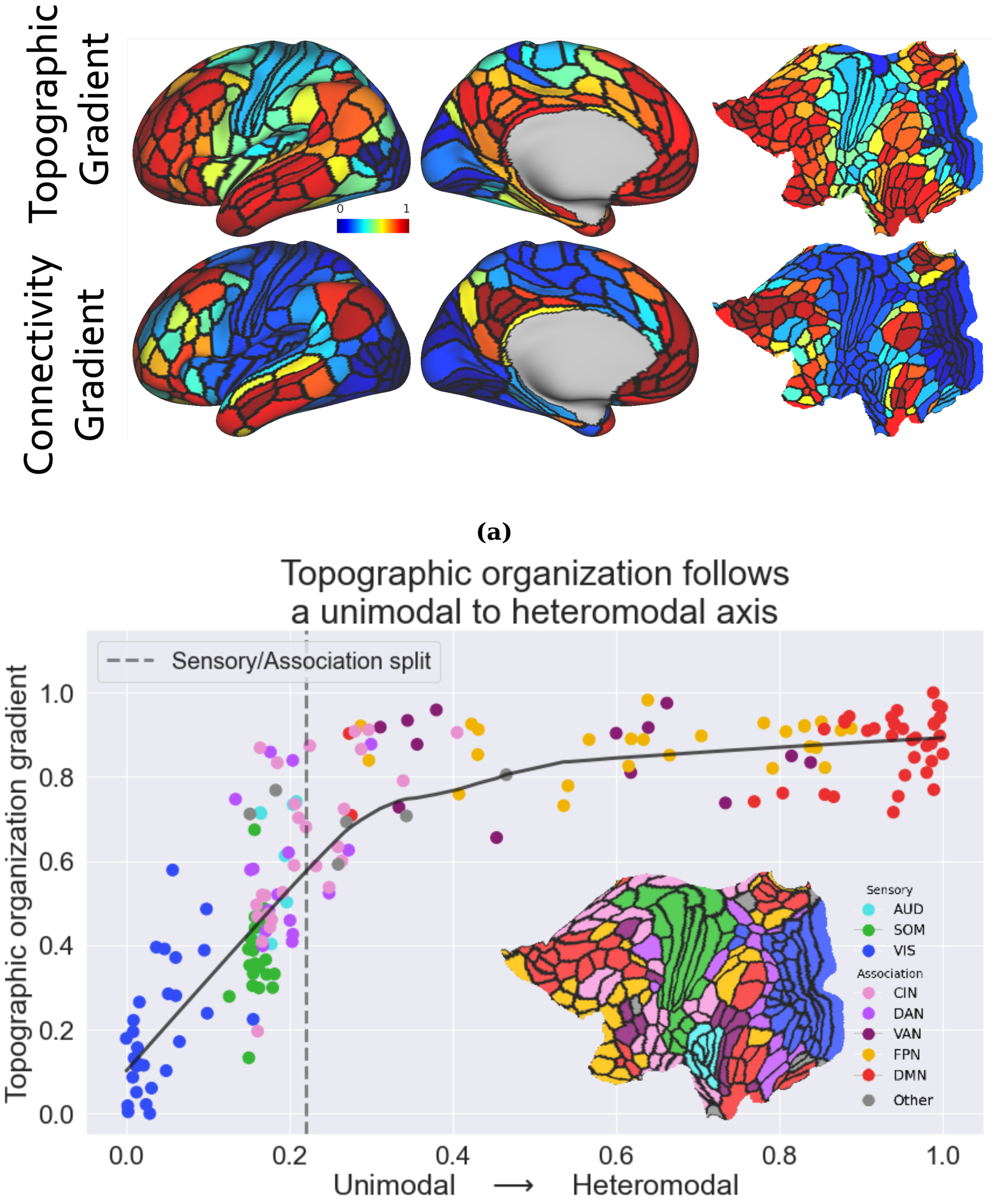
Figure 7.9: Magnitude of topographic organization by functional network. We bin topographic organization estimates, ν, by functional network (A), and examine organization as a function of the sensory vs. association split (B). Statistical significance computed using Mann-Whitney U-test. ****: $p \leq 1.0\text{e-}04$

sensory-to-association organization pattern, indicating that cortical regions from the same functional networks induced similar degrees of topographic organization on the same target regions. We expect that cortical regions closer to each other in the functional processing hierarchy (e.g. areas in the primary visual cortex) would induce more similar patterns of topographic organization on the same target region, as opposed to regions from different levels in this hierarchy. Empirically, our findings support this interpretation.

In Figure 7.10(b), we plot the primary topographic gradient against the primary functional connectivity gradient, and note that we can identify a rough sensory-association split in the topographic organization. These results reflect the data shown in Figure 7.9(a), where association areas displayed more organized maps amongst all other association areas, while sensory areas only displayed strong organization with areas from the same functional network (e.g. VIS-VIS, AUD-AUD, SOM-SOM).

7.3.4 Connectivity variance is unrelated to connectivity strength

In contrast to the topographic organization statistic, ν, representing the "linearity" of mappings between source and target region pairs, we found that the variance ratio statistic, η, yielded inconclusive and insignificant results (Figures 9.15, 9.16, 9.18, and 9.17). We found that the magnitude of η was almost entirely unrelated to the geodesic distance between pairs of cortical regions when considering all pairs of regions simultaneously (Figure 9.15), when aggregating across functional networks (Figure 9.16), and within each independent network (Figure 9.18). The relationship between functional connectivity strength and η was weak, with functional connectivity explaining no more than 5% of the variance in η. Apart from the auditory cortex, where geodesic distance between region pairs significantly explained roughly 20% (p=0.02) of the variance in this ratio, all other functional networks were inconclusive.



(a)

(b)

Figure 7.10: Gradients of topographic organization and functional connectivity projected onto the cortical surface (A), and plotted against one another (B). We also show the functional network assignments of each cortical area, as defined by [16].

7.4 Discussion

Motivated by neural tracing work in mice, and preliminary functional MRI studies in humans, that identified retinotopic organization between pairs of primary and secondary visual areas, here we sought to examine the pairwise organization for the entire cortex. Other studies have examined connectivity in the somatosensory cortex and found that subregions of the primary sensory cortices (e.g those corresponding to the lower limbs, trunk, ocular, face areas, etc.) are more functionally connected to one another and are also characterized by similar histological feature profiles. As such, we hypothesized that we could extend the results examining pairwise topographic organization in the visual cortex, to other primary sensory areas (such as the somatosensory and auditory cortices), and possibly to other association and higher-order cortical regions.

As discussed in Chapter 6, prior work has also examined functional connectivity amongst cortical areas associated with higher-order attention networks and the default mode networks [19], and found that there exists an attention "core" associated with a "multiple demand" function. These core regions tend to co-activate together, irrespective of the task. However, each still preferentially activates with respect to certain task domains (working memory, gambling, etc.). We hypothesize that these results can explain, in part, our findings that cortical areas corresponding these multiple-demand cores show strong inter-network topographic organization, relative to more weakly inter-connected sensory areas. Interestingly, these results corroborate our previous findings from Chapter 6, Figure 6.12 where we examined the similarity of local gradients induced on a region by single target areas. Those results show that target areas from the same functional networks tend to induce similar patterns of spatial organization on a source area – for example, two areas from the visual cortex will tend to induce more comparable patterns of spatial organization on a source region, relative to an area from the visual cortex and the default mode area). In Figure 6.12, we found that target areas from the attention networks (DAN, VAN, FPN) and default mode network (DMN) tend to induce similar patterns of spatial organization on target areas as well.

Taken together, these results support the idea of a certain "rigidity" to the spatial organization of regions in these higher-order transmodal regions. By rigidity, we mean that their organization patterns are more "fixed", and are less dependent on which *other* cortical area is connected to it. Irrespective of which target area is currently connected to a region in these multiple-demand networks, the resulting induced connectopy is more consistent. These conclusions are supported by the fact that *global* connectopies induced on source regions in the multiple demand networks – e.g. the aggregate organization pattern characterizing connectivity to the whole brain – were more similar to the local connectopies in these same multiple-demand networks, relative to the sensory networks.

One limitation of our approach concerns the use of BOLD signals to relate the organization between pairs of regions. By applying the methodology of [22] to coordinate maps between two cortical areas, we are significantly reducing the dimensionality of the data by simply identifying the center of connectivity "mass" – e.g. a single spatial coordinate – rather than considering the entire correlation profile. Our approach is an attempt to extend the neural tracing work to the functional domain, but inherently results in a loss of information. Theoretically, this work could be extended to structural connectivity data, which more closely resembles tract-tracing work. However, we are still left with issue that tractography algorithms generate a *distribution* of termination points for each source voxel or vertex.

Similarly, in this work, we restricted our analysis to group-average connectivity data. In order to determine whether these findings exist at the individual level, it would be prudent to apply these approaches to subject-level connectivity data.

Chapter 8

Future work: learning cortical maps from multi-modal MRI signals

8.1 Background

As we discussed in Chapters 3 and 4, while cortical mapping studies have primarily relied on functional connectivity to drive cortical segmentation via supervised (such as those methods discussed in this dissertation) and unsupervised learning approaches, it remains an open question for how best to incorporate diffusion tractography or other vertex-wise features into this pipeline. Analogous to using functional connectivity features, individual voxels and vertices can be characterized by their neuronal projection patterns, which can be computed using various streamline tracking algorithms. We propose to extend the approaches developed in Section 4 describing how to learn cortical parcellations, and adapt these tools to incorporate both functional and structural connectivity in such as a way as to take advantage of intrinsic structure of each data type.

Previous work has shown a strong relationship between structural and functional connectivity. [127] found that structural connectivity strength explained about 48% of the variation in functional connectivity strength. After accounting for the geodesic distance between cortical areas, the explained variation increased to 69%. Similarly, [127, 129] found that structural connectivity within areas of the Default Mode Network was predictive of the functional connectivity between those same pairs of regions. While structural connectivity can reliably infer functional connectivity strength, the inverse relationship does not necessarily hold – cortical areas can exhibit strong functional connectivity even in the absence of direct neuronal connections. It remains an open question of how to utilize functional connectivity to make inference on the structural connectivity relationships between cortical areas.

Some cortical features, such as myelination, correlate well with functional specialization. For example, cortical thickness increases from the primary sensory areas to association sensory areas (e.g. 3b → 3a → 1 → 2), while myelination has been shown to decrease along this same hierarchy. Further, [130] showed that within sub-divisions of the sensory cortex corresponding to lower limb, trunk, upper limb, ocular, and facial areas, gradations in myelination correlated strongly with functional specialization (computed using analogous non-linear decomposition approaches as described in Chapter 6), providing evidence for a histology-function relationship, at least in S1. Similarly, recent work by [131] identified a face-specific area in the anterior temporal lobe of the fusiform face area in a face vs. place contrast of visual object recognition – the spatial extent of these activation areas correlated strongly with gradients of myelination by [52] at the border of the para-hippocampal cortex. These results lend further evidence to the idea of a coupling between cellular organization and functional specialization.

While most of the focus for learning cortical maps has been on utilizing resting-state and task-based fMRI data to delineate functionally homogenous cortical regions, prior work has shown that structural connectivity can be used to delineate the cortex as well. For example, [40] developed a whole-brain spectral clustering approach to simultaneously map the cortical at multiple resolutions using diffusion MRI. Similarly, [41] applied spectral clustering to tractography data of the inferior parietal lobe (IPL), and identified

7 IPL sub-regions with unique functional and structural connectivity profiles. These 7 sub-regions corresponded strongly with maps of the inferior parietal area previously identified by [5] via histology studies. These methods delineate the cortex on the basis of homogenous structural connectivity profiles. However, reconciling parcels computed from tractography with those computed using resting and task-based functional connectivity remains an open question. In some parts of the cortex, areal boundaries computed using each modality may coincide, but this is not always the case. Often, various heuristics are applied to "match" parcels computing using each modality – most of these *ad hoc* approaches utilize some form of spatial overlap between parcels, while others use more sophisticated information-based approaches to match feature distributions between modalities [37].

As we demonstrated in Chapters 3 and 4, a naive (albeit commonly used) approach to combining different cortical signals is to simply treat each signal as an independent feature, and concatenate these features column-wise in a single matrix at training and testing time. While efficient and straightforward, this approach ignores the structure inherent in the individual signals. Additionally, depending on the sheer number of independent variables or measurements per modality or signal type, one modality may come to dominate the parcel prediction or definition [132]. This is especially evident when applying kernel-based methods that compute pairwise similarities between feature vectors (see Figure 4.12 – in this case, models trained on connectivity alone perform comparably to those that incorporate functional connectivity, spatial location, and univariate cortical measures).

To that end, we frame our proposed future work around two questions:

1. Can we predict one modality or signal, given another?
2. Can we combine multiple imaging modalities / feature types more intelligently to learn cortical parcellations?

We propose to extend the methodology developed in Chapter 4 – by incorporating recent developments in the field of neural networks called *transformer networks*, these proposed methods can offer new insight into the structural-functional connectivity relationship, and can offer improved approaches to fusing imaging modalities for the purpose of image segmentation or classification.

8.2 Methods

8.2.1 (Variational) Auto-Encoders

We propose to explore the use of graph variational auto-encoders to address the first question. Auto-encoders have been used in brain imaging for a variety of purposes, including image denoising, data augmentation, and modality prediction. However, to the best of our knowledge, an analogous coupled auto-encoder approach has not been extended to the analysis of fMRI, DTI, or other brain imaging modalities, nor has the idea of coupled graph auto-encoders been applied to MRI research. Recent work by [93] used coupled auto-encoders in a non-graphical framework to jointly learn embeddings of high-dimensional transcriptomic and patch-seq electrophysiology data. They found that cross-modal reconstructions could accurately capture cell-type-specific gene expression and electrophysiology features. Further, after accounting for the unique structure in each modality, [93] were able to more-precisely identify unique neuronal cell-types by applying unsupervised clustering algorithms to these joint transcriptomic / patch-seq embeddings.

Assume we are given an input feature matrix X. Auto-encoders are neural networks that consist of an encoding step, $\mathcal{E}(X) \mapsto Z$, whereby the network learns a low-dimensional embedding of the input data, and a decoding step $\mathcal{D}(Z) \mapsto \hat{X}$, where the network learns to reconstruct the input data, given the encoded features, with a loss function $\mathcal{L}(X, \hat{X})$ that captures the difference between the input X and the predicted input $\hat{X}$.

Variational auto-encoders (VAE) are a more flexible variant of auto-encoders that apply

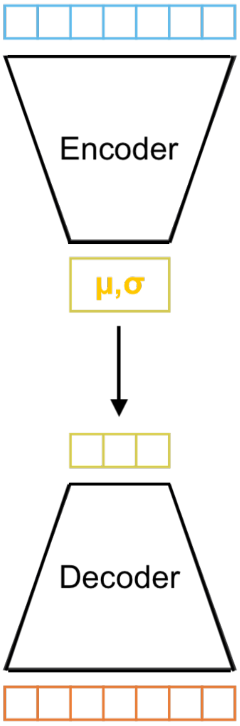


Figure 8.1: Variational auto-encoder framework, where the encoder maps input data, X, to an embedding space, sampled from a parameterized distribution (here, a Gaussian distribution), while the decoder reconstructs the input to recreate the input data as $\hat{X}$.

an information theoretic approach to learning latent embeddings Z. By incorporating a variational learning step, these models can be described as "generative" models – we can sample the embedding space to generate new samples $\hat{X}$ that were not included during training time. Briefly, we assume that the input data X is sampled from some unknown distribution $\mathcal{P}(X;\Theta)$. During the encoding stage, we generate embeddings corresponding to some other distribution $\mathcal{Q}(Z|X)$. While we cannot ever learn P, we can parameterize $\mathcal{Q}(Z|X)$ using some parametric form (such as a Gaussian distribution), such that the embeddings Z are samples from this distribution. We thus learn an encoder-decoder model to generate embeddings of Z of X such that Z look like they are sampled from a Gaussian distribution.

VAEs minimize the KL-divergence between the parametric posterior distribution

$\mathcal{Q}(Z|X)$ and the unknown posterior distribution $\mathcal{P}(Z|X)$ as:

$$D_{KL}\Big(q(z|x) \parallel p(z;x)\Big) = E_{z\sim q}\Big[log\frac{(q(z|x))}{p(z|x)}\Big] \tag{8.1}$$

$$= -\int q(z|x)log\frac{p(z|x)}{q(z|x)} \tag{8.2}$$

After applying Bayes' rule and expanding the logarithms, we have

$$= -\int q(z|x)log\frac{p(x|z)p(z)}{q(z|x)p(x)} \tag{8.3}$$

$$log(p(x)) \geq -D_{KL}(q(z|x) \parallel p(z)) + E_{q(z|x)}\Big[log(p(x|z)\Big] \tag{8.4}$$

$$\tag{8.5}$$

where $p(x)$ is the evidence of the data, which we cannot compute, and the right hand side is the evidence lower bound (ELBO), which can be maximized using gradient descent. In short, the model learns to approximate the unknown posterior distribution $p(z|x)$ with a known, parameterized distribution $q(z|x)$. The Kullback-Leibler term on the right is a regularization term, because it forces the approximation to take a specific parametric form, while the expectation is a reconstruction term that represents the likelihood of the decoder output, given the parameterized embedding. The parameterized embedding is useful, primarily because one can sample from the embedding space (a parameterized distribution with known parameters – most often, 0 mean and spherical unit variance), to generate new samples from the data space.

