Spring 2021

Latent Factor Analysis of High-Dimensional Brain Imaging Data

Siyuan Gao
Yale University Graduate School of Arts and Sciences, rickygao213@gmail.com

Abstract

Latent Factor Analysis Of High-dimensional Brain Imaging Data

Siyuan Gao

2021

Recent advances in neuroimaging study, especially functional magnetic resonance imaging (fMRI), has become an important tool in understanding the human brain. Human cognitive functions can be mapped with the brain functional organization through the high-resolution fMRI scans. However, the high-dimensional data with the increasing number of scanning tasks and subjects pose a challenge to existing methods that aren't optimized for high-dimensional imaging data. In this thesis, I develop advanced data-driven methods to help utilize more available sources of information in order to reveal more robust brain-behavior relationship. In the first chapter, I provide an overview of the current related research in fMRI and my contributions to the field. In the second chapter, I propose two extensions to the connectome-based predictive modeling (CPM) method that is able to combine multiple connectomes when building predictive models. The two extensions are both able to generate higher prediction accuracy than using the single connectome or the simple average of multiple connectomes, suggesting the advantage of incorporating multiple sources of information in predictive modeling. In the third chapter, I improve CPM from the target behavioral measure's perspective. I propose another two extensions for CPM that are able to combine multiple available behavioral measures into a composite measure for CPM to predict. The derived composite measures are shown to be predicted more accurately than any other single behavioral measure, suggesting a more robust brain-behavior relationship. In the fourth chapter, I propose a nonlinear dimensionality reduction framework to embed fMRI data from multiple tasks into a low-dimensional space. This framework helps reveal the common brain state in the multiple available tasks while also help discover the differences among these tasks. The results also provide valuable insights into the variable prediction performances based on connectomes from different tasks. In

the fifth chapter, I propose another hyerbolic geometry-based brain graph edge embedding framework. The framework is based on Poincaré embedding and is able to more accurately represent edges in the brain graph in a low-dimensional space than traditional Euclidean geometry-based embedding. Utilizing the embedding, we are able to cluster edges of the brain graph into disjoint clusters. The edge clusters can then be used to define overlapping brain networks and the derived metrics like network overlapping number can be used to investigate functional flexibility of each brain region. Overall, these work provide rich data-driven methods that help understand the brain-behavioral relationship through predictive modeling and low-dimensional data representation.

Latent Factor Analysis Of High-dimensional Brain Imaging Data

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Siyuan Gao

Dissertation Director: Dustin Scheinost

June 2021

Contents

Acknowledgments

Graduate school is never easy for me, but I enjoyed every day of it. Coming from a mathematics major in college, I have never imagined my research will be 'studying the brain'. I still remembered the day when I was looking for papers to read for the Physical and Chemical Basis of Biosensing course, I found the super cool functional connectome fingerprinting paper that eventually lead me to the Connectivity Lab in MRRC. I feel so blessed that I found that paper, but more importantly, I found the perfect research group for me.

First and foremost, I would like to thank my advisor, Dustin Scheinost. I enjoy every conversation we had, either about research or about life. I am really grateful to have a mentor who can discuss with me how to make the for loop quicker in Matlab and also how to drive Mario kart faster on Nintendo. You are the true role model that I look up to and keep me motivated. Without you, I wouldn't enjoy doing research as much as a hobby. Thank you for making my 5 years of PhD journal so colorful!

Todd Constable, you are my first mentor that leads me to the field of functional connectivity. I can still remember the first day that I went to your office and chatted with you about this field that I had no experience with. I enjoyed your high-level conversation in the group meeting every Friday and the grilled steak in the group party every half a year. You showed me how to balance work and life as an already famous researcher and it will influence me throughout my life.

Gal Mishne, I am so grateful that I attended your talk that afternoon. I remember that I was so excited the whole talk and imagined what cool research I can do with diffusion maps. I am glad we did it. Thank you for keeping my passion for math. Math becomes so much more alive when I am discussing it with you than reading papers or textbooks. You also influenced me on how to make nice figures and writing good papers. My PhD will be so different without you as an important mentor.

Xenios Papademetris, don't know if you remembered, you are actually the first person I talked to in MRRC. I went to your office during campus visit before deciding which PhD program to

choose. I remembered we had long discussions about what does it mean to get a PhD. Time flies! I am really grateful to having you on my committee and thanks for everything I learned from you.

To all the other members of the MINDS Lab and Constable Lab, I cannot express how grateful I am with you as my family in New Haven. You really made my PhD a colorful journal that is not only filled with lonely science. It is amazing to stay in the lab for five years so I get to know initially Abigail Greene, Corey Horien, Eve Lake, Mehraveh Salehi, Xilin Shen, Stephanie Noble to later Alex Dufford, Alice Hahn, Angeliki Pollatou, Erin Yeagle, Hannah Peterson, Javid Dadashkarimi, Link Tejavibulya, Matt Rosenblatt, Max Rolison, Michael Farruggia, Qinghao Liang, Saloni Mehta, Sarah Lichenstein, Silvia Gini, Wei Dai, An Qu, David O'Connor, Dingjue Ji, Francesca Mandino, Javid Dadashkarimi, Kangjoo Lee, Wendy Luo. This also included the wonderful people that I met and collaborated in the school of Medicine, Daniel Barron, Sarah Yip and Katie Garrison, Kartiga Selvaganesan, Yonghyun Ha, Fuyuze Tokoglu, Mariela Rance. I am really grad we are building a big family in here and hope our path will come across in the future!

I would also like to thank all the professors and staff members that help me finish my PhD study smoothly at Yale, including Richard Carson, Steven Zucker, James Duncan, Ronald Coifman and GSAS, OISS, MRRC, BME staff. I would also like to thank Cameron Craddock for serving as my external reader.

Lastly, I would like to thank all my friends at Yale for making my out-of-lab life so fantastic-Xinxin Nie, Xuechen Zhou, Sheng Xu, Wenmian Hua, Mingrui Xu and the list goes on!

I would like to thank my parents Fengying Gu and Guocong Gao for all the support throughout the years. I can still feel the joy that we shared when I got admitted into the Ph.D. program and I cannot imagine finishing it without you two as my strongest supporters.

And finally I would like to thank my best friend, and my other part of life, Wenye, for all the moments we shared since college. You are my motivation for me to achieve everything and I am really looking forward to all the best things that will happen to us in the future.

List of Figures

Chapter 1

Introduction

1.1 Modeling the brain as a network

The human brain is a complex biological system that is capable of performing various functions. In order to successfully provide coherent control over the actions of human, hundred billions of neurons across a range of spatial and temporal scales co-activate in complex patterns. To link these large-scale neural activity patterns with the corresponding cognitive behaviors, proper data summarization tool is needed to prevent practitioners from being overwhelmed by the sheer amount of data.

Functional magnetic resonance imaging (fMRI) is a neuroimaging method to measure brain activity by detecting changes associated with blood flow. It can non-invasively investigate the function of the human brain in healthy and disease groups. Using fMRI under the resting state or during tasks, neural connections of the brain can be represented by using brain connectomes. As a mapping of the functional coherence of different brain regions, brain connectome measures the cross-correlation of the brain regions' time series. It has been used in revealing robust individual differences in patterns of neural activity that predict continuous behavioral measures and clinical symptoms (Yoshida et al. [2017], Shen et al. [2017], Poldrack et al. [2016c]). The predictive features derived from these models have helped researchers better understand the underlying

functional differences and thus hold great clinical values. While the majority of the literature has focused on using functional connectivity (FC) from only a single scanning condition (e.g., resting-state) to predict a single cognitive score of interest (e.g, the Penn Matrix Reasoning Test score), more data are available and remain unused or only used independently. There is ample evidence that different cognitive conditions amplify individual differences in FC in a distinct, complementary manner (Finn et al. [2017b], Geerligs et al. [2015]). Thus, methods that incorporate FC information from a spectrum of scanning conditions and cognitive measures into a single predictive model may represent the best performing and most generalizable methods for prediction of behavior from FC data.

As described above, the entire scan can be collapsed into a single FC that represents the average neural activity pattern. However, investigating the neural patterns in finer temporal scales such as quantifying moment-to-moment changes in brain activation or connectivity is also gaining attentions (Allen et al. [2014b], Monti et al. [2017b], Shine et al. [2019]). A main goal of these works is to find representative brain states—or distinct, repeatable patterns of brain activity or connectivity—as the reference to quantify these brain dynamics. Focusing on a few specific states operationalizes the characterization of brain dynamics into computational tractable problems. However, a raw 3D fMRI volume contains hundreds of thousands of voxels (e.g., $> 600k$ for the Human Connectome Project data) and for a single task design, we often have multiple different runs and participants as the fMRI Signal-To-Noise Ratio (SNR) can be limited. Compared with the static FC analyses that collapse the temporal dimension, dynamic fMRI analyses face more challenges in terms of the dimensionality as the temporal variation cannot be ignored. Thus to analyze these high- and multi-dimensional data, proper data-driven approach that reduces the data dimension is needed. To collapse information across different runs/participants, taking the Euclidean mean of fMRI time series or FCs within the same task design is the simplest approach to aggregate information. However, whether it is a proper approach is questionable due to the 'curse of dimensionality' (e.g., high-dimensional points are more uniformly distant from each other in

terms of Euclidean metrics). In order to reduce the dimensionality, linear dimensionality reduction approaches like principal component analysis (PCA) are still the mainstream method possibly due to its simplicity in both implementation and interpretation. Low-dimensional spaces built with these linear methods have been observed in a variety of neural recordings and animal models (C. elegans Ca2+ imaging (Nichols et al. [2017]), human electrophysiological FC (Stitt et al. [2017]), human fMRI BOLD signals (Shine et al. [2019]). However, the rich repertoire of available fMRI tasks probably lie on a bigger portion of the neural manifold, which can cause challenge for the linear approximations (Cunningham and Byron [2014], Gallego et al. [2017]).