In order to leverage the spatial dependencies between data points, we can utilize graph neural network models for the encoding phase, $Z = \mathcal{E}_i(X) = \text{GNN}(X, A)$. In a typical auto-encoding framework, we utilize a decoder to reconstruct the input as $\hat{X} = \mathcal{D}_i = \text{GNN}(Z, A)$, to minimize the difference between X and $\hat{X}$. Recent developments in the field of graph-based learning have enabled learning the topological structure of the graph as well, such that, rather than reconstructing the input features, we reconstruct the input adjacency structure A. This involves utilizing an inner-product decoder, typically formulated as $\hat{A} = \sigma(ZZ^T)$, where σ is a non-linearity function.

8.2.2 Coupling auto-encoders to learn joint embeddings

Assume we are given m unique modalities (for example, functional connectivity, structural connectivity, cortical scalar metrics, etc.). We learn a VAE model $\mathcal{M}_i$ for each modality, each with an encoder, $\mathcal{E}_i$, and decoder, $\mathcal{D}_i$ for $i = 1 : m$. For example, in [93], $m = 2$ (transcriptomics and patch-seq features). In the coupled auto-encoder framework, the m encoding models are linked at the deepest embedding state, where the learned embeddings $Z_1 \dots Z_n$ are regularized to be similar to one another. Assuming S samples in each feature matrix X_i, [93] characterized the loss function of these coupled networks as

$$L(X;\Theta) = \sum_{s \in S} \left[\sum_{i=1}^{M} \alpha_i \| x_{si} - \mathcal{D}_i(\mathcal{E}_i(x_{si})\| + \lambda \sum_{j,k \in M} \frac{\|\mathcal{E}_j(x_{si}) - \mathcal{E}_k(x_{si})\|_2^2}{f_{j,k}(X)} \right] \tag{8.6}$$

This loss is characterized by two components that are summed. The first component, on the left, captures the reconstruction loss of a model $\mathcal{M}_i$. The second component, on the right, captures the differences between learned embeddings of all model pairs. The separate loss components are weighted independently: α_i weights the reconstruction error, while λ weights the similarity error, representing an inherent tradeoff between the reconstruction loss and similarity loss. The denominator of the similarity loss, $f_{j,k}(X)$, scales the similarity loss for the same sample, from different feature sets. Since the learned embeddings are sampled from a parametric distribution, and are forced to be similar across feature types, we can predict one feature set, given another.

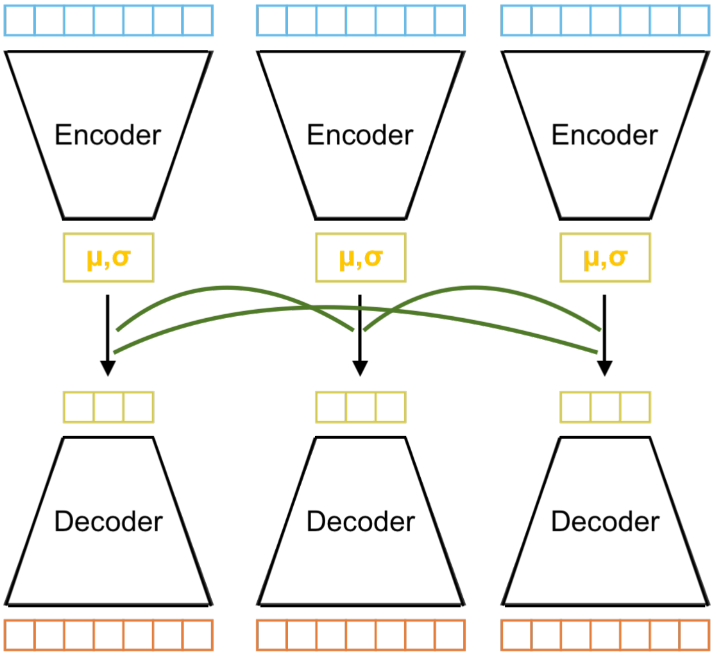


Figure 8.2: A schematic of the coupled variational auto-encoder approach, using three different modalities X, Y, and Z. In the coupled framework, each individual encoder-decoder learns a unique embedding that is regularized to be similar across models.

8.2.3 Transformers

In a typical natural languge processing (NLP) problem, tokens (vectorized representations of characters or strings corresponding to items in a document) are shown to a model, such as a recurrent neural network (RNN) or long-short-term memory network (LSTM) *sequentially*. The model iteratively builds a representation of input signals, as the signals are shown to it. However, numerous studies have shown that these models are unable to "remember" information at different time scales (i.e. from both the beginning and end of a sequence, especially for long sequences). In their seminal paper, [75] introduced the "self-attention" mechanism via a "Transformer" network as a way to alleviate this problem. Transformers are composed of an encoding and decoding phase. During the encoding step, while learning embeddings for each sample (token) of a sequence, the encoder has access to all other samples in the sequence to inform learning the new encoded embedding. This is in contrast to RNN and LSTM networks, where the model only has access to data from the previous $t-1$ samples in the sequence. Instead of providing data sequentially to a model, Transformers examine all signals simultaneously, allowing the model to identify relationships between each token. Ordering or sequential information can be learned or encoded as an additional input signal, via positional embeddings. Recently, transformers have been applied to image processing applications such as image classification, and more recently, feature fusion. We propose to utilize the transformer architecture for the purpose of multi-modal fusion of imaging signals in a cortical classification framework.

Assume, we are given an input signal $X \in \mathbb{R}^{N \times P}$, where each row corresponds to a signal (token) embedding. Self-attention is applied by encoding each token with unique embeddings: queues Q, keys K, and values V.

$$Q = XW_q \tag{8.7}$$

$$K = XW_k \tag{8.8}$$

$$V = XV_v \tag{8.9}$$

where $W_{q,k,v} \in \mathbb{R}^{P \times D}$ are learned linear transformations that project the token embeddings to a new space of dimension M and rows q_i, k_i, and v_i are the new embeddings for sample x_i. Self-attention is then computed as

$$Z = softmax\Big(\frac{QK^T}{\sqrt{M}}\Big)V \tag{8.10}$$

where the dot product, QK^K, captures the similarity between different tokens.

8.2.4 Applying graph auto-encoders to neuroimaging data

As mentioned in [127], structural connectivity can be used to accurately predict functional connectivity strength, while the inverse is not guaranteed to be true, at least using simple linear models. We are interested in two question: 1) can we predict functional connectivity from structural connectivity (and vice versa), and 2) can we *intelligently* combine feature types and then learn cortical parcellations from these fused features?

Predicting structural and functional connectivity strength

Inner product decoders attempt to predict graph adjacency structures from the hidden embeddings learned by the encoding step. The complexity of the inner product decoding step is $\mathcal{O}(N^2)$, where N is the number of graph nodes. For our purposes, each surface mesh has upwards of 32k nodes, especially in the case when working with many graphs – as such, a dense inner-product decoder represents a computationally prohibitive step. To alleviate this complexity, we propose to train graph auto-encoders *after* on *regionalized* connectivity matrices – that is, we'll try to predict connectivity between *regions*, rather than *vertices*. For structural and functional features, assume we have connectivity matrices S and F, respectively. Index $S_{i,j} \in [0,1]$ represents the probability of streamline connections between cortical region i and j as calculated using a fiber tracking algorithm. Index $F_{i,j} \in [-1,1]$ represents the functional connectivity between region i and j, computed using the Pearson correlation. We'll model this problem as a *coupled* graph-reconstruction problem, where each modality has its own specific encoder-decoder pair $(\mathcal{E}_S, \mathcal{D}_S)$ and $(\mathcal{E}_F, \mathcal{D}_F)$.

We'll attempt to reconstruct the weighted adjacency matrix of each modality using an inner-product decoder, each with a modality-specific non-linearity function. For functional connectivity, given $Z_F = \mathcal{E}_F(X_F, A)$, we'll use an inner product decoder defined as $\hat{A}_F = \mathit{tanh}(Z_F Z_F^T)$ in order to map edge values into [-1,1]. For structural connectivity, given $Z_S = \mathcal{E}_S(X_S, A)$, we'll use an inner product decoder defined as $\hat{A}_S = \mathit{sigmoid}(Z_S Z_S^T)$, in order to map edge values in [0,1]. Because the embeddings are learned jointly, such that $Z_S \approx Z_F$, we can predict the weighted adjacency structure of one modality, given the other, and thereby assess the ability of GNNs to accurately map between the connectivity profiles of two modalities.

In [127], the authors explored the relationship between functional and structural connectivity using a linear model framework, assuming no spatial dependencies between different cortical regions. They examined the predictive ability of their model as a function of increasing parcellation resolution. We propose to follow a similar analysis to examine how cross-modal prediction relates to parcellation resolution. We hypothesize that cross-modal prediction accuracy will decrease with increasing parcellation resolution due to increasing spatial dependencies between data points.

Fusing multi-modal features for cortical parcellation

We showed in Chapter 4 that graph neural networks are particularly appealing tool for the purpose of cortical mapping because they can account for the complex spatial dependencies between data points. These algorithms are effective tools for training statistical models that can learn cortical parcellations on the basis of functional connectivity. As mentioned in Section 1.1, identifying cortical areas on the basis of multi-modal *in vitro* and *in vivo* features yielded more accurate cortical areal boundaries that captured variation in different architectural, transcriptomic, and functional signals. However, the question of how to "merge" or "fuse" these features remains inconclusive.

The coupled auto-encoder approach represents one possible solution to this problem, in that joint embeddings can capture the structure inherent to each modality, while enforcing that these embeddings adhere to one another. These joint embeddings can then be used for image or cortical classification.

For each encoder m, we learn an embedding $Z_m = \mathcal{E}(X) \in \mathbb{R}^{N \times D}$. By design, using a coupled-auto-encoder approach, these features will share redundant information. We propose to concatenate these features column-wise in a matrix $Z \in \mathbb{R}^{N \times M*D}$ which is passed through a linear layer to optimally combine these redundant embeddings, which can then be passed further along through convolutional layers in a classification framework as in Chapter 4. Contrastingly, the joint embeddings can be used in an unsupervised learning approach (as described in Section 1.1) to *learn* parcels rather than predict them with respect to a pre-existing cortical atlas.

However, we can also "fuse" these features using the transformer architecture described above, in Section 8.2.3. Assume for each cortical node v_i, we sample M different modalities, of dimensions $x_{i,1} \in \mathbb{R}^{m_1}$, $x_{i,2} \in \mathbb{R}^{m_2} \dots x_{i,M} \in \mathbb{R}^{m_M}$, which we pass through a modality specific linear layer to generate M embeddings each of dimension D, such that for N cortical nodes, our full node-wise feature matrix $X \in \mathbb{R}^{N \times M \times D}$. Each node v_i is characterized by a matrix $X_i \in \mathbb{R}^{M \times D}$. We apply the self-attention mechanism to this matrix, as:

$$Q_i = X_i W_q \tag{8.11}$$

$$K_i = X_i W_k \tag{8.12}$$

$$V_i = X_i W_v \tag{8.13}$$

and compute the self-attention as

$$Z_i = softmax\Big(\frac{Q_i K_i^T}{\sqrt{D}}\Big) V_i \tag{8.14}$$

and then take the sum over the rows of Z to apply the feature fusion on a node-by-node basis. These fused features can then be passed through a general neural network architecture to perform node or graph classification.

Both of these approaches make use of software we have already developed in-house.

With some modification of the source code, testing these methods for the purpose of cross-modal prediction and modality fusion would be relatively straightforward.

Chapter 9

Supplementary Figures

9.1 Chapter 2

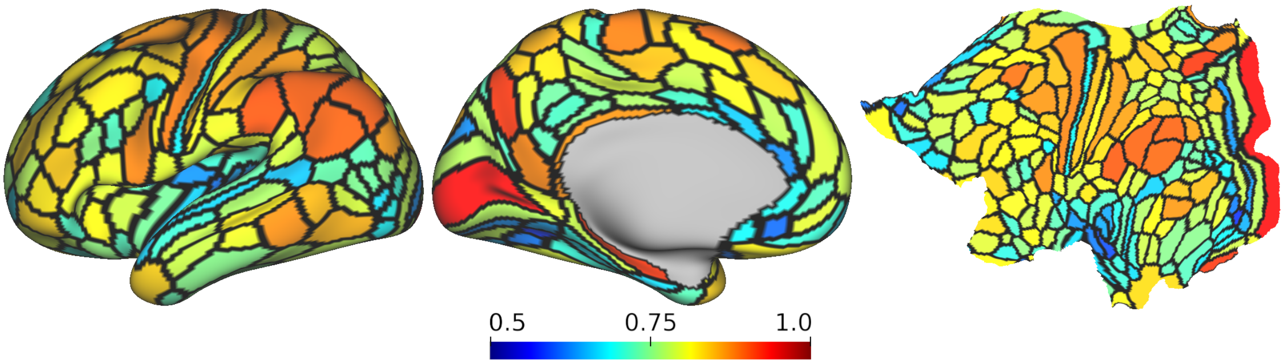


Figure 9.1: Mean regional classification accuracy. Red: most accurate. Blue: least accurate. We see that the primary visual cortex is classified accurately most frequently, as are areas in the parietal cortex near the angular and supramarginal gyri. Interestingly, the optimal model performs notably worse in area 3a in the primary sensory cortex, relative to directly-adjacent areas like 3b and 4. Similarly, the optimal model performs worst in the primary auditory cortex and in the higher-order visual areas related to object recognition.

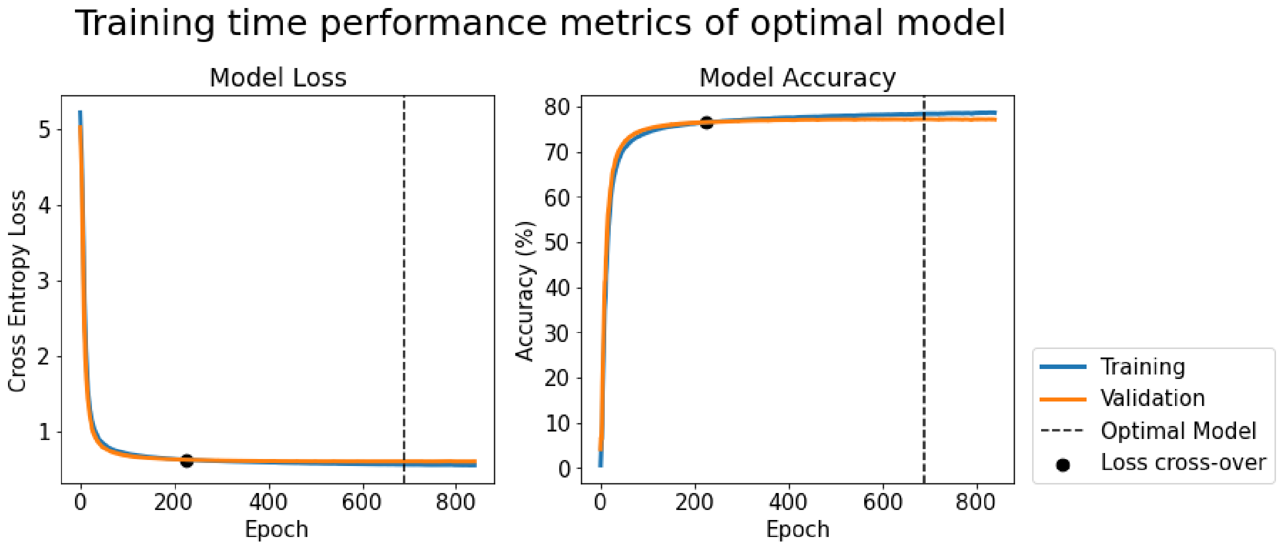


Figure 9.2: Training metrics of optimal model as a function of training epoch. Blue: training. Orange: validation. The dashed black line indicates the training epoch for which validation loss is minimized – the model at this epoch is retained as the "optimal" model. The black dot indicates the point at which the training accuracy is greater than the validation accuracy.