Instead of using dimensionality reduction methods to directly reduce the fMRI scan's dimension, another way of organizing the high-dimensional data is through organizing the brain regions into large-scale brain networks. There is ample evidence that certain areas form local hierarchical relations and also large-scale circuits without clear hierarchical relations also exist (Selemon and Goldman-Rakic [1988], Cavada and Goldman-Rakic [1989], Hubel and Wiesel [1962]). There are also studies that explored the organization of large-scale distributed networks in the human cerebral cortex using resting-state fMRI and various community detection methods (Power et al. [2011b], Holmes et al. [2011], Garcia et al. [2018]). The identified brain networks are more densely connected within-network compared with cross-network. Robust canonical brain networks are identified across different subjects and show distinct FC patterns. However, most of the studies are based on non-overlapping node-based community detection methods and there are couple of limitations that arise from this setting. Firstly, the non-overlapping constraint only allows each brain region to be assigned to only one community and thus neglects the flexibility of our brain. It is unlikely that our brain should be simply parcellated into a discrete number of non-overlapping networks (Mesulam [1998]). To address the stringent network membership issue, multiple computational approaches that allow each brain region to be associated with more than one network have been proposed and adopted to discover overlapping network structures that are both meaningful and replicable (Yeo et al. [2014], Lee et al. [2016], Yeo et al. [2015]). Moreover, these node-based

community detection methods neglected inter-community connections by nature and did not give explicit network structures. A method that gives flexible brain region assignments and explicit network structures is thus needed.

1.2 Summary and contributions of this thesis

In this thesis, I propose different methods to help better understand our brain-behavioral relationship through fMRI data. The methods that I develop aim for several different research topics in cognitive neuroscience: behavioral measure prediction, low-dimensional fMRI data representation and brain network organization. And each of the topics can be related with one dimension of the 3D fMRI tensor data: subject, time and region. The following chapters will cover the details of the methods including related mathematical theories and the relevant experimental results that demonstrate the application and advantage of each method.

Chapter 2-3 investigates the subject dimension. It focuses on methods that integrate multiple sources of information in connectome-based predictive modeling (CPM) (Shen et al. [2017]) to improve the overall prediction accuracy. Being able to predict behavioral measures from fMRI data, especially FC, is meaningful as it assesses the 'brain-behavior' relationship and help determine where in the brain or what in the behavior is relevant to each other. As a validated and widely-used behavioral measure prediction framework, CPM takes the functional connectivity matrix from each subject as the input feature and validates the brain-behavior relationship via cross-validation. It has the advantage of generating accurate prediction with a simple model setup. However, the model is not designed to aggregate multiple connectomes or behavioral measures from a single subject. This has become more important as more and more fMRI datasets now include multiple scans and behavioral measures to increase the robustness and flexibility of the experiment.

In Chapter 2, we propose two different extensions over the original CPM framework to incorporate multiple connectomes. The first extension is based on canonical correlation analysis (CCA) where different connectomes are projected and merged together by CCA so that the combined con-

nectome is the mostly correlated with the target behavioral measure. This CCA merging step is performed prior and independent to CPM so that it can be applied to any other prediction method. By using the CCA-merged FC for prediction, it achieves higher accuracy than using any of the single FC alone. However, as the added CCA step brings more flexibility to the framework, it is also limited as the whole framework is optimized in two independent stages. Thus we propose another holistic model that is able to incorporate multiple connectomes along with building the prediction models. Through the use of regularized regression, we show that the model not only achieves the highest prediction accuracy, but also requires less parameter-tuning. Overall, these two extensions make CPM more flexible when considering the rapid speed of new datasets being generated and also indicates that more FCs per subject will help push the brain data closer to the observed behavior and improve the behavioral measure prediction if proper methods are used.

In Chapter 3, we investigate another side of CPM, the target behavioral measures. Two methods are proposed to combine multiple behavioral measures in the CPM framework. Both of the two methods are performed in an independent stage from CPM. We show that by finding the composite behavioral measure that is more aligned with the FC, we can achieve higher prediction accuracy. From a neuroscience perspective, this also enables us to find the composite measures that can be better explained by the FC, which pushes the 'brain-behavior' relationship closer from another side.

Chapter 4 focuses on the time dimension of the fMRI data tensor and calculates low-dimensional representation of the brain dynamics. Instead of viewing the brain as a rather static graph, we investigate how the brain changes dynamically within and across different scanning conditions. As the fMRI data is high-dimensional (number of voxels or brain regions is usually large), it is hard to understand or represent the data without proper dimensionality reduction. However, most of the current approaches used in fMRI data are linear dimensionality reduction methods, e.g., principal component analysis (PCA). With the number of available tasks in each dataset increasing, jointly embedding the fMRI time series from multiple tasks into a common space becomes necessary

and requires methods that can reveal the underlying manifold structure (Gao and Ganguli [2015]). Moreover, as the fMRI scans normally include a wide-range of tasks (e.g., motor, working memory and relational), the underlying structure is less likely to be linear. Thus, although PCA can provide satisfying result when embedding fMRI data from only one task, it may not reveal a clear structure when multiple tasks of data are included (Cunningham and Byron [2014]). To overcome the limitations of linear dimensionality reduction methods, we propose a novel non-linear manifold learning framework for the 3-dimensional fMRI data. This framework is able to generate a low-dimensional population-wise time series embedding. It is based on the diffusion maps algorithm (Coifman and Lafon [2006b]) and operates in a two-stage way to integrate similarity information from each subject. We apply this framework to data including over 300 subjects, each with 6 different scanning tasks. The joint low-dimensional embedding of this dataset reveals four brain states that are common to all the tasks. It is also able to reveal brain graph topological information that was previously discovered only by explicitly forming the dynamic functional connectivity (Shine et al. [2018]). These results indicate the advantage of using nonlinear methods when dealing with the increasingly complicated fMRI data and also validate the possibility of using an end-to-end framework that is able to reveal similar information as the hand-tuned framework.

Chapter 5 also focuses on the low-dimensional representation of the brain data. However, instead of embedding the fMRI time series, it focuses on the region dimension of the 3D tensor data and the edges between brain regions are embedded into the low-dimensional space. In terms of embedding the brain graph, the graph nodes (brain regions) are often embedded. Downstream tasks like clustering the brain regions into the functionally coherent clusters or brain networks can then be performed in this low-dimensional space. However, clustering edges of the brain graphs only starts to gain attention recently (Faskowitz et al. [2020]). Compared with clustering the nodes, clustering the edges naturally allows each brain region to be associated with multiple brain networks. Moreover, as the edge-level, instead of node-level, analyses are often performed (e.g., CPM, test-retest reliability), defining brain networks based on edges may potentially lead

to more accurate network definitions. However, when the popular linkage-clustering based edge clustering algorithm (Ahn et al. [2010]) is applied to the brain graph data, it is hard to get satisfying networks due to the inconvenience of controlling the cluster numbers, which inevitably leads to shattered or composite brain networks that require additional merging or splitting steps. In order to increase the flexibility of the edge clustering, we utilize the hierarchical structure of the brain edges and propose a Poincaré embedding-based edge clustering algorithm. By first embedding all the edges into a hyperbolic space, which is more capable of representing the tree-structured data, a subsequent k-medoids algorithm can be followed to cluster edges into k networks. The algorithm is applied on the resting-state FC and similar sets of networks (e.g., default mode network, motor network) are obtained from the clustering, which validates that the edge-clustering can provide similar node-level information. However, with the added flexibility of the framework, each nodes can now be associated with multiple networks and the node's number of associated networks can potentially reflect its functional flexibility. This framework can also be used as a versatile approach to investigate the edge-level information not only limited to clustering. By comparing it with traditional embedding methods like multidimensional scaling, we show that Poincaré embedding is more appropriate when representing edges of the graph.

1.3 Published components of this thesis and contributions

Chapter 2 and chapter 3 have been published. Chapter 4's algorithm is published while the extended application of the algorithm is under review. Chapter 5's algorithm is published and the extended application of the algorithm is in preparation for publication. The following authors contribute to each chapter:

Chapter 2: Abigail S. Greene, Dustin Scheinost, R. Todd Constable

Chapter 3: Xilin Shen, R. Todd Constable, Dustin Scheinost, Daniel S. Barron, Javid Dadashkarimi, Abigail S. Greene, Marisa N. Spann, Stephanie Noble, Evelyn Lake, John Krystal

Chapter 4: Gal Mishne, Dustin Scheinost

Chapter 5: Gal Mishne, Dustin Scheinost

Chapter 2

Combining Multiple Connectomes Improves Predictive Modeling of Phenotypic Measures

2.1 Introduction

Advanced functional magnetic resonance imaging (fMRI) techniques, particularly functional connectivity analyses, are revealing robust individual differences in patterns of neural activity that predict continuous phenotypic measures (Dubois and Adolphs [2016b], Rosenberg et al. [2018]). Predictive modeling of the associations between phenotypic measures and the functional organization of an individual's brain improves generalization of results to novel individuals and increases their eventual clinical utility. Recent work has used functional connectivity matrices, or connectomes, to predict a wide range of phenotypic measures, including fluid intelligence (Finn et al. [2015b]), brain maturity (Dosenbach et al. [2010]), and sustained attention (Rosenberg et al. [2016]). However, most of the current state-of-the-art algorithms only build predictive models based on a single connectome for each individual (Dadi et al. [2019]). This approach neglects the complementary information contained in connectomes from different sources and reduces prediction performance.

While functional connectivity is usually calculated from data acquired during rest, task conditions better reveal individual differences (Finn et al. [2017b], Vanderwal et al. [2017]) and improve

phenotypic prediction (Greene et al. [2018b], Rosenberg et al. [2016]). Further, the observed improvement in predictive power appears to be task specific, suggesting that task conditions are likely better at generating models of phenotypes related to the circuits they perturb (Greene et al. [2018b], Rosenberg et al. [2016]). Overall, it is unlikely that a single task can be developed that is optimal for all phenotypes. Instead, methods that incorporate functional connectivity information from a spectrum of tasks into a single predictive model may yield the best performance and most generalizable method for predicting phenotypic measures from connectomes.

In order to combine different task connectomes into a single predictive model in a principled way, we propose a novel prediction framework, termed multidimensional connectome-based predictive modeling. Two algorithms, each with their own strengths and limitations, are provided to illustrate the advantage of utilizing multiple connectomes. Both are based on the previously validated connectome-based predictive modeling (CPM) method (Shen et al. [2017]). The first utilizes canonical correlation analysis (CCA), while the second utilizes ridge regression. CCA combines multiple task connectomes by finding the projection direction which maximizes correlation between the combined connectomes and behavioral measure(s) to be predicted. In contrast, ridge regression directly incorporates the large number of edges in multiple connectomes through regularization.

Using two large open-source datasets with multiple tasks—the Human Connectome Project (HCP) (Van Essen et al. [2013a]) and the Philadelphia Neurodevelopmental Cohort (PNC) (Satterthwaite et al. [2016]), we validate and compare our two algorithms against performing CPM on each task connectome independently, CPM on a general functional connectivity (GFC) (Elliott et al. [2019]) matrix created by concatenating time series from all task conditions to create a single connectome for an individual, ridge regression on GFC and CPM with a naïve extension to multiple connectomes where each edge for each task is selected independently.

In all, our contribution in this paper is two-fold. First, we propose the combination of multiple connectomes from different task conditions in one predictive model. Second, we developed two

algorithms for dealing with multiple connectomes and show that they outperformed the validated single connectome predictive model.

2.2 Methods

2.2.1 Connectome-based Predictive Modeling (CPM)

CPM (Shen et al. [2017]) is a validated method for extracting and pooling the most relevant features from connectivity data in order to construct linear models to predict phenotypic measures (Figure 2.1a). Briefly, edges of connectivity matrices that are significantly correlated with the phenotypic measure of interest are selected. The selected features are then pooled (e.g. averaged) and linear regression is used to predict the phenotypic measure in novel individuals. It is designed for single connectome-based prediction, but can be easily extended to multiple-connectome scenarios as each edge is selected independently. However, this simple approach is not equipped to efficiently incorporate the increasing number of features introduced by multiple connectomes.

2.2.2 Multidimensional Connectome-based Predictive Modeling

Although CPM can be extended to leverage multiple connectomes, a specially designed framework will better utilize the complementary information in different brain connectivity patterns driven by corresponding task conditions. Here, we present two realizations of this framework to illustrate the feasibility and advantage of combining multiple connectomes for prediction.

CCA Connectome-based Predictive Modeling (cCPM)

Canonical Correlation Analysis (CCA): For two sets of observation matrices $\mathbf{X}$ and $\mathbf{Y}$, assuming that the variables are correlated, CCA seeks linear combinations of the columns of these two matrices that maximize correlation between them. In other words, we want to find vectors a and b such that the random variables $\mathbf{Xa}$ and $\mathbf{Yb}$ maximize the correlation. Assuming that $\mathbf{X}$ and $\mathbf{Y}$ are

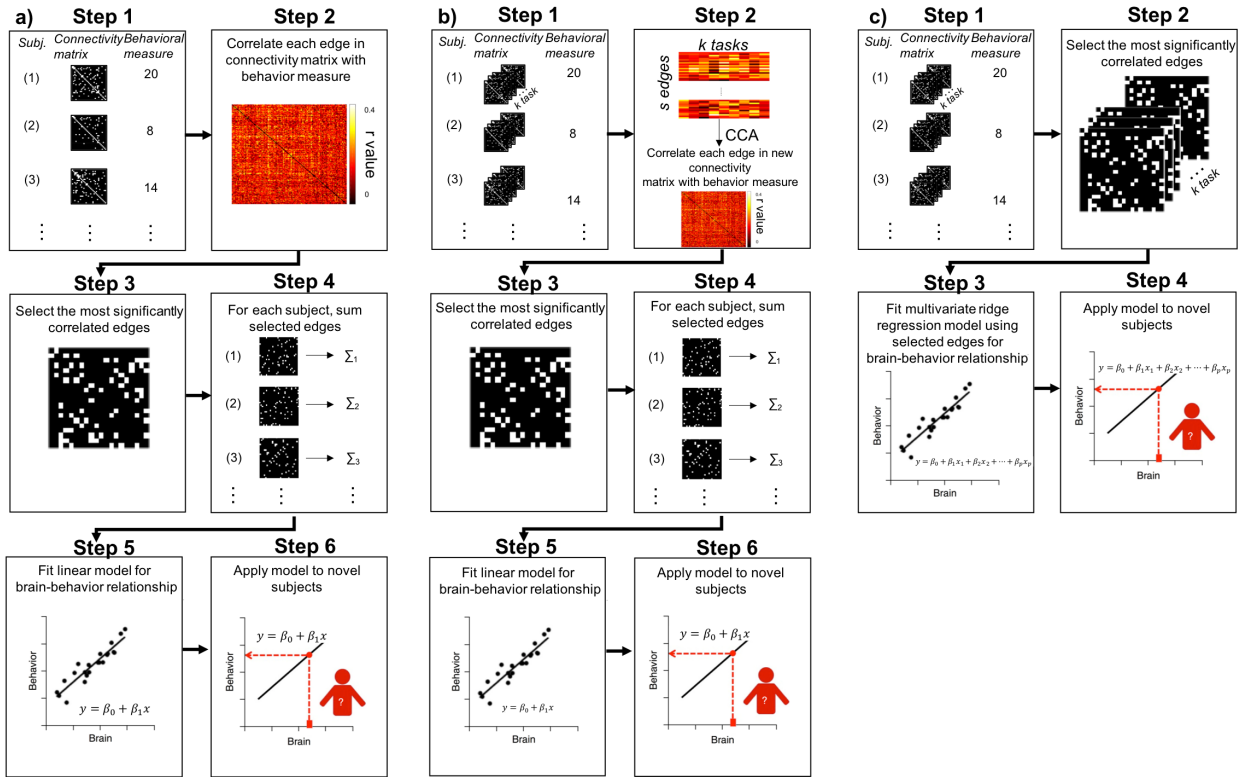


Figure 2.1: **Algorithm flow chart for three major models mentioned.** a) The original CPM flow chart b) cCPM extends CPM to handle multiple connectomes per individual by replacing the correlation step in CPM with a canonical correlation analysis (CCA) step. c) rCPM extends CPM to handle multiple connectomes per individual by replacing the pooling (i.e. averaging) and linear regression step with a ridge regression step.

centered such that each column of either matrix has mean zero, the correlation to be maximized can be expressed by the following equation:

$$\rho = \frac{((\mathbf{Xa})^T(\mathbf{Yb}))}{\sqrt{([(\mathbf{Xa})^T(\mathbf{Xa})][(\mathbf{Yb})^T(\mathbf{Yb})])}}$$

CCA Connectome-based Predictive Modeling: The cCPM (Gao et al. [2018b]) pipeline consists of six steps (Figure 2.1b).

In the first step, individuals are divided into training and testing sets using 10-fold cross-validation. We denote the number of individuals for training as N^{train} and number of individuals for testing as N^{test}.

In the second step, edges are combined. For the k-th edge, we have a matrix $\mathbf{E}_k \in R^{N^{train}*M}$

containing edge strength of the k-th edge for all the individuals in the training set. Each row of the matrix $\mathbf{E}_k$ denotes each training individual's k-th edge's different strengths under M different tasks. Using CCA, we can find the canonical coefficients $\mathbf{w}_k \in R^M$ for each edge. As each edge matrix $\mathbf{E}_k$ corresponds to the observation matrix $\mathbf{X}$ in the above definition equation for CCA, these coefficients $\mathbf{w}_k$ correspond to the vector $\mathbf{a}$, and the observation matrix $\mathbf{Y}$ will store the behavioral measures. We then combine connectomes from all tasks into a total connectivity matrix using different canonical correlations, $\mathbf{E}_k^{total} = \sum_{m=1}^{M} \mathbf{E}_k(:, m)\mathbf{w}_k(m)$, where the m-th column of $\mathbf{E}_k$ is denoted as $\mathbf{E}_k(:, m)$. Within each single task, each edge is demeaned across different individuals so that each column of $\mathbf{E}_k$ has mean 0.

In the third step, we assign the combined edges that are significantly correlated with the behavioral measures to the "correlated network" (CN). The significance of the correlation is found from the CCA. Here, we assume that CCA always maximizes the positive correlation between combined edge strength and behavioral measure as the sign of the canonical coefficients can trivially be changed to maximize the positive correlation. Various significance thresholds for feature selection can be used.

In the fourth step, we calculate "network strength" $\mathbf{s}^{CN}$ by pooling (i.e. summing) the strength of all CN edges in each individual's total connectivity matrix, yielding a summary value $\mathbf{s}^{CN}$ for each individual:

$$\mathbf{s}^{CN} = \sum_k (B(k) \sum_m \mathbf{E}_k(:, m)\mathbf{w}_k(m))$$

where $\mathbf{s}^{CN}$ is the vector of summary values, $B(k)$ is the indicator of whether the k-th edge passes the thresholding for CN.

In the fifth step, we use linear regression $y = \beta_0 + \beta_1 s^{CN}$ to model the association between "network strength" and the phenotypic measure in N^{train} individuals.