9.2 Chapter 3

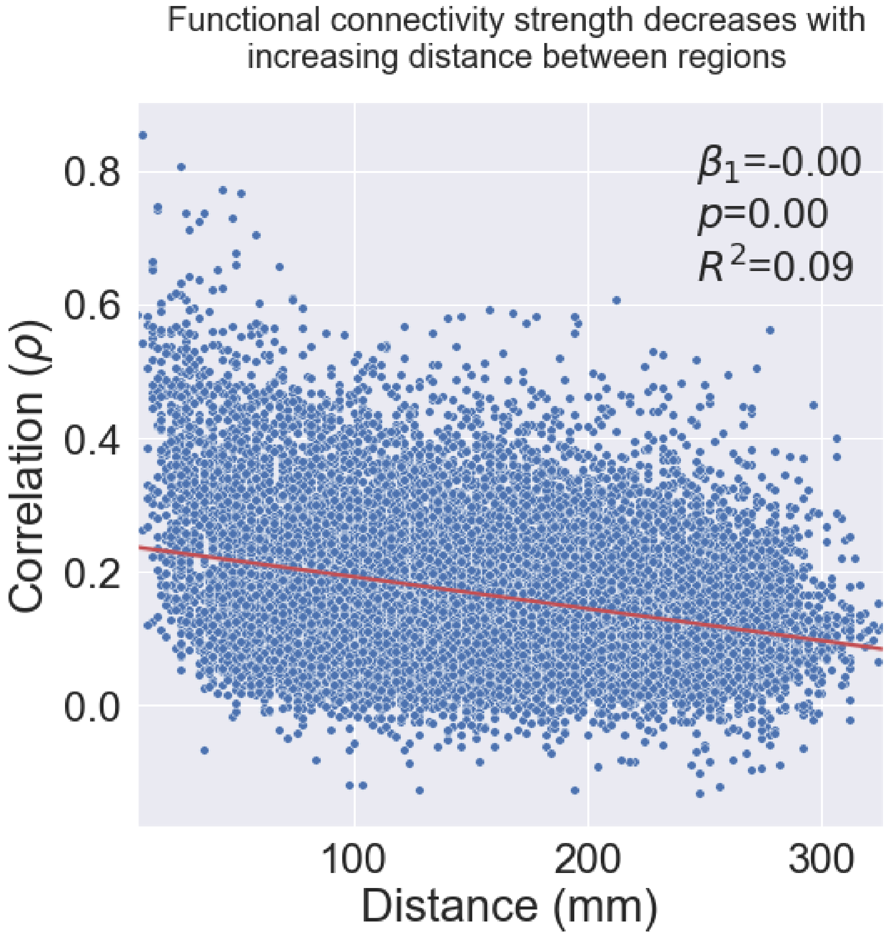


Figure 9.3: Functional connectivity strength between region pairs decreases with increasing distance between region pairs.

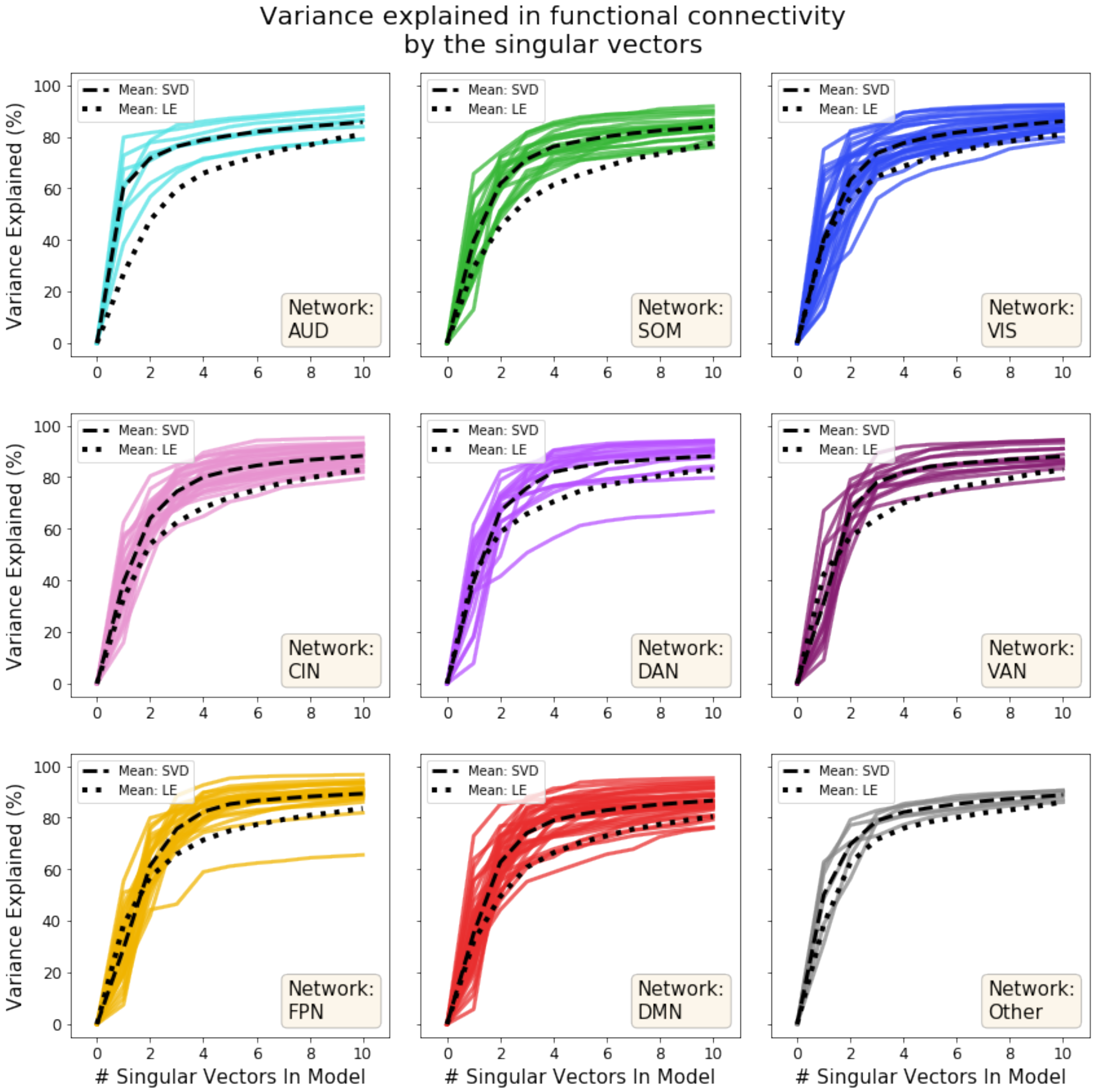


Figure 9.4: Variance explained by the first k global singular vectors in group-averaged connectivity profiles. Each individual plot in the 3-by-3 grid shows the curves for all areas in a single functional network. Each curve corresponds to the variation explained by a single cortical area. The dashed black line represents the mean of all curves for a single functional network. The dotted black line represents the mean when using Laplacian eigenmaps. Data points where the number of gradients is equal to 0 represent the variance explained by the intercept vector.

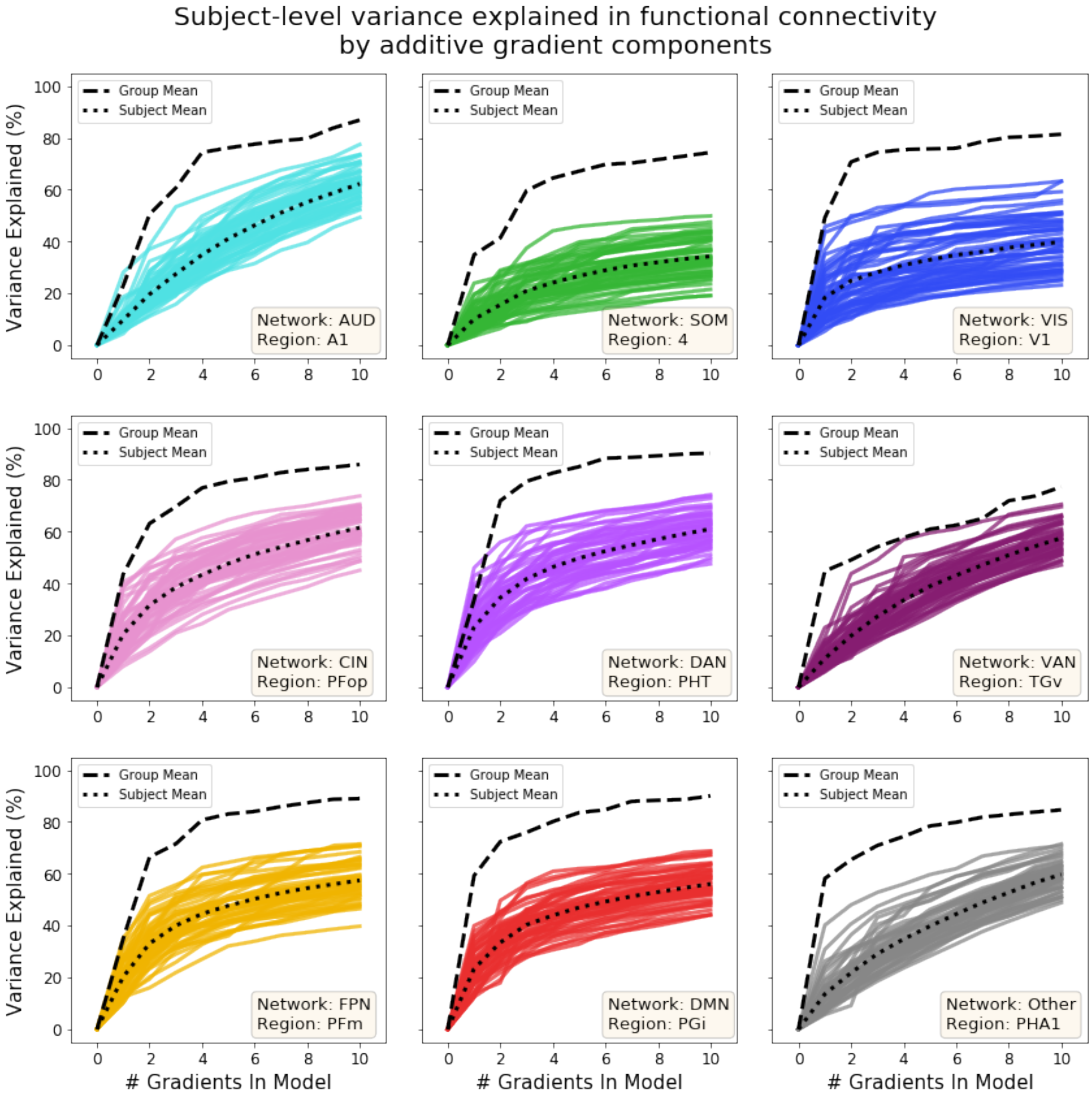


Figure 9.5: Variance explained by the first k global gradients for single-subject connectivity profiles. Each individual plot in the 3-by-3 grid shows the curves for all 154 subject, for a single cortical area in a single functional network. The specific cortical areas are: A1, 4, V1, PFop, PHT, TGv, PFm, PGi, PHA1. The dotted black line shows the mean of all subject-level curves. The dashed black line, shows the curve for each respective area, computed using the group-average gradients in Figure 6.8

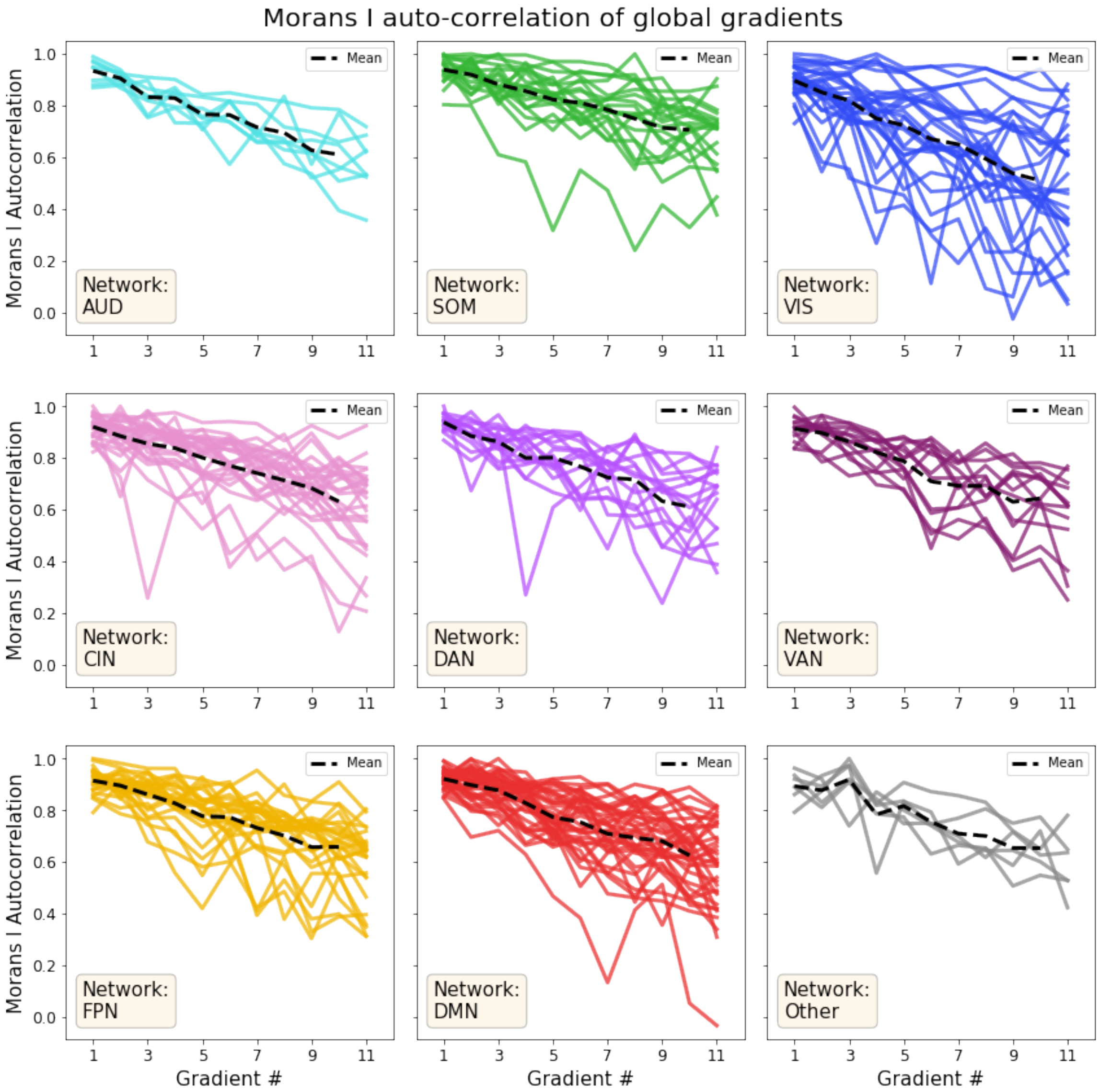


Figure 9.6: Auto-correlation of global gradients. We compute Moran's I spatial auto-correlation for the k-th global gradient for each cortical area, and plot auto-correlation value as a function of gradient order. Moran's I captures the degree of spatial dependence of a signal on a graph domain. For maps with smoothly, slowly varying signals (like the first global gradient), auto-correlation will be high. For a map of pure noise, auto-correlation will be low. We bin data by the functional network of the cortical area. The dashed black line shows the mean of all curves.

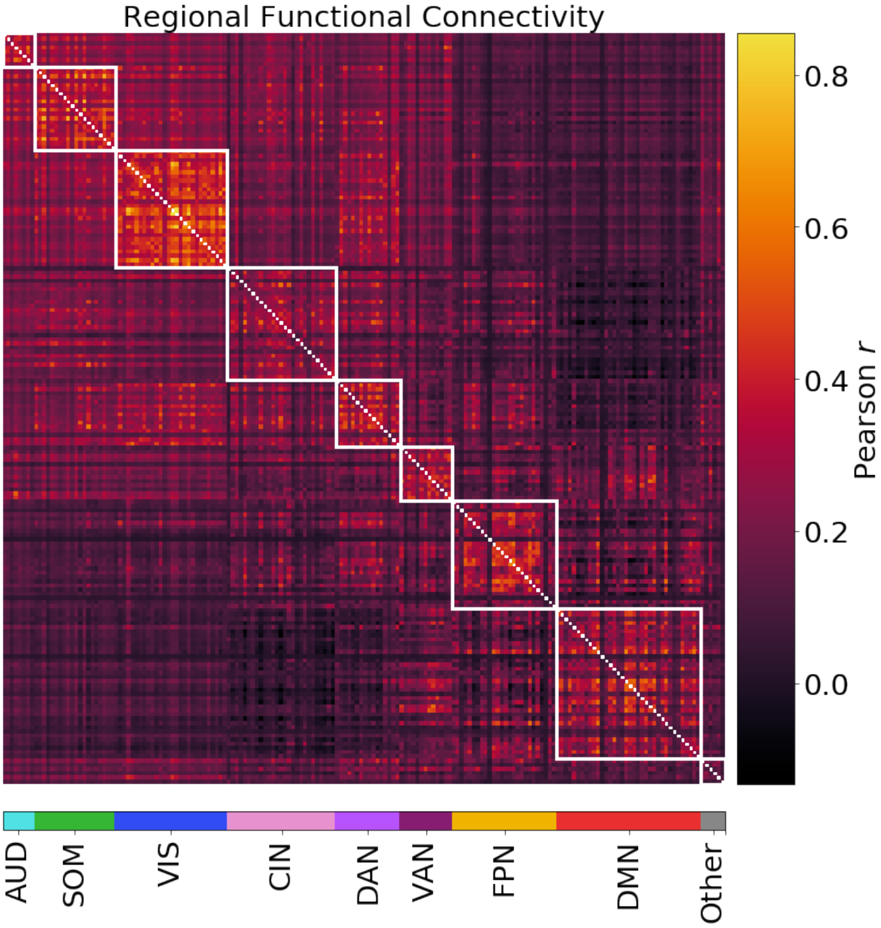


Figure 9.7: Regional functional connectivity matrix of group-averaged subject-level correlation matrices. White blocks corresponds to regions belonging to the same functional network, as defined by [16].