In the sixth step, the "network strength" is calculated for the excluded N^{test} individuals, and is submitted to the corresponding regression model to generate phenotypic measure estimates for those testing individuals. This process is repeated iteratively, with different individuals in the

training and testing sets.

Ridge regression Connectome-based Predictive Modeling (rCPM)

Ridge regression: In ordinary least-squares (OLS) regression, a greater number of independent variables compared to the number of observations leads to an ill-posed problem and overfitting. To solve this ill-posed problem, regularization on regression coefficients can be applied to shrink the coefficients. Ridge regression shrinks the regression coefficients by imposing a L2-norm penalty on their size. Compared with OLS regression, the coefficients from ridge regression minimize a penalized residual sum of squares,

$$\hat{\beta}^{ridge} = \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

where λ is the complexity parameter that controls the shrinkage strength: $\lambda = 0$ gives rise to the unregularized OLS, while increasing λ shrinks the coefficients towards zero. If we write the criterion in the above equation in matrix form,

$$RSS(\lambda) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) + \lambda\beta^T\beta$$

the ridge regression solutions can be solved by

$$\hat{\beta}^{ridge} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$$

where I is the identity matrix. Compared with the solution for OLS,

$$\hat{\beta}^{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

adding a positive constant to the diagonal of $\mathbf{X}^T\mathbf{X}$ before inversion makes the problem nonsingular, even if $\mathbf{X}^T\mathbf{X}$ is not of full rank.

Ridge regression CPM: Based on ridge regression, we modify the original CPM framework to better suit the high-dimensional nature of connectivity data (Figure 2.1c) (Gao et al. [2019a]). Specifically, due to the positive semi-definite nature of a functional connectivity matrix, the edges are not independent. Ridge regression is more robust than OLS in this case.

Instead of summing selected edges and fitting a one-dimensional OLS model, we directly fit a ridge regression model with training individuals using the selected edges from all the tasks and apply the model to testing individuals in the cross-validation framework. λ parameter in the ridge regression is chosen by another inner 10-fold cross-validation which uses only the training individuals. The largest λ value that has a mean squared error (MSE) within one standard error of the minimum MSE is chosen. In the Results, we show that rCPM is not sensitive to λ.

2.2.3 Experiment setup

Datasets: We applied all algorithms (see Competing methods below) to the Human Connectome Project (HCP) 900 Subject Release and the Philadelphia Neurodevelopmental Cohort (PNC) first study release. These data releases were the only releases available at the time that this work began.

Phenotypic measure: In both datasets, a matrix reasoning test—a measure of fluid intelligence (gF)—was used as the phenotypic measure for prediction. In the HCP dataset, a 24-item version of the Penn Progressive Matrices test was used; this test is an abbreviated form of Raven's standard progressive matrices (Bilker et al. [2012]). In the PNC dataset, 24- and 18-item versions of the Penn Matrix Reasoning Test were used (Bilker et al. [2012], Moore et al. [2015]). Integer scores indicate number of correct responses (HCP: PMAT24_A_CR, range=5–24, mean=17.53, s.d.=4.45, median=19; PNC: PMAT_CR (phv00194834.v1.p1.c1), range=0–23, mean=12.27, s.d.=4.04, median=12).

HCP participants: From this dataset, we restricted our analyses to those individuals who participated in all nine fMRI conditions (seven task, two rest), whose mean frame-to-frame displacement was less than 0.1 mm and whose maximum frame-to-frame displacement was less than 0.15

mm (see HCP imaging parameters and preprocessing), and for whom gF measures were available ($n = 515$; 241 males; ages 22-37). This conservative threshold for exclusion due to motion was used to mitigate the substantial effects of motion on functional connectivity; following this exclusion, there was no significant correlation between motion and gF for most conditions (all $p > 0.05$, Bonferroni corrected) except the Social task, right-left (RL) phase encoding run ($rs = -0.16(p = 0.00017)$), the Relational task, left-right (LR) phase encoding run ($rs = -0.15(p = 0.0008)$), and the Emotion task, RL phase encoding run ($rs = -0.14(p = 0.0017)$).

PNC participants: From this dataset, we used behavioral, structural imaging, and functional imaging data. We restricted our analyses to those individuals who participated in all three fMRI runs (two task, one rest), on whom registration was successful (nine individuals were excluded for failed registrations), whose mean frame-to-frame displacement was less than 0.1 mm and whose maximum frame-to-frame displacement was less than 0.15 mm (as for the HCP dataset, and with the same motivation), and for whom gF measures were available ($n = 571$; 251 male, ages 8–21). Following exclusion for motion, there was no significant correlation between motion and gF for any condition (all $p > 0.05$, Bonferroni corrected).

fMRI processing: fMRI data were processed with standard methods and parcellated into 268 nodes using a whole-brain, functional atlas defined in a separate sample (see (Greene et al. [2018b]) for more details). Task functional connectivity was calculated based on the "raw" task timecourses, with no regression of task-evoked activity: the mean timecourses of each node pair were correlated and correlation coefficients were Fisher transformed. Matrices were generated for both the LR and RL phase encoding runs in the HCP data, and these matrices were averaged for each condition, thus generating one 268×268 connectivity matrix per individual per task condition. Nodes that have missing coverage during any individual's scan were excluded from all individuals (9 nodes in HCP and 18 nodes in PNC were excluded). These matrices were used to generate cross-validated predictive models of gF.

Competing methods: We compared cCPM and rCPM to four simpler CPM-based approaches.

For the first approach, we performed CPM on each task independently as previously demonstrated (Greene et al. [2018b]). For the second approach, we performed CPM on a general functional connectivity (GFC) matrix created by averaging brain connectivity information across all task conditions (Elliott et al. [2019]). As time courses are z-score normalized before creating connectomes, averaging connectomes is similar to first, concatenating time courses and, then, correlating them. However, as the time length of each task is different, by concatenating time series first avoids biasing FC estimates toward the shortest tasks. Here, we generated GFC as suggested in the original paper by first concatenating and then correlating time courses. For the third approach, we compared ridge regression with CPM on GFC matrices. This is similar to performing rCPM on GFC matrices. However, we have chosen not to use the "rCPM" term here to avoid confusion between using GFC and our more direct way of combining multiple connectomes. Finally, for the fourth approach, we used a naïve extension to CPM, where all task connectomes were vectorized and concatenated to create a feature space that contained all task data for CPM. In contrast to the cCPM and rCPM approaches, this naïve implementation does not consider any shared or unique information offered by each task. Corrected resampled t-tests (Bouckaert and Frank 2010) were used to compare competing methods.

Internal validation: 10-fold cross-validation was used to train all models. In 10-fold cross-validation, the sample was randomly divided into 10, approximately equal-sized groups; on each fold, the model was trained on 9 groups and tested on the excluded 10th group. This process was iteratively repeated 10 times, with each group excluded once. This procedure was repeated for 100 random divisions. CPM was performed with a range of p-value edge selection thresholds from 0.001 to 0.5. Model performance was evaluated by the cross-validated R^2,

$$R^2_{CV} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

R^2_{CV} can be negative (Scheinost et al. [2019]) and negative values were set to 0. In this paper, $\sqrt{R^2_{CV}}$ is reported as it is comparable to, but less biased than, the normally used Pearson corre-

lation between observed and predicted measures when using cross-validation. $\sqrt{R^2_{CV}}$ is averaged over the cross validation folds.

External validation: Additionally, we trained models using one of the datasets (either HCP or PNC) and applied the model to the other dataset. For external validation, we only used the Emotion and Working Memory tasks from HCP for consistency with the available task data from the PNC. To fairly compare between models, CPM was performed with the 50%, 5% and 1% of edges with lowest p-values. Model performance was evaluated by Pearson correlation coefficient ($r_{Pearson}$) between the predicted and observed gF measures.

Quantification of task contribution: To quantify the contribution of each task to a given predictive model, we calculated the m-th task's average weight (labeled W_m) to the model as

$$W_m = \sum_k \mathbf{B}(k)\beta_k^m std(\mathbf{E}_k(:,m))$$

where $\mathbf{B}(k)$ indexes whether the k-th edge is selected, $std(\mathbf{E}_k(:,m))$ represents the standard deviation of the k-th edge in the m-th task and β_k^m represents the weight learned by cCPM or rCPM for the k-th edge in the m-th task. To make the results more interpretable, W_m are then normalized to have sum 1, $\sum_m W_m = 1$, so that it represents each task's contribution proportion in the whole model.

Similarly, as certain tasks may contain redundant information for prediction, we adopted forward feature selection to select the optimal combination of tasks. Forward feature selection finds the optimal combination of tasks by adding each of the tasks in a stepwise way; in each step, the task that improves prediction the most will be added to the selected task list. The optimal combination is found when any additional task won't lead to further improvement in prediction. We found the optimal task combination for HCP with both cCPM and rCPM. Both of the two algorithms are performed with a p-value threshold of 0.1.

Sensitivity to hyperparameters: Although all of the tested approaches are relatively simple in terms of choosing hyperparameters before training the model, all are dependent on the chosen

p-value threshold for edge selection. To evaluate the sensitivity of each approach to this hyperparameter, we repeated analyses with p-value thresholds of 0.001, 0.005, 0.01, 0.05, 0.1, and 0.5. The rCPM approach has an additional hyperparameter: the penalty weighting parameter λ. To show that the rCPM models are also not sensitive to the choice of λ, we fixed λ at the average chosen value in cross-validation and varied it in 10% steps to test whether fixing and perturbing λ changes the prediction performance of rCPM.

Data and code availability statement: The HCP data used in this study are publicly available on the ConnectomeDB database (https://db.humanconnectome.org). The PNC data used in this study are publicly available on the database of Genotypes and Phenotypes (dbGaP accession code phs000607.v1.p1); a data access request must be approved to protect the confidentiality of individuals. MATLAB scripts to run the cCPM and rCPM analyses can be found at (`https://github.com/YaleMRRC/CPM`). BioImage Suite tools used for analysis and visualization can be accessed at (`https://bioimagesuiteweb.github.io/webapp/connviewer.html`). MATLAB scripts written to perform additional post-hoc analyses are available from the authors upon request.

2.3 Results

2.3.1 Combining multiple connectomes improves prediction accuracy compared with single connectome-based prediction

As shown in Figure 2.2, all models that incorporate task data significantly predicted fluid intelligence, whereas the models based only on rest did not predict fluid intelligence better than simply predicting the population mean (i.e. R^2_{CV} did not differ from zero). In both datasets, rCPM (HCP: 0.436 ± 0.0072, PNC: 0.356 ± 0.0078) outperformed (HCP: $p = 7.37 * 10^{-4}$, PNC: $p = 0.0721$, corrected resampled t-test comparing rCPM to the next best performing method, GFC-ridge) all competing approaches. GFC-ridge (HCP: 0.387 ± 0.0068, PNC: 0.329 ± 0.151) and cCPM (HCP:

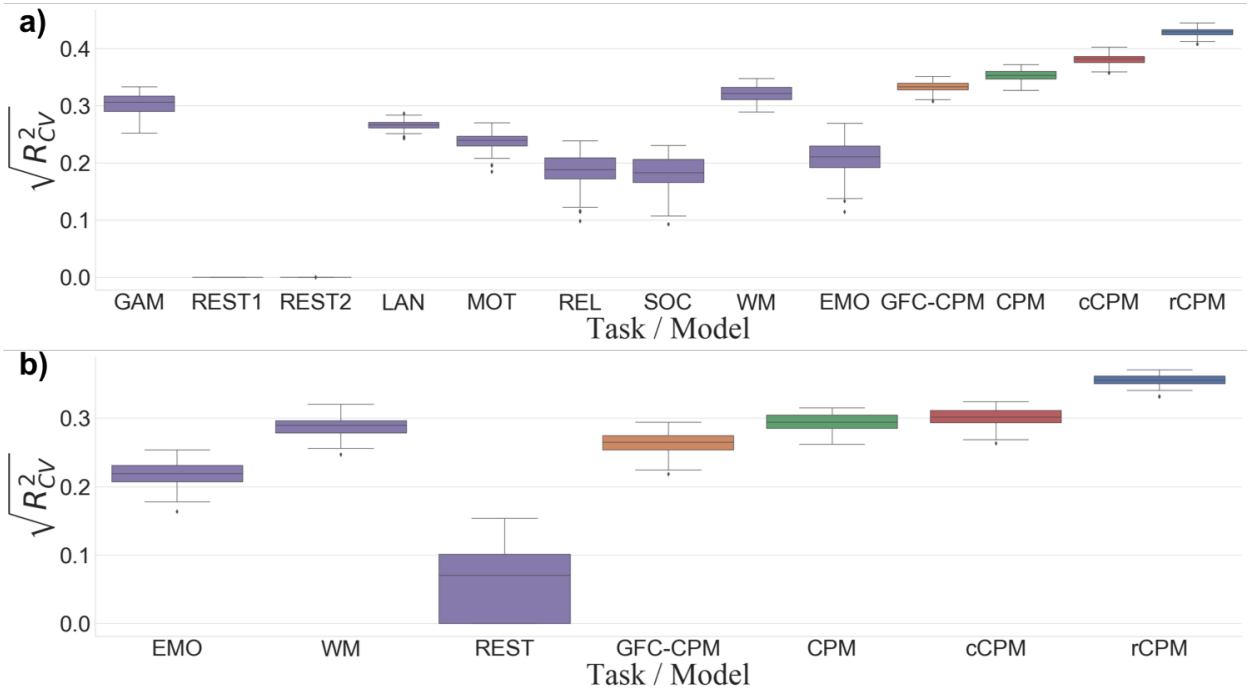


Figure 2.2: **Comparison of the predictive modeling approaches' ability to predict an individual's gF.** a) HCP dataset. b) PNC dataset. Purple box plots show results from CPM on a single task. The orange, green, red, and blue box plots show results from combining multiple task connectomes using GFC-CPM, GFC-ridge, CPM, cCPM, and rCPM, respectively. Box plots show cross-validated $\sqrt{R^2_{CV}}$ with the error bars representing the 25th and 75th percentiles, respectively. Values below (or above) the 25th (or 75th percentiles) are shown as *. The best results across different edge selection thresholds are shown. Task acronyms: GAM: Gambling, LAN: Language, MOT: Motor, REL: Relational, SOC: Social, WM: Working Memory, EMO: Emotion.

0.386 ± 0.0084, PNC: 0.301 ± 0.0122) performed similarly to each other (HCP: $p = 0.43$, PNC: $p = 0.18$) in both the two datasets and outperformed the other competing methods. The naïve CPM implementation (HCP: 0.354 ± 0.0094, PNC: 0.293 ± 0.0130) and CPM using a GFC matrix (HCP: 0.333 ± 0.0086, PNC: 0.263 ± 0.0151) had similar performance to models built from the best-performing single task, working memory (HCP: 0.322 ± 0.0134, PNC: 0.288 ± 0.0144). We also tested the performance of ridge regression on single task connectomes. While the overall prediction performance increases when using ridge regression, rCPM still significantly outperforms ridge regression on single task connectomes (HCP: $p = 0.0012$, PNC: $p = 0.0146$, comparing rCPM to the next best performing single task with ridge regression).

2.3.2 Different tasks contribute differentially to the model

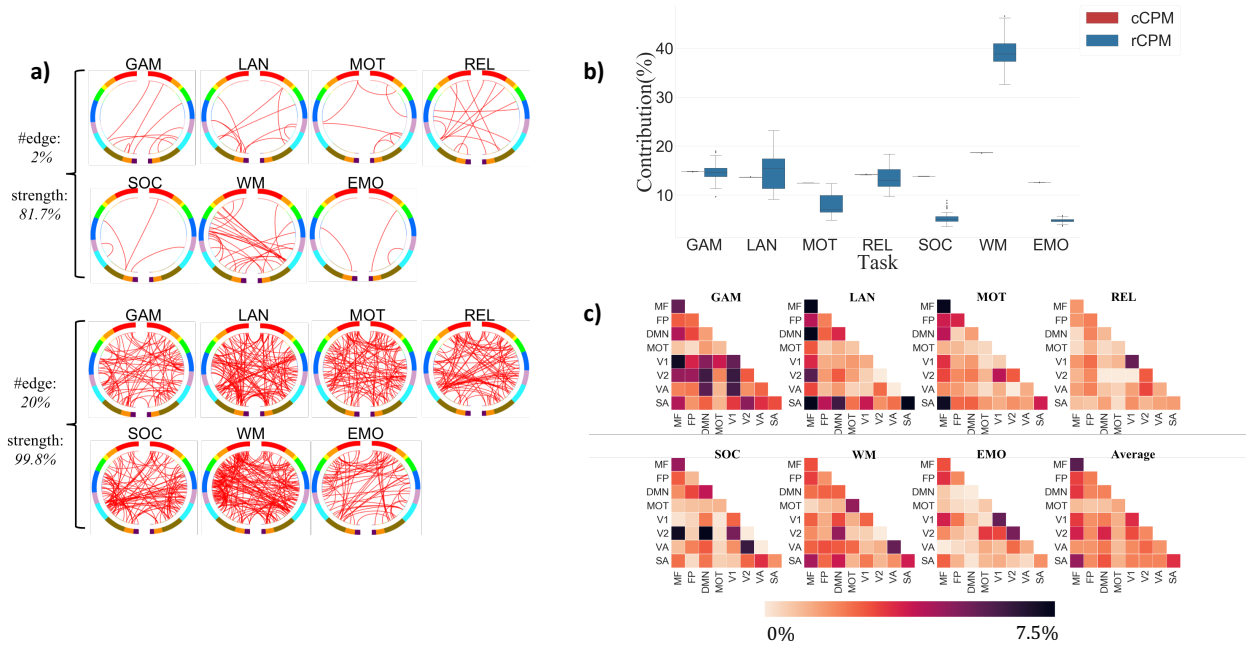


Figure 2.3: **Different tasks' contributions to the model.** a) Visualization of the selected edges for different tasks in the model. Top row represents 2% and bottom row represents 20% of total number of selected edges. 81.7% and 99.8% of feature contribution (combined sum of each feature's regression coefficient times its standard deviation) in regression are possessed respectively by those networks. Anatomical acronyms: PFC = Prefrontal, MOT = MotorStrip, INS = Insula, PAR = Parietal, TEM = Temporal, OCC = Occipital, LIM = Limbic, CER = Cerebellum, SUB = Subcortical, BSM = Brainstem. b) Different tasks' average contribution fraction to the cCPM and rCPM model. c) Different tasks' contributions to the model, summarized at the network level. Network acronyms: MF=Medial Frontal, FP=Frontoparietal, DMN=Default Mode Network, MOT=Motor Cortex, V1=Visual I, V2=Visual II, VA=Visual Association, SA=Salience.