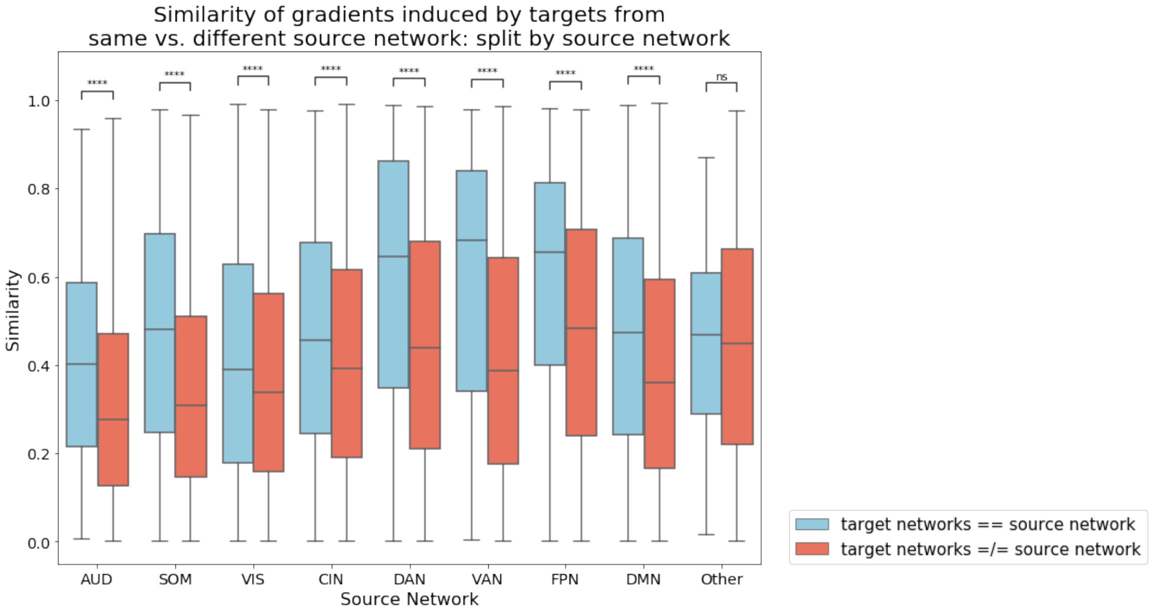


Figure 9.8: Similarities of the first target-specific gradients, delineated by the source region functional network. Mann-Whitney U-Test test statistic p-values shown at the top of network. **** = p < 0.0001, **ns** = not significant.

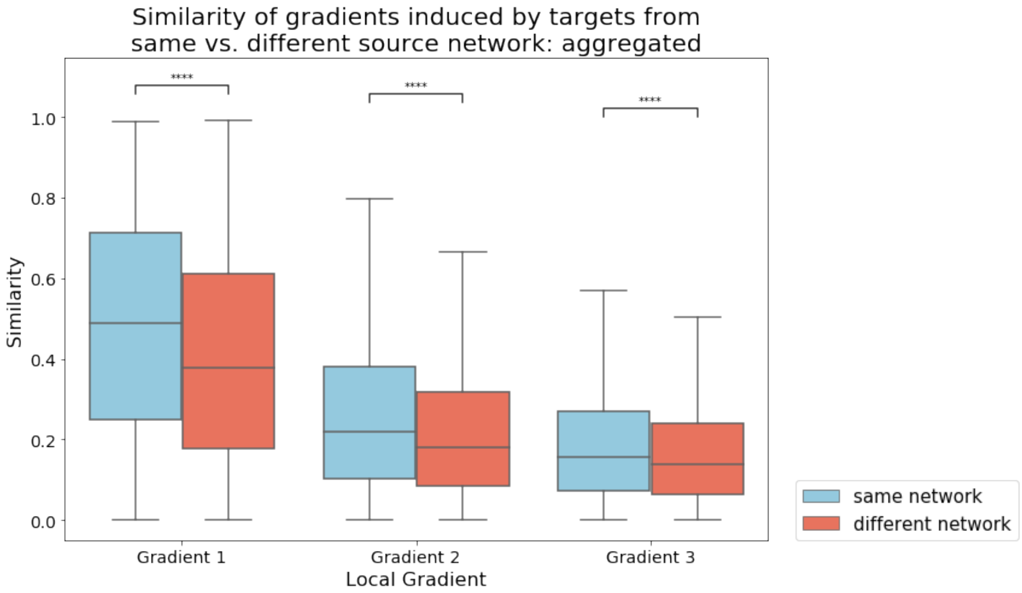


Figure 9.9: Similarity of induced target-specific gradients, when the source network is the same or different from that of both target areas.

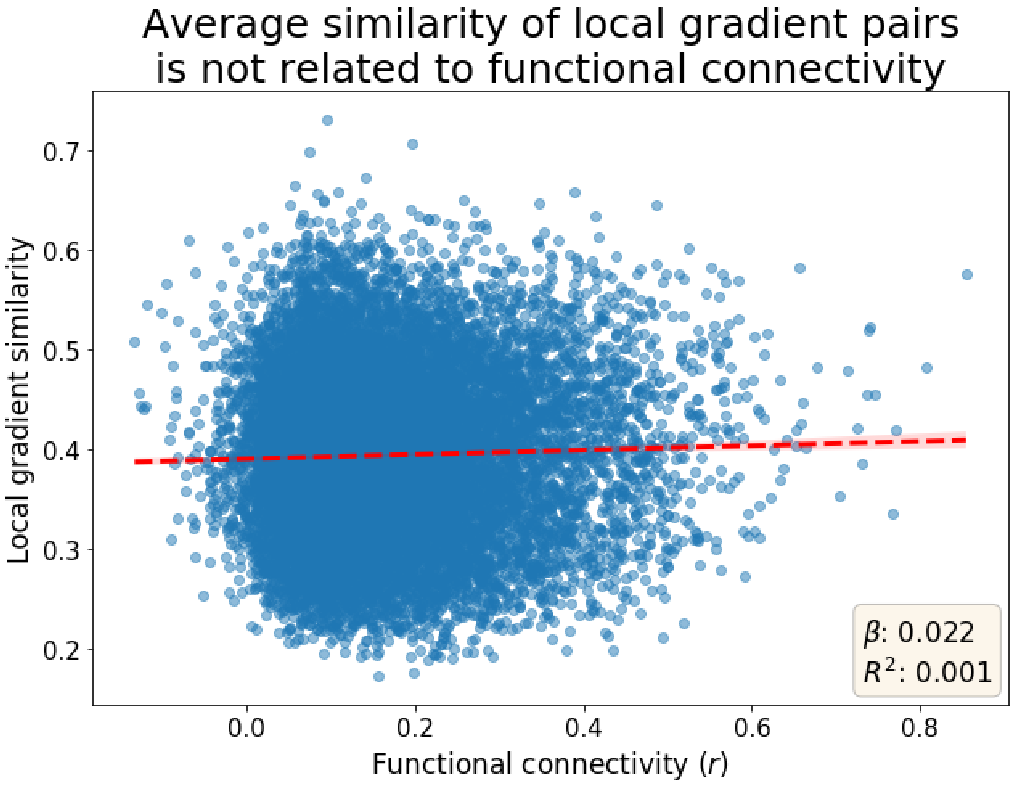


Figure 9.10: Regression of similarities between target-specific gradients, on the functional connectivity strength between the two target areas.

9.3 Chapter 4

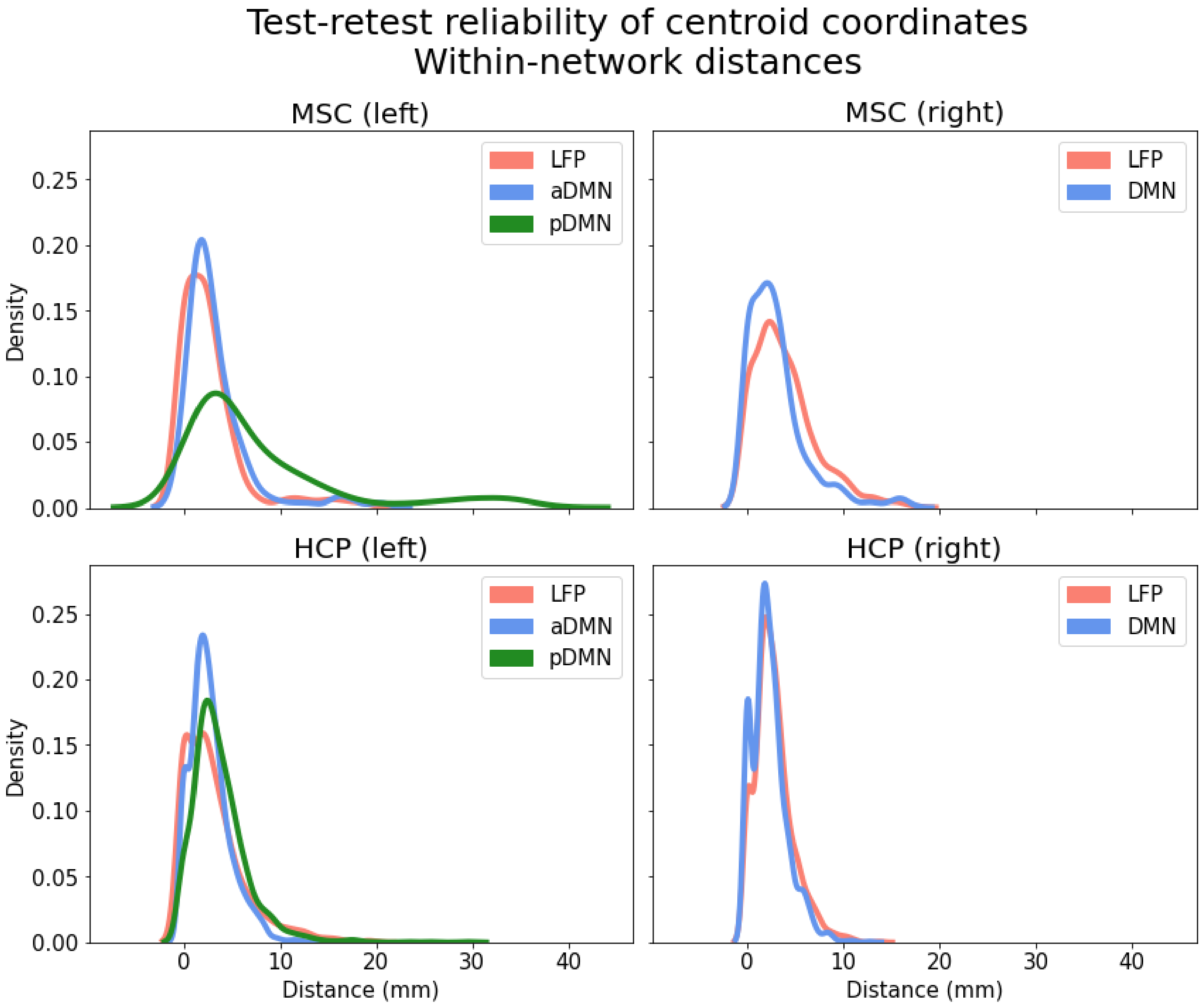


Figure 9.11: Within-network distances for the HCP and MSC control datasets. Results are averaged across 100 HCP subjects (4 sessions per subject) and 10 MSC subjects (10 sessions per subject).

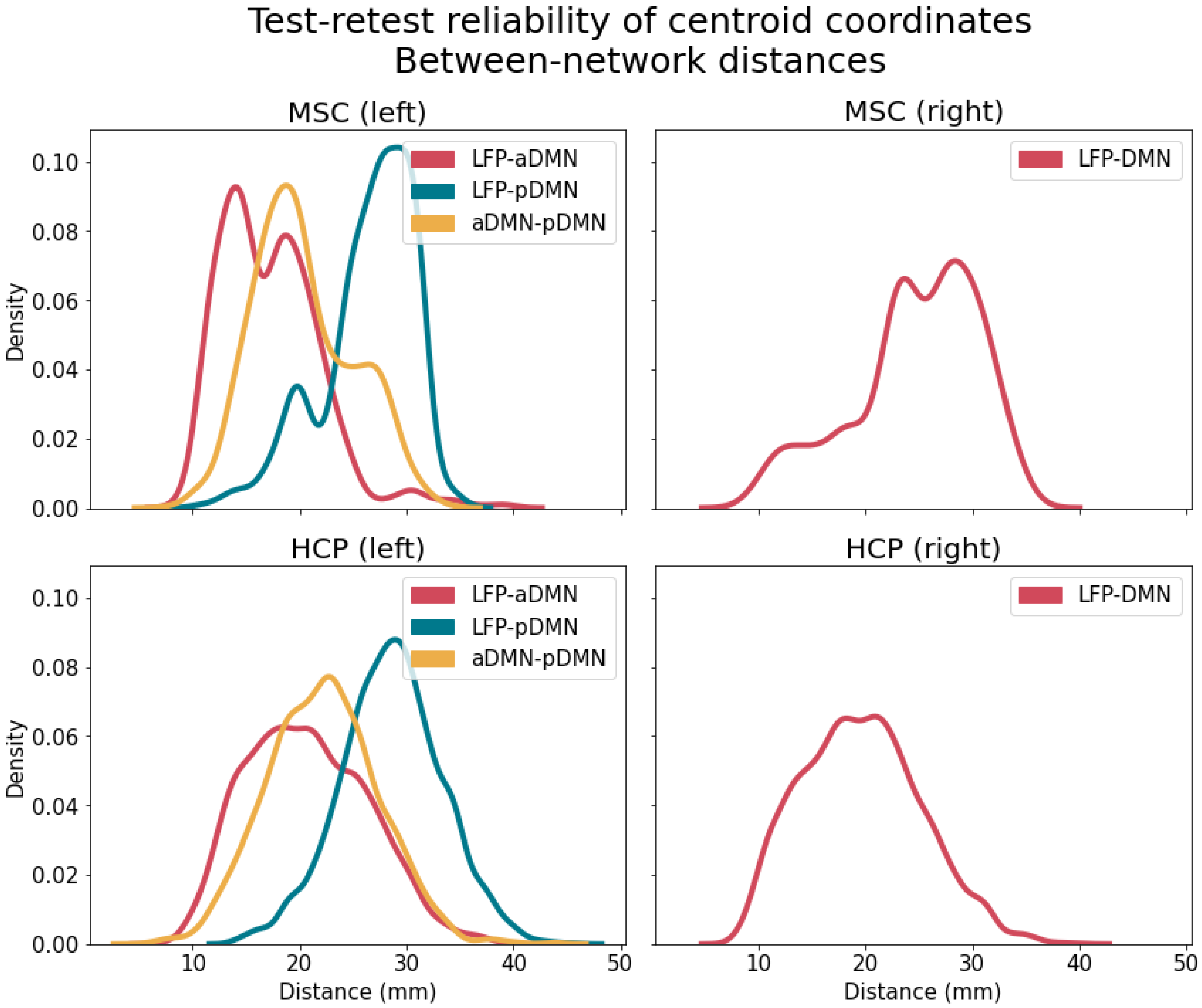


Figure 9.12: Between-network distances for the HCP and MSC control datasets. Results are averaged across 100 HCP subjects (4 sessions per subject) and 10 MSC subjects (10 sessions per subject).