As shown in Figure 2.3, different tasks contribute different numbers of edges to the final model. Tasks that are more predictive by themselves (e.g., Working Memory) tend to contribute more edges to the model, while less predictive tasks (e.g., Emotion) contribute fewer edges (Figure 2.3a). Similarly, in terms of contribution W_m to the regression model, tasks that are predictive by themselves contribute more to the overall predictive model while less-predictive tasks contribute less (Figure 2.3b). Additionally, tasks appear to select different edges in different networks for prediction (Figure 2.3c). In the figure, percentage of edges in each network pair that are selected by the model are shown. While the presented results are mostly for rCPM (Figure 2.3ac), the same

trend is observed for cCPM (Figure S2). As shown in Figure S3, the predictive utility of a task is not dependent on the length of the task.

a)

Step	$\sqrt{R^2_{CV}}$
{WM}	0.315
{WM, MOT}	0.351
{WM, MOT, REL}	0.374
{WM, MOT, REL, EMO}	0.385
{WM, MOT, REL, EMO, GAM}	0.390
{WM, MOT, REL, EMO, GAM, SOC }	**0.391**
{WM, MOT, REL, EMO, GAM, SOC, LAN}	0.380

b)

Step	$\sqrt{R^2_{CV}}$
{WM}	0.358
{WM, MOT}	0.400
{WM, MOT, REL}	0.418
{WM, MOT, REL, LAN}	0.425
{WM, MOT, REL, LAN, SOC}	0.431
{WM, MOT, REL, LAN, SOC, GAM}	**0.433**
{WM, MOT, REL, LAN, SOC, GAM, EMO}	0.421

Figure 2.4: **Forward task selection for cCPM and rCPM.** a) shows the results for cCPM while b) shows the results for rCPM. The optimal task combinations for the two algorithms are both using 6 tasks, where cCPM excludes the Language task while rCPM excludes the Emotion task. However, the algorithm performance of using all 7 available tasks is not significantly worse than using 6 tasks (cCPM: $p = 0.38$, rCPM: $p = 0.40$), and overall, including more tasks significantly improves prediction.

Finally, as shown in Figure 2.4, using stepwise forward task selection, the second-best-performing individual task was not added to the combined model until step 5 or 6 (for rCPM and cCPM, respectively). Similarly, the best-performing model did not use all tasks. Together, these results suggest that certain tasks may contain redundant information for prediction and that an optimal combination of tasks with complementary information is needed to maximize prediction performance.

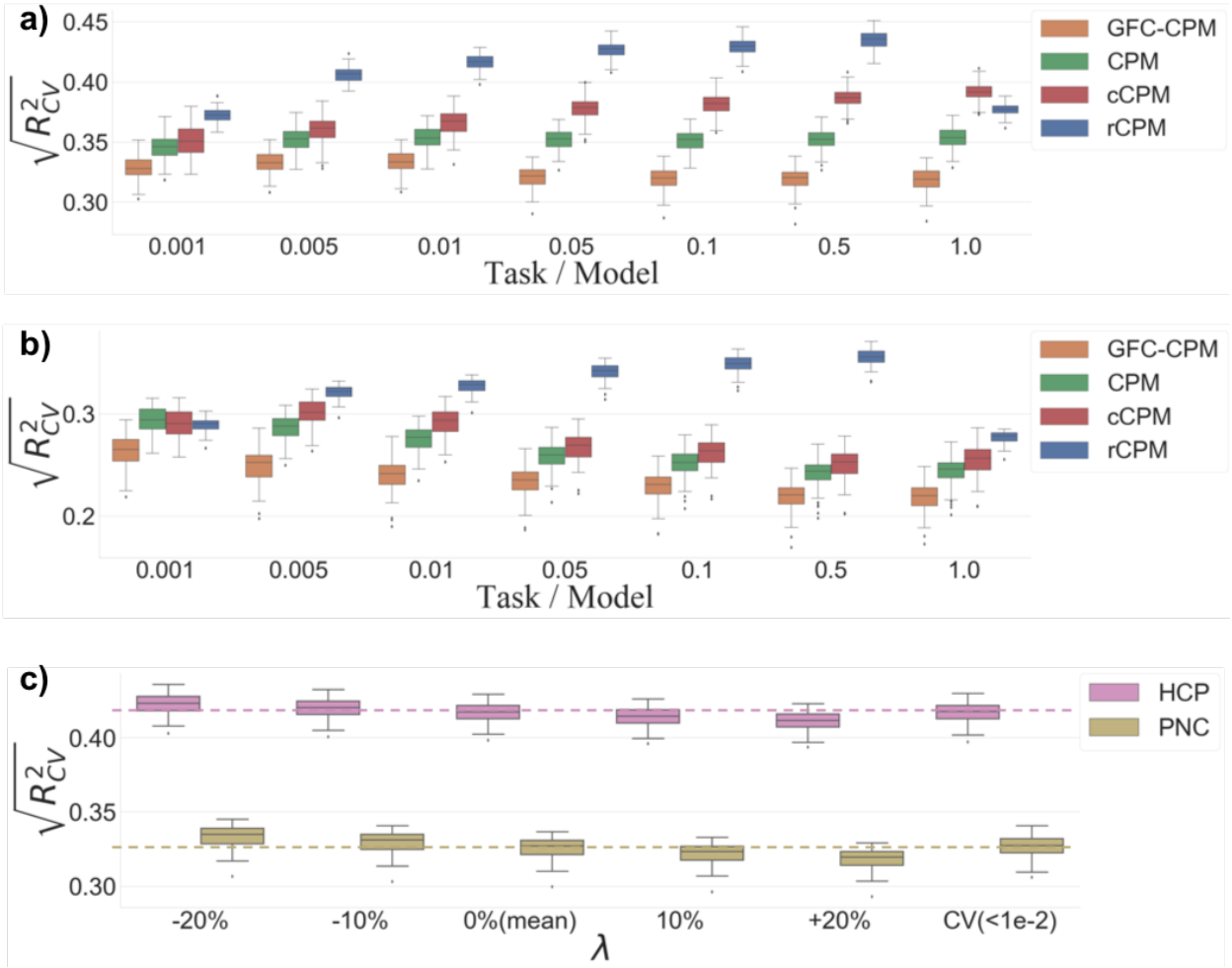


Figure 2.5: **Models' performance with various hyperparameters.** a) Varying edge selection threshold for the HCP dataset. b) Varying edge selection threshold for the PNC dataset. c) Varying penalty weighting parameter for rCPM. In a) and b) edge selection threshold=1.0 represents no edge selection. Horizontal line indicates prediction accuracy with λ chosen by inner cross validation.

2.3.3 Evaluation of hyperparameters on model performance

As shown in Figure 2.5ab, the performance of all approaches varies as a function of the p-value threshold in the edge-selection step. The p-value threshold controls the number of edges retained in the models. A lower p-value threshold represents more stringent edge selection and fewer edges will enter the final model. The GFC-CPM, GFC-ridge, the naïve CPM and cCPM all exhibit a decrease in performance as the p-value threshold is increased (i.e. more retained features), except

for cCPM and GFC-ridge on the HCP dataset. In contrast, rCPM exhibits better performance as the p-value threshold is increased. As a result, the improvement in prediction performance offered by rCPM over the competing approaches is at its maximum at a higher p-value threshold (i.e. $p = 0.1$ or $p = 0.5$). However, the extra computation cost induced by more features should also be considered. As shown in Figure 2.5c, results from rCPM are insensitive to the choice of λ over the tested range.

2.3.4 Models trained on one dataset can be generalized to another dataset

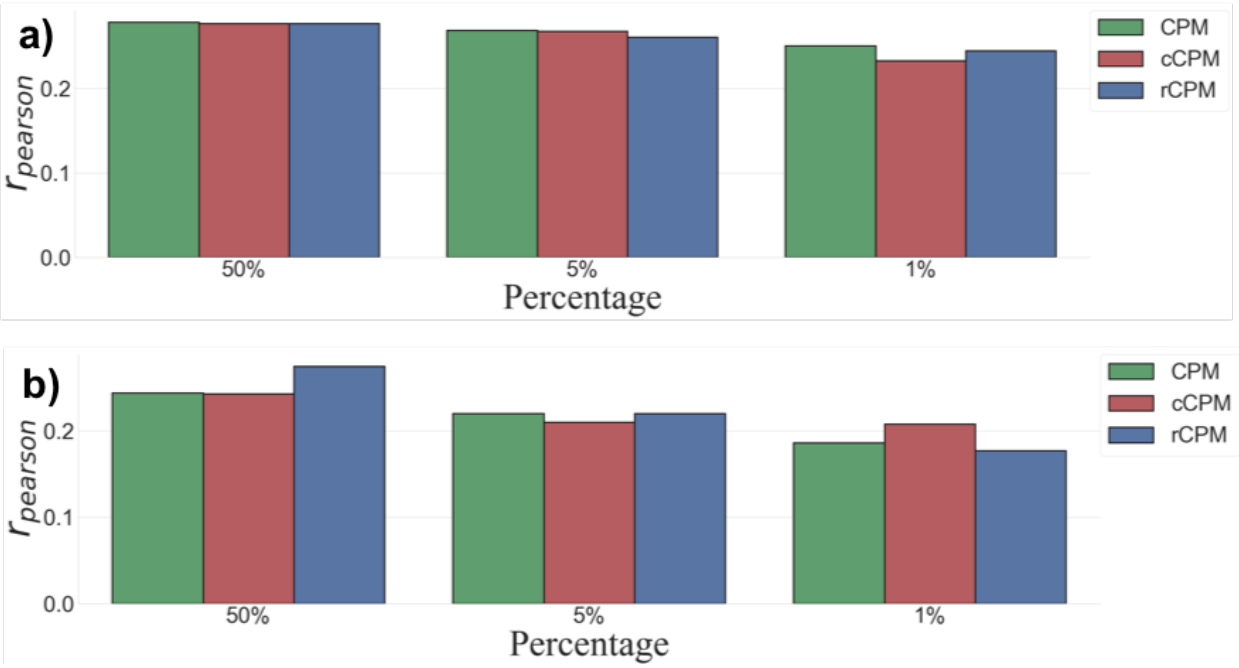


Figure 2.6: **Different models' generalizability to independent, external datasets.** a) Trained on HCP and applied to PNC. b) Trained on PNC and applied to HCP. b) Trained on PNC and applied to HCP. The results are presented as Pearson correlation between predicted and actual measures. Models trained on either HCP or PNC datasets can significantly predict gF in the other dataset.