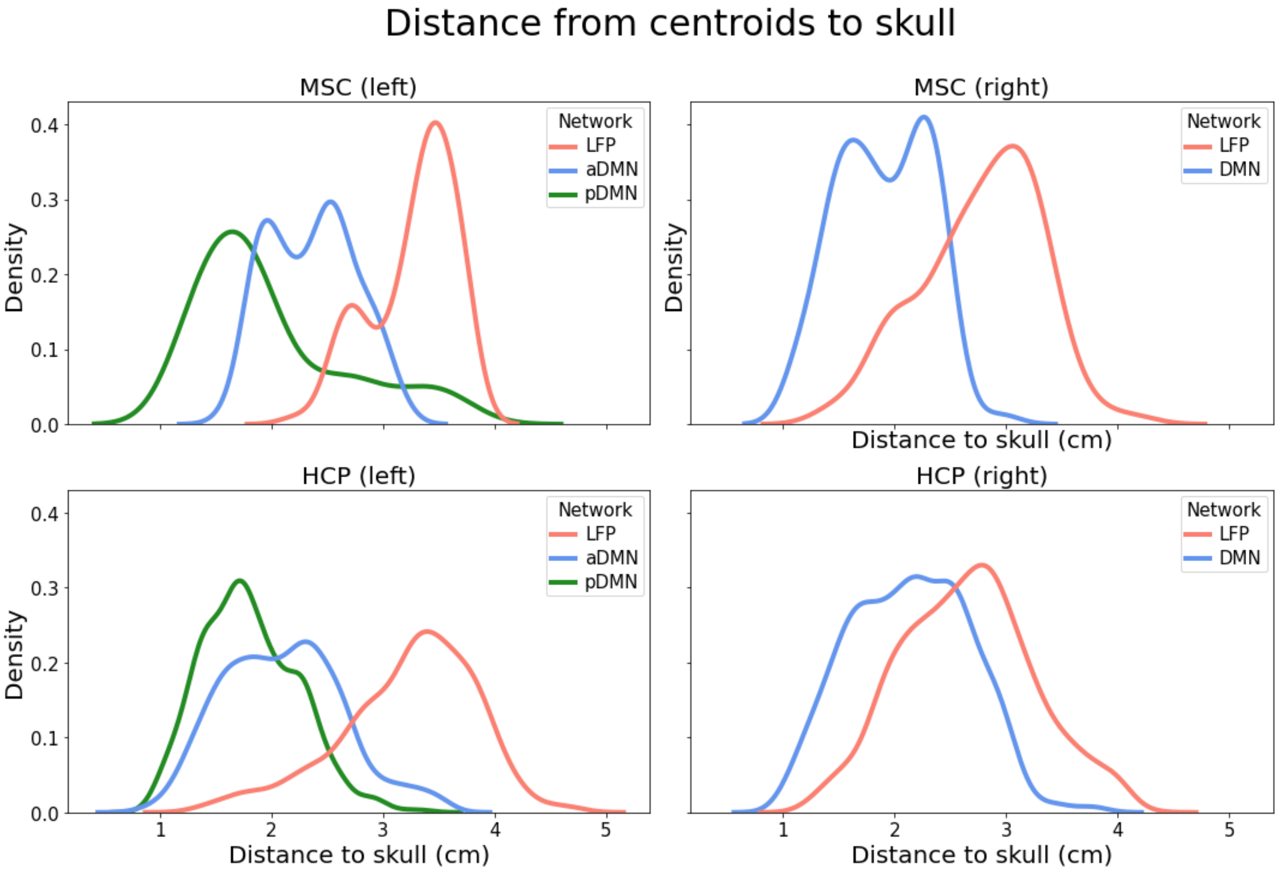


Figure 9.13: Distance to skull for each network centroid for both the HCP and MSC datasets. Results are averaged across 100 HCP subjects (4 sessions per subject) and 10 MSC subjects (10 sessions per subject).

9.4 Chapter 5

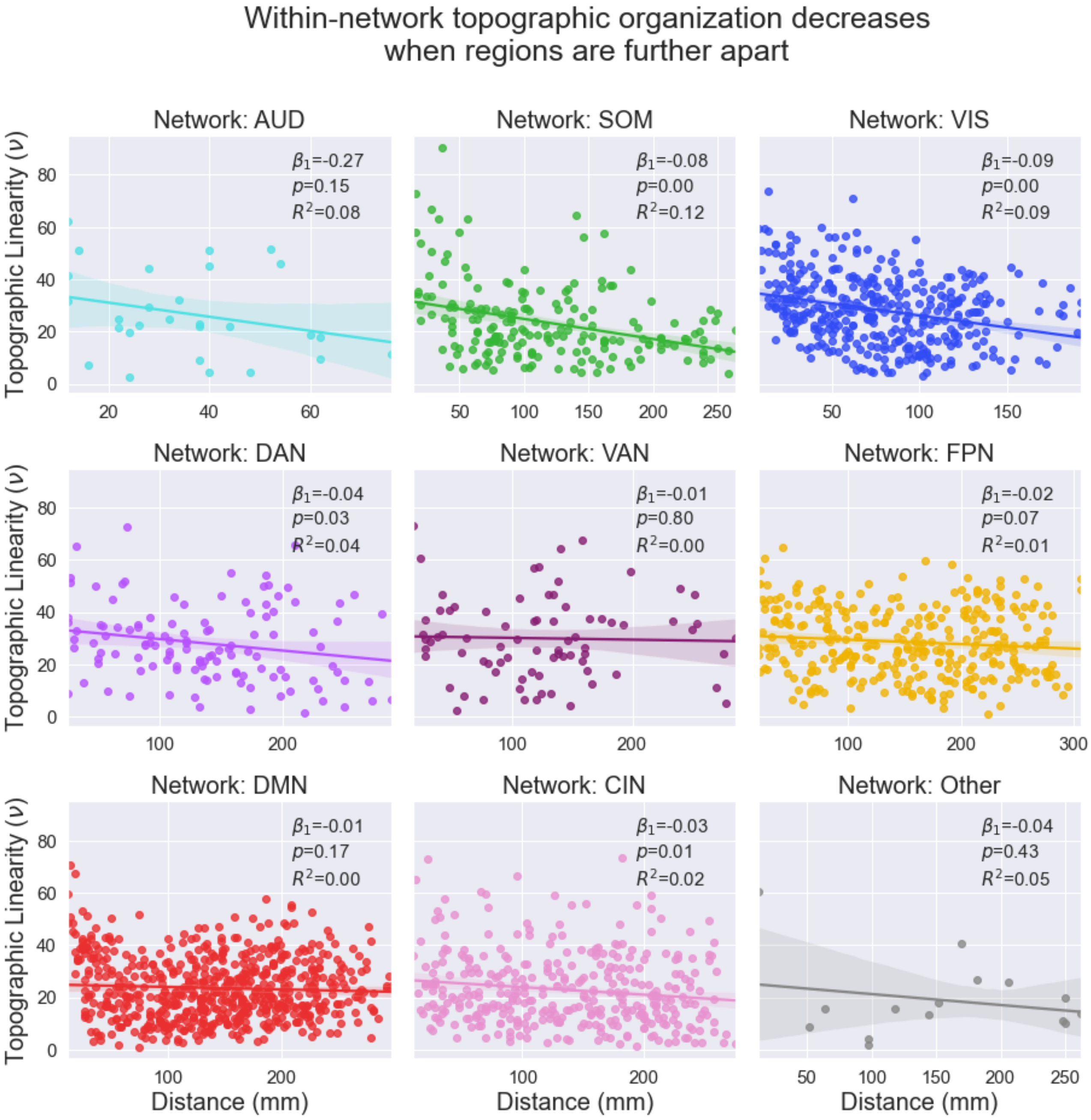


Figure 9.14: Topographic organization between region pairs is only weakly related to geodesic distance between regions, when considering regions from the same functional network.

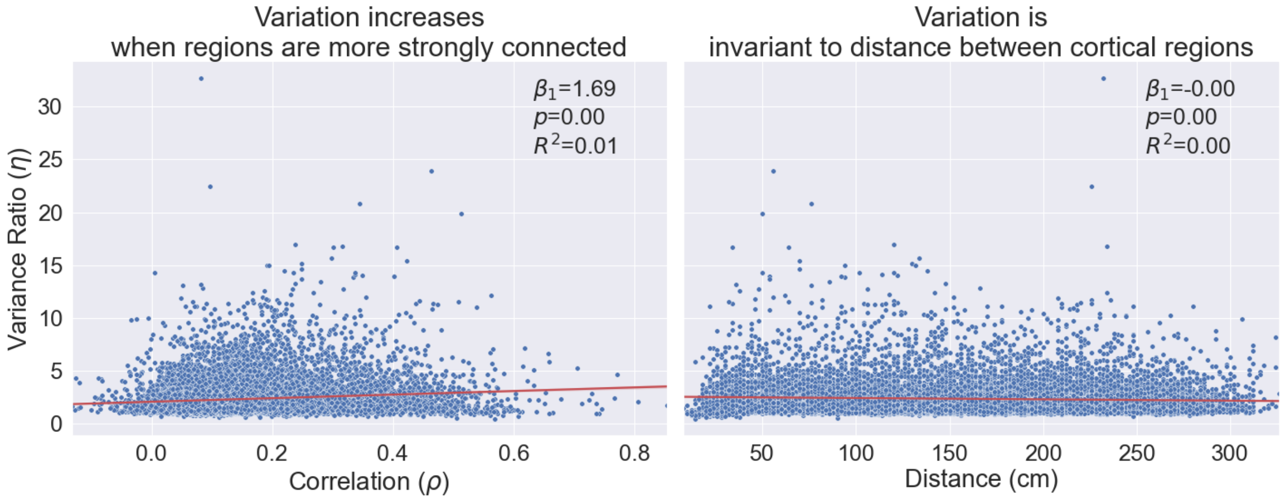


Figure 9.15: Pairwise variance is unrelated to both functional connectivity strength and geodesic distance at the region level.

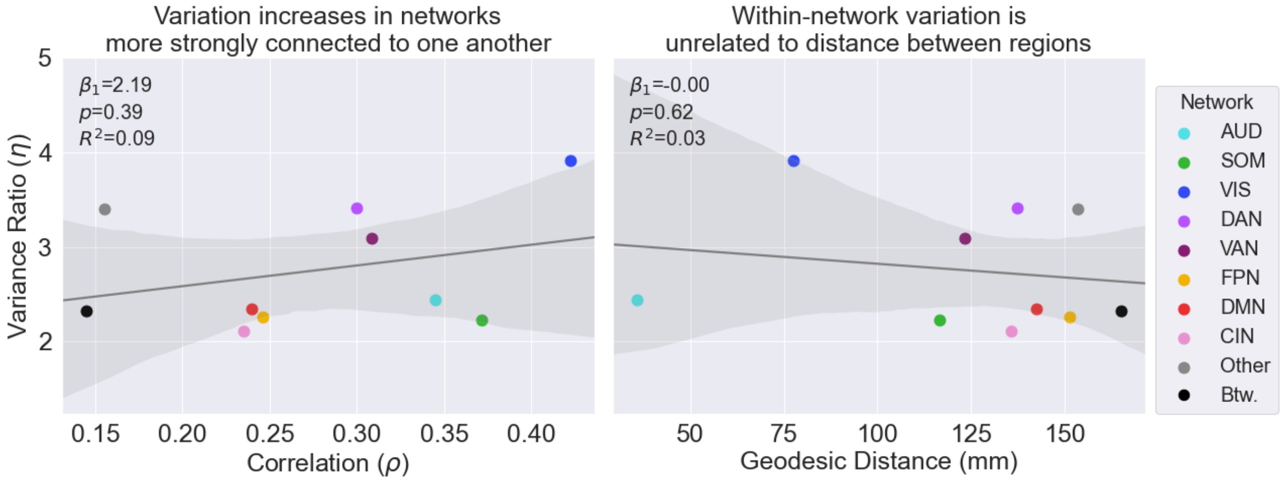


Figure 9.16: Pairwise variance is unrelated to both functional connectivity strength and geodesic distance at the network level.

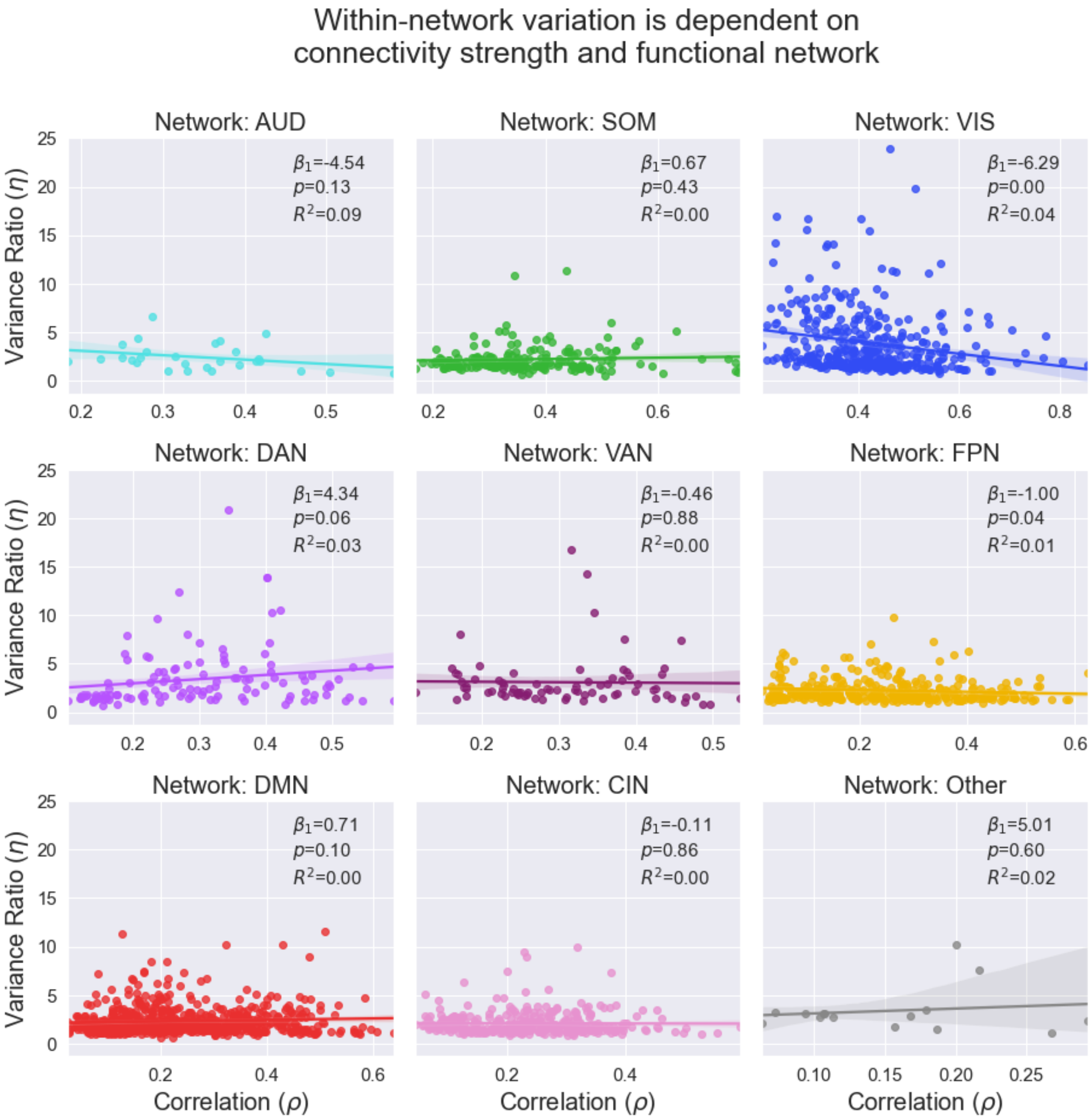


Figure 9.17: Pairwise variance is unrelated to functional connectivity within networks.