Figure 2.6 shows that our models generalized to independent, external datasets. By showing the results under different edge sparsity levels, we validate that model generalizability does not decrease with more edges as more features often lead to overfitting. Actually, models based on more edges still outperform models based on fewer edges. GFC's generalizability is also tested

here as it can potentially utilize all the available task scans in HCP dataset and apply it to PNC, which has fewer tasks. However, we didn't see improvement in generalizability by including more tasks.

2.3.5 Exploratory comparison of ridge regression to lasso (least absolute shrinkage and selection operator) and elastic net

Given the improved performance of rCPM, we explored the performance of CPM approaches based on two other regularized regression approaches, lasso and elastic net. Instead of imposing a $L2$-norm penalty on regression coefficient β as in ridge regression, lasso uses a $L1$-norm penalty as $\lambda \sum_{j=1}^{p} |\beta_j|$. This encourages coefficients to be set to zero, while ridge regression only shrinks the size of the coefficients. Thus, lasso is often preferred for feature selection when sparsity is preferred. However, lasso tends to select only a small number of variables when the sample size is small. To overcome this limitation, elastic net regularization combines both the L1-norm penalty and L2-norm penalty. An α hyperparameter is used to balance between the weight of the two kinds of penalty. So the overall objective function for elastic net can be written as,

$$\hat{\beta}^{EN} = \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda(\sum_{j=1}^{p} \frac{1-\alpha}{2}\beta_j^2 + \alpha \sum_{j=1}^{p} |\beta_j|)$$

For this objective function, the bigger α is, the more lasso-type shrinkage will be put on the coefficients. When α=1, it is the same as lasso; when $\alpha = 0$, it is the same as ridge regression. Implementations of these approaches are identical to rCPM with the exception that the ridge regression step is replaced with either lasso or elastic net.

In our experiments, neither elastic net nor lasso performed as well as ridge regression (HCP: $p = 4.4 * 10^{-5}$, PNC: $p = 0.0035$). For both datasets, $\alpha = 0$ always produces the best prediction (Figure 2.7), suggesting—again—the importance of including a large number of edges in a predictive model to best reflect distributed patterns of functional connectivity. In this chapter,

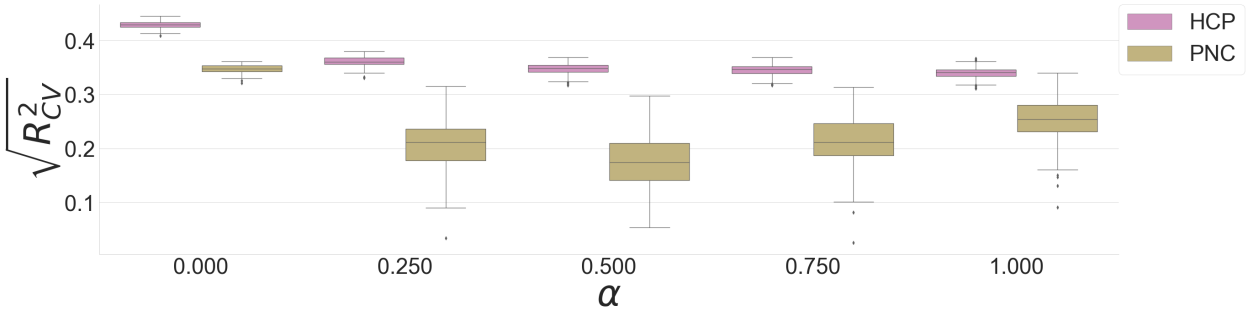


Figure 2.7: **Comparison of ridge regression with Elastic Net and Lasso.** α, the weighting parameter between ridge and lasso-type regularization, is varied across different values. $\alpha = 0$ is the same as ridge regression while $\alpha = 1$ is the same as lasso. Ridge regression generates the most accurate prediction in both the HCP and PNC datasets.

we proposed cCPM and rCPM, two general connectome-based prediction frameworks that utilize complementary information in different task connectomes to improve phenotype prediction. We tested the two algorithms on two open-source datasets, HCP and PNC, to predict fluid intelligence using all the available task connectomes. rCPM shows superior performance in prediction for within-sample prediction compared to the competing methods, including cCPM, though all methods performed similarly for out-of-sample prediction. By looking at the contribution of specific tasks to the final prediction and the stepwise forward optimal task combination selection, we found that different tasks contribute differentially to the final predictive model. In contrast to other competing methods, rCPM performed better when the number of features included in the ridge regression step was large and included features only weakly correlated with intelligence. Last but not least, although the model takes in a large number of features, we validated its generalizability and robustness by using a range of hyperparameters and testing models on external datasets. Overall, cCPM and rCPM provide a powerful framework to combine all available functional imaging data into a single predictive model.

The major contribution of this chapter is showing that combining connectomes in an appropriate manner improves prediction. Although task-based connectomes improve prediction performance compared to resting-state connectomes (Greene et al. [2018b]), approaches to combine multiple task-based connectomes into a single predictive model are limited. The GFC method is

one approach to combine task-based connectomes (Elliott et al. [2019]). However, by simply averaging over all conditions, GFC loses a large amount of task-specific information, which can be used to increase prediction performance, and performs worse than the other methods that combine multiple connectomes for predictive modeling.

Several regression approaches exist to shrink regression coefficients in the case of highly correlated features, which is common in connectome-based predictive modeling. Two common approaches are principal components regression (PCR), and partial least squares (PLS). Ridge regression shrinks both the high- and low-variance directions of the features, but applies greater shrinkage to the low-variance directions (Krishnan et al. [2011], Mwangi et al. [2014], Zhong et al. [2009]) In contrast, PCR does not shrink the high-variance directions and simply discards the lower-variance directions (Frank and Friedman [1993]). PLS also shrinks the low-variance directions, but may inflate the higher variance directions (Frank and Friedman [1993]). Based on these theoretical considerations and additional experimental observations (see for example (Dubois et al. [2018], He et al. [2018]), ridge regression may be preferred for minimizing prediction error because of its smooth shrinkage (Frank and Friedman [1993]), though we did not test these other approaches in the context of connectome-based predictive modeling. Those related approaches will be tested in future work.

Similarly, ridge regression generated better predictions than approaches that perform shrinkage and feature selection, like lasso and elastic net. While popular in neuroimaging machine learning (Dadi et al. [2019]), lasso has several limitations compared to ridge regression (Zou and Hastie [2005]). When the number of features is greater than the sample size, lasso limits the number of nonzero features to be the same as the sample size, even though additional features may be associated with the phenotype of interest. Similarly, lasso tends to retain only one feature from any set of highly correlated features, shrinking the other features to zero. In cCPM and rCPM (and other connectome-based predictive modeling approaches), the number of features (i.e. edges) is typically greater than the number of individuals and features are highly correlated. While elastic

net regularization attempts to combine the strengths of ridge regression and lasso (Zou and Hastie [2005]), we did not observe any improvements in prediction performance over ridge regression with elastic net.

Despite both approaches incorporating complementary task information, rCPM significantly outperformed cCPM. As only one phenotypic measure was used in these analyses, cCPM simplifies to linear regression. Given this, we chose to follow the CPM design of pooling selected features, rather than constructing a single large matrix with all selected features for linear regression. This single matrix would be rank deficient, resulting in unstable solutions. Ridge regression is a natural answer to this problem as the regularization term allows for stable solutions that minimize the effect of noisy edges on model performance.

RCPM demonstrated improved prediction performance when more features were included in the model. Although rCPM incorporates a modest feature selection step, retaining nearly 60% of the edges (edge number retained with $p < 0.5$) during this step produced the best prediction performance. We attribute these results to ridge regression's ability to shrink noisy and correlated features, thereby, reducing their influence on the model. Moreover, these results suggest that even edges that are not strongly correlated with the behavioral measures still help with prediction, phenotypic information is encoded in spatially distributed connectivity patterns, and (almost) all available information should be used for prediction. These results also align with recent work in "double descent" test risk of predictive models, which suggests a lower test risk when large numbers of features—more than the number of samples—are included (Belkin et al. [2019]).

A natural question that arises is: if each task differentially contributes to the final predictive model, what tasks should be included when designing a study? When using a single task, results suggest that tasks and phenotypes of interest should be matched such that the selected task perturbs brain circuits relevant to the phenotype (Greene et al. [2018b], Rosenberg et al. [2016]). Yet, when combining multiple tasks, including tasks with complementary information appears to be the most beneficial. For example, in the forward feature selection result (Figure 2.4), the second task added

is the motor task, which by itself is a poorer predictor of fluid intelligence than either the gambling or language task. However, the motor task likely provides more complementary information than the gambling or language task and, thus, provides the maximum gain in prediction power when added to the model. To generalize, we hypothesize that a battery of tasks that perturbs multiple and complementary brain circuits will yield better predictive models than a battery of tasks that perturbs a single brain circuit with different, subtle manipulations. While it may not be possible to collect a single battery of tasks that is optimal for all phenotypic information, we suspect that a standard battery could be developed that is good enough in most cases. Future work should develop and compare different batteries of tasks in terms of predictive modeling.

This work has some limitations. First, currently for cCPM and rCPM, all individuals are required to have complete data from all the tasks. As such, using cCPM and rCPM increases the likelihood of removing individuals from the analysis due to missing data. As the utility of predictive models is often dependent on sample size (Cui and Gong [2018], Varoquaux et al. [2017]), cCPM and rCPM may only be applicable to larger datasets that can support removing individuals due to missing data. Future work includes extending cCPM and rCPM to support data imputation methods to handle missing data. Second, cCPM and rCPM may make model interpretation and visualization more challenging. While cCPM and rCPM retain CPM's ability to simply map features back to the brain, the mappings from both of the two algorithms are task specific. Ultimately, the task at hand will determine if the added task information reduces interpretability of the model (see Rule #10 from (Scheinost et al. [2019]) for a greater discussion). RCPM exhibited increased prediction performance when the number of features was larger, and while better prediction is good, the large number of edges can be difficult to visualize and interpret. It should be noted however that this may be a more accurate depiction of the complex systems involved in executing cognitive functions, and our tendency to reduce findings to one or two brain regions is likely grossly oversimplifying this complex system. In other words, the circuits involved in cognition may not lend themselves to easy visualization. Third, cCPM and rCPM can make external validation harder.