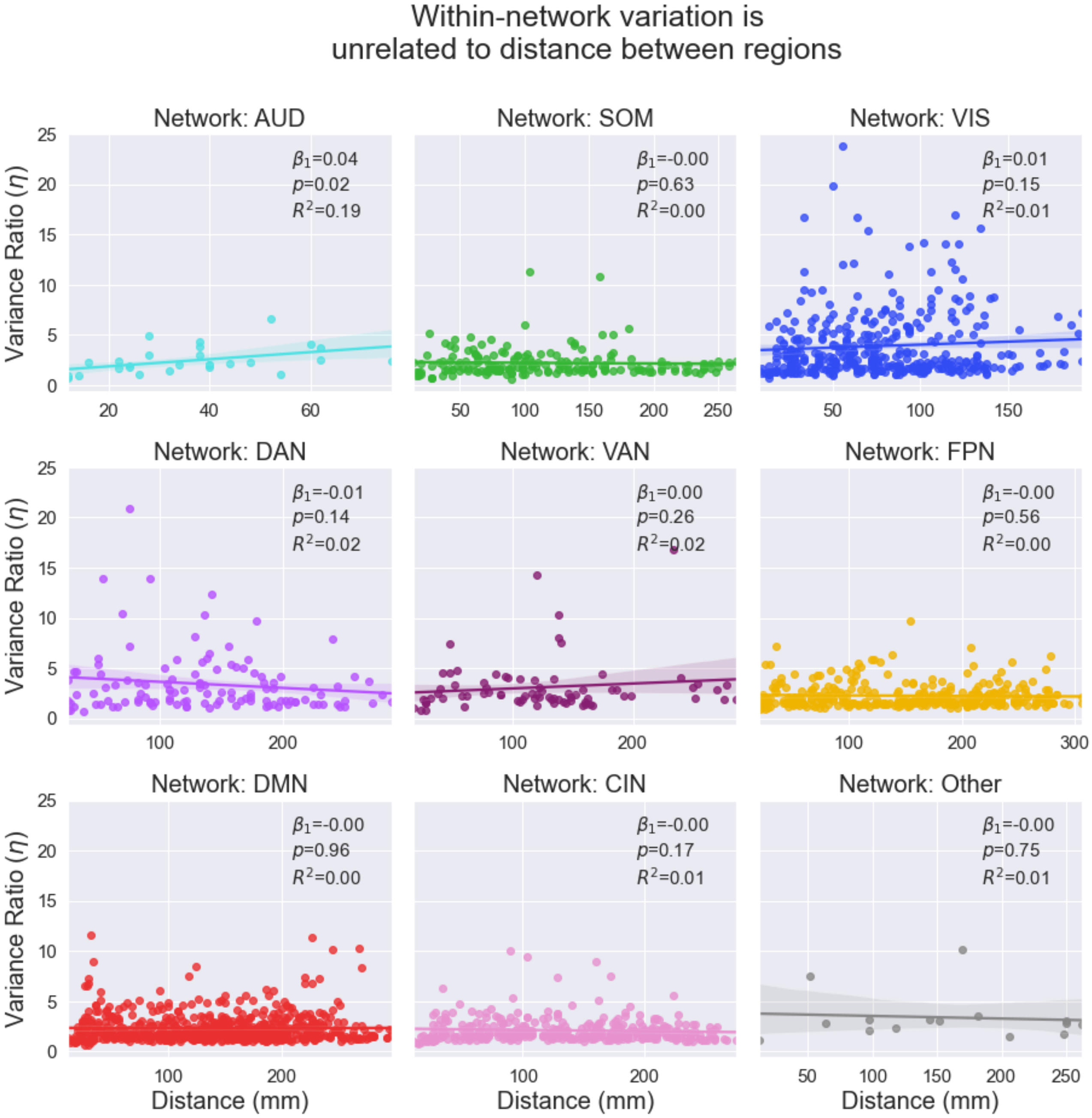


Figure 9.18: Pairwise variance is unrelated to geodesic distance within networks.

Chapter 10

Software

10.1 Code: parcellearning

Here, we provide some instructions for using the **parcellearning** package developed for work presented in Chapter 4, and is available for download here. This package uses a Pytorch backend, and a Deep Graph Library (DGL) frontend.

parcellearning is used for training graph neural networks for the purpose of node classification. The training data consists of subject-level graphs, represented by cortical surface meshes, where graph nodes are vertices in the surface mesh, and graph edges represent local edges between vertices on the cortical surface mesh. Graph nodes are also characterized by independent variables ("features"), which are used to predict a discrete dependent variable ("label") on a node-by-node basis.

10.1.1 Installation

This package can be installed using `pip` as

```
git clone https://github.com/kristianeschenburg/parcellearning.git
cd  ./parcellearning
```

```
pip install .
```

The following examples assume that the downloaded code exists in a directory structure called "./parcellearning".

10.1.2 Building subject-level graphs

Subject-level data is represented using DGL graph objects. Node features and surface mesh data are provided as file paths in a JSON file with a given structure, correspond to the raw data, and are used to create a single-graph object storing this data. The JSON structure is represented as follows (using example file names):

```json
{
    "features": {
        "curvature": "100307.L.curvature.32k_fs_LR.shape.gii",
        "euclidean": "100307.L.midthickness.32k_fs_LR.csv",
        "myelin": "100307.L.MyelinMap.32k_fs_LR.func.gii",
        "regionalized": "100307.L.aparc.a2009s.
  rfMRI_REST1_LR_Atlas_hp2000_clean.Mean.CrossCorr.csv",
        "spectral": "100307.L.midthickness.spectrum.LSA.CPD.csv",
        "sulcal": "100307.L.sulc.32k_fs_LR.shape.gii",
        "thickness": "100307.L.corrThickness.32k_fs_LR.shape.gii"
    },
    "label": "100307.L.CorticalAreas.fixed.32k_fs_LR.label.gii",
    "medial_wall": "L.MedialWall.func.gii",
    "output": "100307.L.rfMRI_REST1_LR_Atlas_hp2000_clean.bin",
    "subject": "100307",
    "surface": "100307.L.midthickness.32k_fs_LR.surf.gii"
}
```

This example structure assumes that all files exist in the *current* working directory, so it should be modified so that the paths point to the correct files. There are 6 required keys:

1. **features**: dictionary of feature names and file paths used as independent variables. The DGL graph object stores numeric features using the same feature names.
2. **label**: path to dependent variable that will be predicted
3. **medial_wall**: contains indices corresponding to medial wall. These indices are excluded during model training.
4. **output**: name of output DGL graph object
5. **subject**: subject ID
6. **surface**: surface file. DGL uses the adjacency structure of this graph

We can then construct the DGL graph object using a script called `build_graph.py` in `./utilities/` as:

```
python build_graph.py --feature-file=${json_file}$ --dType="training"
```

10.1.3 Model schema files

Once we have built the training and validation level graphs, we can train our model. Because `parcellearning` provides different types of network architectures, we also need to provide the training code with a schema file that organizes the required model parameters. This schema file is also saved as a JSON file. We must first aggregate the subject-level training and validation graphs into large training and validation graph objects, which can be done using the DGL function `dgl.load_graphs` to load the subject level graphs, appending these graphs as a list, and then saving the list using the DGL function `dgl.save_graphs`. Here is an example schema file which uses the graph attention network (GAT) architecture:

```
{
  "model": "GAT",
  "data": {
    "out": "./MSMAll/DR/",
    "training": "training.L.MSMAll.quarter.d15.DR.bin",
    "validation": "validation.L.MSMAll.quarter.d15.DR.bin",
    "testing": "",
    "label_table": "L.Glasser.32k_fs_LR.labeltable"
  },
  "variable_parameters": {
    "features": [
      "ica",
      "spectral",
      "curvature",
      "myelin",
      "sulcal",
      "thickness"
    ],
    "response": "label"
  },
  "model_parameters": {
    "in_dim": 22,
    "num_classes": 181,
    "num_heads": [
```

```
      4,
      4
    ],
    "num_layers": 6,
    "num_hidden": 32,
    "residual": "False",
    "feat_drop": 0.1,
    "attn_drop": 0.1,
    "negative_slope": 0.2,
    "activation": "torch.nn.functional.leaky_relu",
    "allow_zero_in_degree": "True"
  },
  "optimizer_parameters": {
    "lr": 0.005,
    "weight_decay": 5e-4
  },
  "training_parameters": {
    "epochs": 1500,
    "n_batch": 40,
    "verbose": "False"
  },
  "stopping_parameters": {
    "mode": "lower",
    "patience": 150
  }
}
```

Once we have built this schema file and the necessary training and validation graphs, we can then train our model. Using the `train.py` file in the corresponding architecture directory, we train the model as follows:

```
python train.py --schema-file=${schema_file}$
```

Files are saved to the directory specified by the "out" schema file key. Continuing with the current schema example, we would generate four files:

1. `gat_parameters.json` : copy of schema file
2. `GAT.earlystop.Loss.pt` : model that optimizes validation loss – we use this for prediction
3. GAT.pt : model that runs to training completion
4. `performance.GAT.json` : training performance metrics including training and validation accuracy and loss

10.1.4 Prediction

Once the model has been trained, we can use the predict.py script in `./utilities/` to predict the label and accuracy metrics of the test graph.

```
python predict.py --schema-file=${schema_file}$ \
                  --graph=${test_graph}$ \
                  --output=${output}$ \
                  --masked \
                  --compute-accuracy
```

where the `--masked` and `--compute-accuracy` options are flags that 1) mask the output using a previously computed spatial prior and 2) compute the accuracy with respect to a ground truth label. We can only use the `--compute-accuracy` flag if the graph object has a "label" key. If we omit these flags, the predicted label is not masked, and accuracy is not computed.

10.2 Code: fieldmodel

fieldmodel is a package for fitting densities over scalar fields that are distributed over meshes, and is an implementation of the connective field modeling methodology described in [22]. We developed this code for work described in Chapter 6. The code is available for download here, with a brief demonstration of the code provided here.

Given a scalar field assigning, let's say, elevation values to latitude-longitude coordinates of Mount Rainier, we could use **fieldmodel** to fit a Gaussian distribution to this elevation profile. The output of this fitting procedure would be 2 parameters: a mean parameter (centered on the true summit coordinates) and a sigma parameter, a measure of how quickly the elevation profile decays as we move away from the summit.

10.2.1 Installation

This package can be installed using `pip` as

```
git clone https://github.com/kristianeschenburg/fieldmodel.git
cd  ./fieldmodel
pip install .
```

10.2.2 Example use

Example data is provided in the `./data/` directory. Example usage is as follows:

```
import numpy as np
import scipy.io as sio
```

```

# X coordinates for plotting
tx_file = '../data/target.X.mat'
tx = sio.loadmat(tx_file)['x'].squeeze()
# Y coordinates for plotting
ty_file = '../data/target.Y.mat'
ty = sio.loadmat(ty_file)['y'].squeeze()

# geodesic distance matrix between vertices / nodes of mesh
dist_file = '../data/distance.mat'
dist = sio.loadmat(dist_file)['apsp']
# scalar field
field_file = '../data/scalar_field.mat'
field = sio.loadmat(field_file)['field'].squeeze()
```

Above, we have loaded the coordinates of the nodes in the mesh, the distance matrix (in this case, geodesic, though Euclidean is also acceptable) characterizing the shortest path distances between all pairs of nodes, and the scalar field to which we'll fit the density.

fieldmodel first identifies local maxima in the scalar field. We provide the parameter `peak_size`, which is the minimum distance required between local maxima. The larger the distance, the fewer local maxima we consider. `hood_size` sets the neighborhood search space around each identified local maxima – the larger this value, the larger the search radius. The parameter `r` is an initial "guess" of the density parameter σ, while `metric` indicates which cost function we use.

```
from fieldmodel import GeodesicFieldModel as GFM
# instantiate field model
G = GFM.FieldModel(r=10, amplitude=False, peak_size=15,
                   hood_size=20, verbose=False, metric='pearson')
```

```
# fit field model
G.fit(data=field, distances=dist, x=tx, y=ty)
# plot results of fitting procedure
G.plot(field='pdf')
```

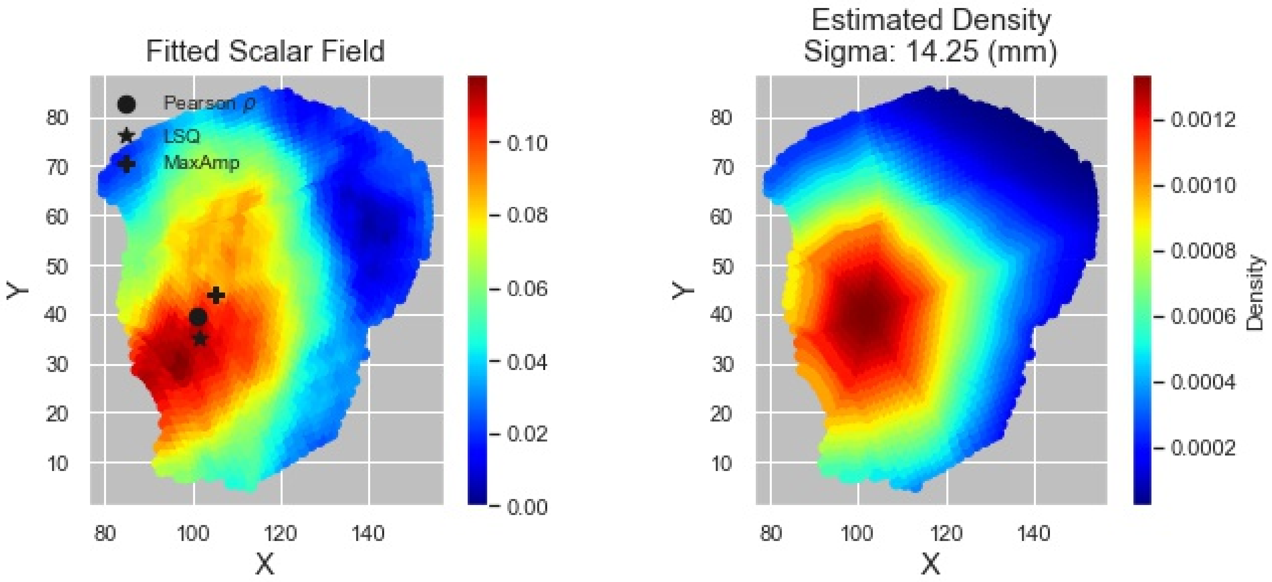


Figure 10.1: Example result of fitting a connective field model. We identify the density peaks computed using three different cost functions, displayed over the initial scalar field: "Pearson", "LSQ", and "MaxAmp". The final density fitted over the scalar field is shown on the right.

References

[1] M. H. Herzog and A. M. Clarke, "Why vision is not both hierarchical and feedforward," *Frontiers in Computational Neuroscience*, vol. 8, p. 135, 10 2014.

[2] C. Paquola, J. Seidlitz, O. Benkarim, J. Royer, P. Klimes, R. A. Bethlehem, S. Lariviere, R. V. d. Wael, J. A. Hall, B. Frauscher, J. Smallwood, and B. C. Bernhardt, "The cortical wiring scheme of hierarchical information processing," tech. rep., McGill University, 5 2020.

[3] K. V. Haak, A. F. Marquand, and C. F. Beckmann, "Connectopic mapping with resting-state fMRI," *NeuroImage*, vol. 170, pp. 83–94, 4 2018.

[4] N. Palomero-Gallagher, F. Hoffstaedter, H. Mohlberg, S. B. Eickhoff, K. Amunts, and K. Zilles, "Human Pregenual Anterior Cingulate Cortex: Structural, Functional, and Connectional Heterogeneity," *Cerebral Cortex (New York, NY)*, vol. 29, p. 2552, 6 2019.

[5] S. Caspers, A. Schleicher, M. Bacha-Trams, N. Palomero-Gallagher, K. Amunts, and K. Zilles, "Organization of the Human Inferior Parietal Lobule Based on Receptor Architectonics," *Cereb Cortex*, vol. 23, pp. 615–628, 3 2013.

[6] G. S. Wig, T. O. Laumann, and S. E. Petersen, "An approach for parcellating human cortical areas using resting-state correlations," *Neuroimage*, vol. 93 Pt 2, pp. 276–291, 2014.

[7] T. Blumensath, S. Jbabdi, M. F. Glasser, D. C. Van Essen, K. Ugurbil, T. E. Behrens, and S. M. Smith, “Spatially constrained hierarchical parcellation of the brain with resting-state fMRI,” *Neuroimage*, vol. 76, pp. 313–324, 2013.

[8] K. Amunts, H. Mohlberg, S. Bludau, and K. Zilles, “Julich-Brain: A 3D probabilistic atlas of the human brain’s cytoarchitecture,” *Science*, vol. 369, pp. 988–992, 8 2020.

[9] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, S. M. Smith, and D. C. Van Essen, “A multi-modal parcellation of human cerebral cortex,” *Nature*, vol. 536, no. 7615, pp. 171–178, 2016.

[10] I. Aganj, C. Lenglet, G. Sapiro, E. Yacoub, K. Ugurbil, and N. Harel, “Reconstruction of the Orientation Distribution Function in Single and Multiple Shell Q-Ball Imaging within Constant Solid Angle,” *Magn Reson Med*, vol. 64, no. 2, pp. 554–566, 2010.

[11] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. R. Fischl, J. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. Van Essen, M. Jenkinson, and W.-M. U.-M. H. C. P. Consortium, “The minimal preprocessing pipelines for the Human Connectome Project,” *Neuroimage*, vol. 80, pp. 105–124, 2013.

[12] H. Lombaert, L. Grady, J. R. Polimeni, F. Cheriet, and A. A. Martinos, “FOCUSR: feature oriented correspondence using spectral regularization–a method for precise surface matching,” *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 9, pp. 2143–2160, 2013.

[13] T. S. Coalson, D. C. Van Essen, and M. F. Glasser, “The impact of traditional neuroimaging methods on the spatial localization of cortical areas,” *PNAS*, vol. 115, no. 27, 2018.

[14] E. C. Robinson, S. Jbabdi, M. F. Glasser, J. Andersson, G. C. Burgess, M. P. Harms, S. M. Smith, D. C. Van Essen, and M. Jenkinson, “MSM: a new flexible framework for Multimodal Surface Matching,” *Neuroimage*, vol. 100, pp. 414–426, 2014.

[15] E. C. Robinson, K. Garcia, M. F. Glasser, Z. Chen, T. S. Coalson, A. Makropoulos, J. Bozek, R. Wright, A. Schuh, M. Webster, J. Hutter, A. Price, L. Cordero Grande, E. Hughes, N. Tusor, P. V. Bayly, D. C. Van Essen, S. M. Smith, A. David Edwards, J. Hajnal, M. Jenkinson, B. Glocker, D. Rueckert, and S. Lukes Hospital, “Multimodal Surface Matching with Higher-Order Smoothness Constraints,” *Neuroimage*, vol. 167, pp. 453–465, 2018.

[16] M. Demirtas, J. B. Burt, M. Helmer, J. L. Ji, B. D. Adkinson, M. F. Glasser, D. C. Van Essen, S. N. Sotiropoulos, A. Anticevic, and J. D. Murray, “Hierarchical Heterogeneity across Human Cortex Shapes Large-Scale Neural Dynamics,” *Neuron*, vol. 101, pp. 1181–1194, 3 2019.

[17] R. Vos de Wael, O. Benkarim, C. Paquola, S. Lariviere, J. Royer, S. Tavakol, T. Xu, S.-J. Hong, G. Langs, S. Valk, B. Misic, M. Milham, D. S. Margulies, J. Smallwood, and B. C. Bernhardt, “BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets.,” *Communications biology*, vol. 3, p. 103, 3 2020.

[18] D. S. Margulies, S. S. Ghosh, A. Goulas, M. Falkiewicz, J. M. Huntenburg, G. Langs, G. Bezgin, S. B. Eickhoff, F. X. Castellanos, M. Petrides, E. Jefferies, and J. Smallwood, “Situating the default-mode network along a principal gradient of macroscale cortical organization,” *PNAS*, vol. 113, no. 44, 2016.

[19] M. Assem, M. F. Glasser, D. C. Van Essen, and J. Duncan, “A Domain-General Cognitive Core Defined in Multimodally Parcellated Human Cortex,” *Cerebral Cortex*, vol. 00, pp. 1–20, 2020.

[20] M.-M. Mesulam, “From sensation to cognition,” *Brain*, vol. 121, pp. 1013–1052, 1998.

[21] Q. Wang and A. Burkhalter, “Area map of mouse visual cortex,” *Journal of Comparative Neurology*, vol. 502, pp. 339–357, 5 2007.

[22] K. V. Haak, J. Winawer, B. M. Harvey, R. Renken, S. O. Dumoulin, B. A. Wandell, and F. W. Cornelissen, “Connective field modeling,” *NeuroImage*, vol. 66, pp. 376–384, 2 2013.

[23] J. F. O’Rawe and H. C. Leung, “Topographic Mapping as a Basic Principle of Functional Organization for Visual and Prefrontal Functional Connectivity,” *eNeuro*, vol. 7, 1 2020.

[24] J. M. Huntenburg, P.-L. Bazin, A. Goulas, C. L. Tardif, A. Villringer, and D. S. Margulies, “A Systematic Relationship Between Functional Connectivity and Intracortical Myelin in the Human Cerebral Cortex,” *Cerebral Cortex*, vol. 27, pp. 981–997, 2017.

[25] J. B. Burt, M. Demirtas, W. J. Eckner, N. M. Navejar, J. Lisa Ji, W. J. Martin, A. Bernacchia, A. Anticevic, and J. D. Murray, “Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography,” *Nature Neuroscience*, vol. 21, pp. 1251–1259, 2018.

[26] E. M. Gordon, T. O. Laumann, B. Adeyemo, J. F. Huckins, W. M. Kelley, and S. E. Petersen, “Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations,” *Cereb Cortex*, vol. 26, no. 1, pp. 288–303, 2016.

[27] K. Brodmann and L. J. Gary, *Brodmann’s localisation in the cerebral cortexa: the principles of comparative localisation in the cerebral cortex based on cytoarchitectonics.* New York, NY: Springer, 2006.

[28] C. Vogt and O. Vogt, “Allgemeine Ergebnisse unserer Hirnforschung,” 1919.

[29] V. B. Mountcastle, P. W. Davies, and A. I. Berman, “Response properties of neurons of cat somatic sensory cortex to peripheral stimuli,” tech. rep., Johns Hopkins University, 1956.

[30] D. H. Hubel and T. N. Wiesel, “Laminar and columnar distribution of geniculo-cortical fibers in the macaque monkey,” *Journal of Comparative Neurology*, vol. 146, pp. 421–450, 12 1972.

[31] N. Palomero-Gallagher and K. Zilles, “Cortical layers: Cyto-, myelo-, receptor- and synaptic architecture in human cortical areas,” *NeuroImage*, vol. 197, pp. 716–741, 8 2019.

[32] S. Caspers, S. Geyer, A. Schleicher, H. Mohlberg, K. Amunts, K. Zilles, and O. Vogt, “The human inferior parietal cortex: Cytoarchitectonic parcellation and interindividual variability,” *Neuroimage*, vol. 33, pp. 430–448, 2006.

[33] R. S. Desikan, F. Segonne, B. R. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany, “An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest,” *Neuroimage*, vol. 31, pp. 968–980, 2006.

[34] C. Destrieux, B. R. Fischl, A. M. Dale, and E. Halgren, “A sulcal depth-based anatomical parcellation of the cerebral cortex,” *Human Brain Mapping*, 2009.

[35] D. C. Van Essen, M. F. Glasser, V. E. DC, and G. MF, “Parcellating Cerebral Cortex: How Invasive Animal Studies Inform Noninvasive Mapmaking in Humans,” 8 2018.

[36] V. B. Mountcastle, “The columnar organization of the neocortex,” tech. rep., Johns Hopkins University, 1997.

[37] R. L. Tungaraza, S. H. Mehta, D. R. Haynor, and T. J. Grabowski, “Anatomically Informed Metrics for Connectivity-based Cortical Parcellation from Diffusion MRI,” *IEEE J Biomed Health Inform*, vol. 19, no. 4, pp. 1375–1383, 2015.

[38] S. Arslan, S. Parisot, and D. Rueckert, “Joint Spectral Decomposition for the Parcellation of the Human Cerebral Cortex Using Resting-State fMRI,” *Inf Process Med Imaging*, vol. 24, pp. 85–97, 2015.

[39] S. Parisot, S. Arslan, J. Passerat-Palmbach, W. M. Wells, and D. Rueckert, “Groupwise parcellation of the cortex through multi-scale spectral clustering,” *Neuroimage*, vol. 136, pp. 68–83, 2016.

[40] S. Parisot, S. Arslan, J. Passerat-Palmbach, W. M. Wells, and D. Rueckert, “Tractography-Driven Groupwise Multi-scale Parcellation of the Cortex,” *Inf Process Med Imaging*, vol. 24, pp. 600–612, 2015.

[41] J. Wang, L. Fan, Y. Zhang, Y. Liu, D. Jiang, Y. Zhang, C. Yu, and T. Jiang, “Tractography-based parcellation of the human left inferior parietal lobule,” *Neuroimage*, vol. 63, 2012.

[42] S. Jbabdi, M. W. Woolrich, and T. E. Behrens, “Multiple-subjects connectivity-based parcellation using hierarchical Dirichlet process mixture models,” *NeuroImage*, vol. 44, no. 2, pp. 373–384, 2009.

[43] C. Baldassano, D. M. Beck, and L. Fei-Fei, “Parcellating connectivity in spatial maps.,” *PeerJ*, vol. 3, no. 2, p. e784, 2015.

[44] K. Amunts and K. Zilles, “Primer Architectonic Mapping of the Human Brain beyond Brodmann,” *Neuron*, 2015.

[45] B. Fischl, M. I. Sereno, R. B. Tootell, and A. M. Dale, “High-resolution intersubject averaging and a coordinate system for the cortical surface,” *Human Brain Mapping*, vol. 8, no. 4, p. 272, 1999.

[46] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. B. Avants, M. C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey, “Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration,” *NeuroImage*, vol. 46, p. 786, 7 2009.

[47] N. J. Tustison and B. B. Avants, “Explicit B-spline regularization in diffeomorphic image registration,” *Frontiers in Neuroinformatics*, vol. 7, 12 2013.

[48] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, “Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain,” *Medical image analysis*, vol. 12, p. 26, 2 2008.

[49] M. Liu, A. Kitsch, S. Miller, V. Chau, K. Poskitt, F. Rousseau, D. Shaw, and C. Studholme, “Patch-based augmentation of Expectation-Maximization for brain MRI tissue segmentation at arbitrary age after premature birth,” *Neuroimage*, vol. 127, pp. 387–408, 2 2016.

[50] E. M. Gordon, T. O. Laumann, A. W. Gilmore, D. J. Newbold, D. J. Greene, J. J. Berg, M. Ortega, C. Hoyt-Drazen, C. Gratton, H. Sun, J. M. Hampton, R. S. Coalson, A. L. Nguyen, K. B. McDermott, J. S. Shimony, A. Z. Snyder, B. L. Schlaggar, S. E. Petersen, S. M. Nelson, and N. U. Dosenbach, “Precision Functional Mapping of Individual Human Brains,” *Neuron*, vol. 95, pp. 791–807, 8 2017.

[51] B. R. Fischl, A. Van Der Kouwe, C. Destrieux, E. Halgren, F. Segonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. N. Kennedy, V. S. Caviness, N. Makris, B. R.

Rosen, and A. M. Dale, “Automatically parcellating the human cerebral cortex,” *Cereb Cortex*, vol. 14, pp. 11–22, 2004.

[52] M. F. Glasser and D. C. van Essen, “Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI,” *Journal of Neuroscience*, vol. 31, pp. 11597–11616, 8 2011.

[53] C. F. Beckmann, M. DeLuca, J. T. Devlin, and S. M. Smith, “Investigations into resting-state connectivity using independent component analysis,” *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 360, no. 1457, pp. 1001–1013, 2005.

[54] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith, “Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers,” *NeuroImage*, vol. 90, pp. 449–468, 2014.

[55] S. Jbabdi, S. N. Sotiropoulos, A. M. Savio, M. Graña, and T. E. Behrens, “Model-based analysis of multi-shell diffusion MR data for tractography: How to get over fitting problems,” *Magn Reson Med*, vol. 68, no. 6, pp. 1846–1855, 2012.

[56] M. Descoteaux, R. Deriche, D. Le Bihan, J.-F. Mangin, and C. Poupon, “Multiple q-shell diffusion propagator imaging q,” *Med Image Anal*, vol. 15, pp. 603–621, 2010.

[57] T. E. Behrens, M. W. Woolrich, M. Jenkinson, H. Johansen-Berg, R. G. Nunes, S. Clare, P. M. Matthews, J. M. Brady, and S. M. Smith, “Characterization and propagation of uncertainty in diffusion-weighted MR imaging,” *Magnetic Resonance in Medicine*, vol. 50, pp. 1077–1088, 11 2003.

[58] T. E. Behrens, H. Johansen-Berg, S. Jbabdi, M. F. S. Rushworth, and M. W. Woolrich, “Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?,” *NeuroImage*, 2006.

[59] D. J. Felleman and D. C. Van Essen, “Distributed hierarchical processing in the primate cerebral cortex,” *Cereb Cortex*, vol. 1, no. 1, pp. 1–47, 1991.

[60] D. C. Van Essen, D. J. Felleman, E. A. DeYoe, J. Olavarria, and J. Knierim, “Modular and hierarchical organization of extrastriate visual cortex in the macaque monkey,” *Cold Spring Harb Symp Quant Biol*, vol. 55, pp. 679–696, 1990.

[61] W. H. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick, “Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex,” *Journal of Neuroscience*, vol. 17, pp. 2112–2127, 3 1997.

[62] J. C. Horton and D. L. Adams, “The cortical column: a structure without a function,” *Philos Trans R Soc Lond B Biol Sci*, vol. 360, pp. 837–862, 2005.

[63] B. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zollei, J. R. Polimeni, B. R. Fischl, H. Liu, R. L. Buckner, B. T. Thomas Yeo, and L. Zöllei, “The organization of the human cerebral cortex estimated by intrinsic functional connectivity,” *J Neurophysiol*, vol. 106, pp. 1125–1165, 9 2011.

[64] N. S. Gorbach, C. Schutte, C. Melzer, M. Goldau, O. Sujazow, J. Jitsev, T. Douglas, and M. Tittgemeyer, “Hierarchical information-based clustering for connectivity-based cortex parcellation,” *Front Neuroinform*, vol. 5, p. 18, 2011.

[65] B. R. Fischl, N. Rajendran, E. Busa, J. Augustinack, O. Hinds, B. T. Yeo, H. Mohlberg, K. Amunts, and K. Zilles, “Cortical folding patterns and predicting cytoarchitecture,” *Cereb Cortex*, vol. 18, pp. 1973–1980, 2008.

[66] H. Lombaert, A. Criminisi, and N. Ayache, “Spectral Forests: Learning of Surface Data, Application to Cortical Parcellation,” in *MICCAI*, 2015.

[67] V. B. Mountcastle, “Modality and topographic properties of single neurons of cat’s somatic sensory cortex,” tech. rep., Johns Hopkins University, 1956.

[68] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” tech. rep., University of Freiburg, 2015.

[69] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering,” in *Neural Information Processing Systems*, 2016.

[70] T. N. Kipf and M. Welling, “Semi-Supervised Classification With Graph Convolutional Networks,” tech. rep., University of Amsterdam, 2016.

[71] W. L. Hamilton, R. Ying, and J. Leskovec, “Representation Learning on Graphs: Methods and Applications,” in *IEEE Data Engineering*, 2017.

[72] H. Zeng, H. Zhou, A. Srivastava, R. Kannan, and V. Prasanna, “GraphSAINT: Graph Sampling Based Inductive Learning Method,” tech. rep., University of Southern California, 2020.

[73] T. N. Kipf and M. Welling, “Variational Graph Auto-Encoders,” tech. rep., University of Amsterdam, 2016.

[74] Q. Zhao, E. Adeli, N. Honnorat, T. Leng, and K. M. Pohl, “Variational AutoEncoder For Regression: Application to Brain Aging Analysis,” tech. rep., Stanford, 2019.

[75] A. Vaswani, G. Brain, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need,” tech. rep., Google, 2017.

[76] K. Wagstyl, S. Larocque, G. Cucurull, C. Lepage, J. P. Cohen, S. Bludau, N. Palomero-Gallagher, L. B. Lewis, T. Funck, H. Spitzer, T. Dickscheid, P. C. Fletcher, A. Romero, K. Zilles, K. Amunts, Y. Bengio, and A. C. Evans, "BigBrain 3D atlas of cortical layers: Cortical and laminar thickness gradients diverge in sensory and motor cortices," *PLoS Biology*, vol. 18, no. 4, pp. 1–21, 2020.

[77] R. Petersen, P. Aisen, L. Beckett, M. Donohue, A. Gamst, D. Harvey, C. Jack, W. Jagust, L. Shaw, A. Toga, J. Trojanowski, and M. Weiner, "Alzheimer's Disease Neuroimaging Initiative (ADNI) Clinical characterization," *Neurology*, pp. 201–209, 2010.

[78] D. J. Hagler, S. Hatton, M. D. Cornejo, C. Makowski, D. A. Fair, A. S. Dick, M. T. Sutherland, B. J. Casey, D. M. Barch, M. P. Harms, R. Watts, J. M. Bjork, H. P. Garavan, L. Hilmer, C. J. Pung, C. S. Sicat, J. Kuperman, H. Bartsch, F. Xue, M. M. Heitzeg, A. R. Laird, T. T. Trinh, R. Gonzalez, S. F. Tapert, M. C. Riedel, L. M. Squeglia, L. W. Hyde, M. D. Rosenberg, E. A. Earl, K. D. Howlett, F. C. Baker, M. Soules, J. Diaz, O. Ruiz De Leon, W. K. Thompson, M. C. Neale, M. Herting, E. R. Sowell, R. P. Alvarez, S. W. Hawes, M. Sanchez, J. Bodurka, F. J. Breslin, A. S. Morris, M. P. Paulus, W. K. Simmons, J. R. Polimeni, A. Van Der Kouwe, A. S. Nencka, K. M. Gray, C. Pierpaoli, J. A. Matochik, A. Noronha, W. M. Aklin, K. Conway, M. Glantz, E. Hoffman, R. Little, M. Lopez, V. Pariyadath, S. Weiss, D. L. Wolff-Hughes, R. Delcarmen-Wiggins, S. W. Feldstein Ewing, O. Miranda-Dominguez, B. J. Nagel, A. J. Perrone, D. T. Sturgeon, A. Goldstone, A. Pfefferbaum, K. M. Pohl, D. Prouty, K. Uban, S. Y. Bookheimer, M. Dapretto, A. Galvan, K. Bagot, J. Giedd, M. Alejandra Infante, J. Jacobus, K. Patrick, P. D. Shilling, R. S. Desikan, Y. Li, L. Sugrue, M. T. Banich, N. Friedman, J. K. Hewitt, C. Hopfer, J. Sakai, J. Tanabe, L. B. Cottler, S. J. Nixon, L. Chang, C. Cloak, T. Ernst, G. Reeves, D. N. Kennedy, S. Heeringa, S. Peltier, J. Schulenberg, C. Sripada, R. A. Zucker, W. G. Iacono,

M. Luciana, F. J. Calabro, D. B. Clark, D. A. Lewis, B. Luna, C. Schirda, T. Brima, J. J. Foxe, E. G. Freedman, D. W. Mruzek, M. J. Mason, R. Huber, E. Mcglade, A. Prescot, P. F. Renshaw, D. A. Yurgelun-Todd, N. A. Allgaier, J. A. Dumas, M. Ivanova, A. Potter, P. Florsheim, C. Larson, K. Lisdahl, M. E. Charness, B. Fuemmeler, J. M. Hettema, H. H. Maes, J. Steinberg, A. P. Anokhin, P. Glaser, A. C. Heath, P. A. Madden, A. Baskin-Sommers, R. Todd Constable, S. J. Grant, G. J. Dowling, S. A. Brown, T. L. Jernigan, and A. M. Dale, “Image processing and analysis methods for the Adolescent Brain Cognitive Development Study,” *NeuroImage*, vol. 202, 2019.

[79] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lì, and Y. Bengio, “Graph Attention Networks,” in *International Conference on Learning Representation*, 2018.

[80] K. Xu, C. Li, Y. Tian, T. Sonobe, K.-I. Kawarabayashi, and S. Jegelka, “Representation Learning on Graphs with Jumping Knowledge Networks,” tech. rep., MIT, 2018.

[81] C. D. Hacker, T. O. Laumann, N. P. Szrama, A. Baldassarre, A. Z. Snyder, E. C. Leuthardt, and M. Corbetta, “Resting State Network Estimation in Individual Subjects,” *Neuroimage*, vol. 15, no. 82, 2013.

[82] A. J. Asman and B. A. Landman, “Non-local statistical label fusion for multi-atlas segmentation,” *Medical Image Analysis*, vol. 17, pp. 194–208, 2 2012.

[83] A. J. Asman and B. A. Landman, “Hierarchical Performance Estimation in the Statistical Label Fusion Framework,” *Med Image Anal*, vol. 18, no. 7, pp. 1070–1081, 2014.

[84] K. Gopinath, C. Desrosiers, and H. Lombaert, “Graph Convolutions on Spectral Embeddings for Cortical Surface Parcellation,” *Medical Image Analysis*, vol. 54, pp. 297–305, 2019.

[85] G. Cucurull, K. Wagstyl, A. Casanova, P. Velickovic, E. Jakobsen, M. Drozdzal, A. Romero, A. C. Evans, and Y. Bengio, “Convolutional neural networks for mesh-based parcellation of the cerebral cortex,” in *Medical Imaging with Deep Learning*, 2018.

[86] G. Wang, R. Ying, J. Huang, and J. Leskovec, “Improving Graph Attention Networks with Large Margin-based Constraints,” tech. rep., Stanford University, Mountain View, 2019.

[87] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.

[88] S. M. Smith, A. Hyvärinen, G. Varoquaux, K. L. Miller, and C. F. Beckmann, “Group-PCA for very large fMRI datasets,” *Neuroimage*, vol. 101, p. 738, 11 2014.

[89] A. F. Mejia, M. B. Nebel, Y. Wang, B. S. Caffo, and Y. Guo, “Template Independent Component Analysis: Targeted and Reliable Estimation of Subject-level Brain Networks using Big Data Population Priors,” tech. rep., Indiana University, 2019.

[90] M. Wang, D. Zheng, Z. Ye, Q. Gan, M. Li, X. Song, J. Zhou, C. Ma, L. Yu, Y. Gai, T. Xiao, T. He, G. Karypis, J. Li, Z. Zhang, W. Services, A. Shanghai, and A. I. Lab, “Deep graph library: a graph-centric, highly-performant package for graph neural networks,” tech. rep., New York University, 2020.

[91] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. Yeo, “Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI,” *Cerebral Cortex*, vol. 28, pp. 3095–3114, 2018.

[92] C. J. Bajada, R. L. Jackson, H. A. Haroon, H. Azadbakht, G. J. Parker, M. A. Lambon Ralph, and L. L. Cloutman, “A graded tractographic parcellation of the temporal lobe,” *NeuroImage*, vol. 155, pp. 503–512, 7 2017.

[93] R. Gala, A. Budzillo, F. Baftizadeh, J. Miller, N. Gouwens, A. Arkhipov, G. Murphy, B. Tasic, H. Zeng, M. Hawrylycz, U. Sümbül, and U. Samp, “Consistent cross-modal identification of cortical neurons with coupled autoencoders,” *Nature Computational Science*, vol. 1, pp. 120–127, 2 2021.

[94] L. Q. Uddin, K. Supekar, H. Amin, E. Rykhlevskaia, D. A. Nguyen, M. D. Greicius, and V. Menon, “Dissociable connectivity within human angular gyrus and intraparietal sulcus: evidence from functional and structural connectivity,” *Cereb Cortex*, vol. 20, no. 11, pp. 2636–2646, 2010.

[95] E. DuPre, T. Salo, Z. Ahmed, P. A. Bandettini, K. L. Bottenhorn, C. Caballero-Gaudes, L. T. Dowdle, J. Gonzalez-Castillo, S. Heunis, P. Kundu, A. R. Laird, R. Markello, C. J. Markiewicz, S. Moia, I. Staden, J. B. Teves, E. Uruñuela, M. Vaziri-Pashkam, K. Whitaker, and D. A. Handwerker, “TE-dependent analysis of multi-echo fMRI with tedana,” *Journal of Open Source Software*, vol. 6, p. 3669, 10 2021.

[96] P. Kundu, S. J. Inati, J. W. Evans, W. M. Luh, and P. A. Bandettini, “Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI,” *NeuroImage*, vol. 60, pp. 1759–1770, 4 2012.

[97] P. Kundu, N. D. Brenowitz, V. Voon, Y. Worbe, P. E. Vértes, S. J. Inati, Z. S. Saad, P. A. Bandettini, and E. T. Bullmore, “Integrated strategy for improving functional connectivity mapping using multiecho fMRI,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 16187–16192, 10 2013.

[98] C. J. Lynch, J. D. Power, M. A. Scult, M. Dubin, F. M. Gunning, and C. Liston, “Rapid Precision Functional Mapping of Individuals Using Multi-Echo fMRI,” *Cell Reports*, vol. 33, p. 108540, 12 2020.

[99] P. E. Roland and K. Zilles, “Structural divisions and functional fields in the human cerebral cortex 1,” *Brain Research Reviews*, vol. 26, pp. 87–105, 5 1998.

[100] S. O. Dumoulin and B. A. Wandell, “Population receptive field estimates in human visual cortex,” *NeuroImage*, vol. 39, no. 2, pp. 647–660, 2007.

[101] L. Wang, R. E. B. Mruczek, M. J. Arcaro, and S. Kastner, “Probabilistic maps of visual topography in human cortex,” *Cerebral Cortex*, vol. 25, no. 10, pp. 3911–3931, 2015.

[102] A. M. Leaver and J. P. Rauschecker, “Functional topography of human auditory cortex,” *Journal of Neuroscience*, vol. 36, pp. 1416–1428, 1 2016.

[103] C. Paquola, R. Vos De Wael, K. Wagstyl, R. A. Bethlehem, S.-J. Hong, J. Seidlitz, E. T. Bullmore, A. C. Evans, B. Misic, D. S. Margulies, J. Smallwood, and B. C. Bernhardt, “Microstructural and functional gradients are increasingly dissociated in transmodal cortices,” *PLoS Biology*, vol. 17, p. 28, 5 2019.

[104] J. M. Huntenburg, P.-L. Bazin, and D. S. Margulies, “Large-Scale Gradients in Human Cortical Organization,” *Trends in Cognitive Sciences*, vol. 22, pp. 21–31, 2018.

[105] G. H. Patel, D. Kaplan Michael, L. H. Snyder, D. M. Kaplan, and L. H. Snyder, “Topographic organization in the brain: searching for general principles,” *Trends in Cognitive Sciences*, vol. 18, pp. 351–363, 7 2014.

[106] X. Guell, J. D. Schmahmann, J. D. Gabrieli, and S. S. Ghosh, “Functional gradients of the cerebellum,” *eLife*, vol. 7, 8 2018.

[107] G. B. Freches, K. V. Haak, K. L. Bryant, M. Schurz, C. F. Beckmann, and R. B. Mars, “Principles of temporal association cortex organisation as revealed by connectivity gradients,” *Brain Struct Funct*, vol. 1, p. 3, 2020.

[108] Y. Tian and A. Zalesky, “Characterizing the functional connectivity diversity of the insula cortex: Subregions, diversity curves and behavior,” *NeuroImage*, vol. 183, pp. 716–733, 12 2018.

[109] J. Royer, C. Paquola, S. Lariviere, R. Vos de Wael, S. Tavakol, A. J. Lowe, O. Benkarim, A. C. Evans, D. Bzdok, J. Smallwood, B. Frauscher, B. C. Bernhardt, R. V. D. Wael, A. J. Lowe, O. Benkarim, A. C. Evans, D. Bzdok, J. Smallwood, B. Frauscher, and B. C. Bernhardt, “Myeloarchitecture Gradients In The Human Insula: Histological Underpinnings And Association To Intrinsic Functional Connectivity,” *NeuroImage*, vol. 216, p. 116859, 8 2020.

[110] C. Paquola, O. Benkarim, J. DeKraker, S. Lariviere, S. Frässle, J. Royer, S. Tavakol, S. Valk, A. Bernasconi, N. Bernasconi, A. Khan, A. C. Evans, A. Razi, J. Smallwood, and B. C. Bernhardt, “Convergence of cortical types and functional motifs in the mesiotemporal lobe,” *bioRxiv*, p. 2020.06.12.148643, 6 2020.

[111] C. Toro-Serey, S. M. Tobyne, and J. T. McGuire, “Spectral partitioning identifies individual heterogeneity in the functional network topography of ventral and anterior medial prefrontal cortex,” *NeuroImage*, vol. 205, p. 116305, 1 2020.

[112] M. J. McKeown, L. K. Hansen, and T. J. Sejnowsk, “Independent component analysis of functional MRI: What is signal and what is noise?,” *Current Opinion in Neurobiology*, vol. 13, no. 5, pp. 620–629, 2003.

[113] S. M. Smith, C. F. Beckmann, J. Andersson, E. J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D. A. Feinberg, L. Griffanti, M. P. Harms, M. Kelly, T. O. Laumann, K. L. Miller, S. Moeller, S. E. Petersen, J. Power, G. Salimi-Khorshidi, A. Z. Snyder, A. T. Vu, M. W. Woolrich, J. Xu, E. Yacoub, K. Ugurbil, D. C. Van Essen, M. F. Glasser, A. T. Vu, H. Consortium, W. U.-M. H. C. P. Consortium, and K. Uğurbil, “Resting-state fMRI in the Human Connectome Project,” *Neuroimage*, vol. 80, pp. 144–168, 2013.

[114] B. W. Brunton, L. A. Johnson, J. G. Ojemann, and N. J. Kutz, “Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition,” *J Neurosci Methods*, vol. 258, pp. 1–15, 1 2016.

[115] J. Casorso, X. Kong, W. Chi, D. Van De Ville, B. T. Yeo, and R. Liégeois, “Dynamic mode decomposition of resting-state and task fMRI,” *Dynamic mode decomposition of resting-state and task fMRI*, p. 431718, 2018.

[116] J. M. Kunert-Graf, K. M. Eschenburg, D. J. Galas, N. J. Kutz, S. D. Rane, and B. W. Brunton, “Extracting Reproducible Time-Resolved Resting State Networks Using Dynamic Mode Decomposition,” *Frontiers in Computational Neuroscience*, vol. 13, p. 75, 10 2019.

[117] J. Zhang, O. Abiose, Y. Katsumi, A. Touroutoglou, B. C. Dickerson, and L. F. Barrett, “Intrinsic Functional Connectivity is Organized as Three Interdependent Gradients,” *Scientific Reports*, vol. 9, pp. 1–14, 12 2019.

[118] K. Glomb, M. L. Kringelbach, G. Deco, P. Hagmann, J. Pearson, and S. Atasoy, “Functional harmonics reveal multi-dimensional basis functions underlying cortical organization,” tech. rep., University of Lausanne, 7 2019.

[119] S. Atasoy, I. Donnelly, and J. Pearson, “Human brain networks function in connectome-specific harmonic waves,” *Nature Communications*, vol. 7, pp. 1–10, 1 2016.

[120] V. Satopää, J. Albrecht, D. Irwin, and B. Raghavan, “Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior,” tech. rep., Williams College, Williamstown, MA, 2011.

[121] F. Battiston, J. Guillon, M. Chavez, V. Latora, and F. De Vico Fallani, “Multiplex core-periphery organization of the human connectome,” *J. R. Soc. Interface*, vol. 15, 2018.

[122] M. Daianu, E. L. Dennis, N. Jahanshad, T. M. Nir, A. W. Toga, C. R. Jack Jr., M. W. Weiner, P. M. Thompson, and I. the Alzheimer's Disease Neuroimaging, "Alzheimer's Disease Disrupts Rich Club Organization in Brain Connectivity Networks," *Proc IEEE Int Symp Biomed Imaging*, pp. 266–269, 2013.

[123] A. V. Utevsky, D. V. Smith, and S. A. Huettel, "Precuneus is a functional core of the default-mode network," *Journal of Neuroscience*, vol. 34, no. 3, pp. 932–940, 2014.

[124] M. Daianu, N. Jahanshad, T. M. Nir, C. R. Jack Jr., M. W. Weiner, M. A. Bernstein, P. M. Thompson, and I. Alzheimer's Disease Neuroimaging, "Rich club analysis in the Alzheimer's disease connectome reveals a relatively undisturbed structural core network," *Hum Brain Mapp*, vol. 36, no. 8, pp. 3087–3103, 2015.

[125] E. M. Gordon, C. J. Lynch, C. Gratton, T. O. Laumann, A. W. Gilmore, D. J. Greene, M. Ortega, A. L. Nguyen, B. L. Schlaggar, S. E. Petersen, N. U. Dosenbach, and S. M. Nelson, "Three Distinct Sets of Connector Hubs Integrate Human Brain Function," *Cell Reports*, vol. 24, pp. 1687–1695, 8 2018.

[126] D. A. Dawson, J. Lam, L. B. Lewis, F. Carbonell, J. D. Mendola, and A. Shmuel, "Partial Correlation-Based Retinotopically Organized Resting-State Functional Connectivity Within and between Areas of the Visual Cortex Reflects More Than Cortical Distance," *Brain Connectivity*, vol. 6, no. 1, 2016.

[127] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann, "Predicting human resting-state functional connectivity from structural connectivity," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 2035–2040, 2 2009.

[128] J. C. Griffis, A. S. Elkhetali, W. K. Burge, R. H. Chen, A. D. Bowman, J. P. Szaflarski, and K. M. Visscher, "Retinotopic patterns of functional connectivity between V1 and large-scale brain networks during resting fixation," *NeuroImage*, vol. 146, 2017.

[129] M. D. Greicius, K. Supekar, V. Menon, and R. F. Dougherty, “Resting-State Functional Connectivity Reflects Structural Connectivity in the Default Mode Network,” *Cerebral Cortex (New York, NY)*, vol. 19, p. 72, 1 2009.

[130] G. N. Ngo, K. V. Haak, C. F. Beckmann, and R. S. Menon, “Mesoscale hierarchical organization of primary somatosensory cortex captured by resting-state-fMRI in humans,” *NeuroImage*, vol. 235, no. December 2020, p. 118031, 2021.

[131] R. Rajimehr, J. C. Young, and R. B. Tootell, “An anterior temporal face patch in human cortex, predicted by macaque maps,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 6, 2009.

[132] R. D. Markello, G. Shafiei, C. Tremblay, R. B. Postuma, A. Dagher, and B. Misic, “Multimodal phenotypic axes of Parkinson’s disease,” *npj Parkinson’s Disease 2021 7:1*, vol. 7, pp. 1–12, 1 2021.