The external dataset needs to have similar tasks for the model to be applicable. For example, when using external validation between the HCP and the PNC datasets, we limited the tasks from the HCP to only two for compatibility with the PNC dataset. Fourth, cCPM and rCPM did not show improvement for out-of-sample prediction. This result could suggest that cCPM and rCPM overfit for within-sample prediction or that the HCP and PNC datasets differ in important aspects, such as age, that limit generalizability of any model (see Rule #9 from (Scheinost et al. [2019])). Fifth, similarly, due to the large number of edges (features) involved in the prediction models compared with the sample size, overfitting is hard to eliminate completely (Whelan and Garavan [2014]) and this causes the difference between our results with the previous works (Finn et al. [2015b]). However, with the use of both cross-validation and independent dataset generalization, our evaluation results are less-prone to suffer from overestimation. Finally, while we focus on connectivity data derived from multiple fMRI tasks, cCPM and rCPM are agnostic to the type of input data and can easily incorporate structural connectivity data from DTI or from other functional modalities like EEG. In future work, we will explore the inclusion of other measures to further improve prediction.

In summary, we present cCPM and rCPM, two extensions to CPM to handle connectomes from multiple sources. Our results suggest that prediction of phenotypic measures can be improved by including multiple task conditions in computational models, that different tasks provide complementary information for prediction, and that cCPM and rCPM provide two principled methods for modeling such data.

Chapter 3

Combining Multiple Behavioral Measures Improves Predictive Modeling of Phenotypic Measures

3.1 Introduction

Advanced functional magnetic resonance imaging (fMRI) techniques, especially functional connectomics, are revealing robust differences between individuals (Dubois and Adolphs [2016a]). While connectomes are usually calculated from resting-state data, task conditions improve phenotypic predictions (Greene et al. [2018b]). Further, models combining multiple connectomes outperform models built from a single connectome (Gao et al. [2018a]). Whether using connectomes from single or multiple sources, most predictive models only predict a single behavioral measure. Yet, a single behavioral measure is not able to describe an individual's cognitive abilities. A set of similar behavioral measures that describe different aspects of the cognitive ability are often available within the the same datasets. Inspired by the improvement we had by combining multiple connectomes, we aim to improve the prediction also from the other side: combining multiple behavioral measures. To assess the improvement we can have by combining multiple behavioral measures, we propose and test two different frameworks: PCA-CPM and mCCA-CPM, where mCCA stands for multipath-CCA, a novel CCA variant that is able to combine multiple

sources, instead of only two sources.

In the PCA-CPM framework, principal component analysis (PCA) is used to combine multiple behaviors and the first principal component of PCA will be used as the new composite measure for prediction. This new composite measure is expected to be more representative of the overall cognitive ability and more robust than any of the individual cognitive measures used. It is tested on a dataset including different psychiatric disorder cohorts and a general memory related measure is built and predicted by the PCA-CPM framework, which show improved prediction performance. Moreover, this general memory measure can also be used in transdiagnostic prediction, which illustrates the generalizability of the composite behavioral measure.

While the multiple behaviors can be combined by PCA, it is still a unsupervised approach, i.e., the PCA result won't be influenced by the connectome. However, the goal of combining multiple behavioral measures is to find a composite measure that can be better predicted from the functional connectome, and thus reveal a more robust brain-behavior relationship. To this end, we propose another supervised learning framework that combines both multiple behavioral measures and also different connectomes. As mentioned in Chapter 2, multiple connectomes can be combined by canonical correlation analysis (CCA) (Gao et al. [2018a]), separate CCA are performed for each edge. When using a battery of behavioral measures, this procedure creates different latent behaviors for each edge, preventing results from each CCA being pooled. By designing a novel objective function that maximizes all the pairwise edge-behavior correlation, our mCCA-CPM framework simultaneously finds the optimal projection for both a battery of connectomes and a battery of behavioral measures through mCCA, leading to improved prediction performance. We offer a closed-form and an iterative solution for our mCPM framework. Both frameworks are evaluated using data from the Human Connectome Project (HCP) (Van Essen et al. [2013a]). PCA-CPM framework is additionally evaluated on the UCLA Consortium for Neuropsychiatric Phenomics (CNP) dataset (Poldrack et al. [2016a]).

3.2 Methods

3.2.1 Principal Component Analysis (PCA)-CPM

Connectome-based Predictive Modeling (CPM): As introduced in the Chapter 2, CPM (Shen et al. [2017]) is a validated method for extracting and pooling the most relevant features from connectivity data in order to construct linear models to predict behavioral measures. We use the ridge-regression based CPM (Gao et al. [2019a]) in this chapter, i.e., the selected CPM features are fit to learn the regression model coefficients via ridge regression.

PCA-CPM: Compared with the multidimensional-CPM prediction framework introduced in the previous chapter, the only additional goal here is to combine multiple behavioral measures and it is done in an independent step besides CPM. As CPM is normally evaluated by the 10-fold cross validation (CV), in PCA-CPM, we apply PCA on the multiple behavioral measures in the training dataset, get the first principal component along with the loadings for the first component. The ridge-CPM model is then built to predict the first principal component using the FCs within the training dataset. It is worth noting that when using the ridge-CPM model, multiple connectomes can also be utilized to improve the prediction as described in the previous chapter.

3.2.2 multipath Canonical Correlation Analysis(mCCA)-CPM

Canonical Correlation Analysis (CCA): For two sets of random variables $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_m)$, assuming that the variables are correlated, CCA seeks linear combinations of these two sets of random variables that maximize their correlation. In other words, CCA finds vectors $\mathbf{a}^* \in \mathbb{R}^n$ and $\mathbf{b}^* \in \mathbb{R}^m$ such that the random variables $\mathbf{X}\mathbf{a}^*$ and $\mathbf{Y}\mathbf{b}^*$ have maximum correlation. Assuming that $\mathbf{X}$ and $\mathbf{Y}$ are demeaned, i.e., each column of either matrix has mean zero, the procedure can be expressed by the following equation:

$$(\mathbf{a}^*, \mathbf{b}^*) = \underset{\mathbf{a},\mathbf{b}}{\operatorname{argmax}} \frac{(\mathbf{Xa})^\top (\mathbf{Yb})}{\sqrt{[(\mathbf{Xa})^T (\mathbf{Xa})][(\mathbf{Yb})^\top (\mathbf{Yb})]}} \tag{3.1}$$

Multipath Canonical Correlation Analysis (mCCA): Although CCA is able to find projection that maximizes correlation between two sets of variables, maximizing the cross-correlation between multiple datasets is of broader interest. A number of generalizations of CCA for this case exist. However, all versions can be summarized by the specific combination of 5 different objective functions and 4 different constraints (Asendorf [2015]), and almost all of the previous algorithms aim to find solution that maximizes the sum of all the possible inter-dataset correlations. Suppose we have in total M datasets, it will maximize the sum of all the possible $\frac{M\times(M-1)}{2}$ correlations. While in our problem, the aim is to maximize the sum of correlations between one and all the other $M-1$ datasets, or in total $M-1$ correlations. Suppose that there are in total M edges for each connectome, we can then denote the i-th edge's data matrix by $\mathbf{X}_i \in \mathbb{R}^{T\times N}, i = 1, 2, ..., M$, where T represents the number of task functional connectivity matrices each individual has and N is the total number of individuals. Similarly, we denote the behavior matrix by $\mathbf{Y} \in \mathbb{R}^{B\times N}$, where B represents the number of available behavioral measures. The objective function we are maximizing then becomes:

$$J(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_M, \mathbf{b}) = \sum_{i=1}^{M} \mathbf{a}_i^\top \mathbf{R}_{X_iY} \mathbf{b}, \tag{3.2}$$

where $\mathbf{a}_i \in \mathbb{R}^T$, $\mathbf{b} \in \mathbb{R}^B$. $\mathbf{R}_{X_iY} \in \mathbb{R}^{T\times B}$ is the cross-covariance matrices between $\mathbf{X}_i$ and $\mathbf{Y}$:

$$\mathbf{R}_{X_iY} = \sum_{i=1}^{N} (\mathbf{X}_i - \overline{\mathbf{X}}_i)(\mathbf{Y} - \overline{\mathbf{Y}})^\top, \tag{3.3}$$

Here, $\overline{\mathbf{X}}_i = \frac{1}{N}\sum_{j=1}^{N} \mathbf{X}_i(:, j)$ and $\overline{\mathbf{Y}} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{Y}(:, j)$ are the sample-mean for edge data matrix $\mathbf{X}_i$ and behavioral matrix $\mathbf{Y}$. Proper constraint needs to be satisfied on the above objective function

in order to prevent $\mathbf{a}_i$ and $\mathbf{b}$ from increasing unboundedly. Denoting covariance matrix of $\mathbf{X}_i$ and $\mathbf{Y}$ as $\mathbf{\Sigma}_{X_i}$ and $\mathbf{\Sigma}_Y$ separately, we use the constraint:

$$(\sum_{i=1}^{M} \mathbf{a}_i^\top \mathbf{\Sigma}_{X_i} \mathbf{a}_i) + \mathbf{b}^\top \mathbf{\Sigma}_Y \mathbf{b} = 1 \tag{3.4}$$

We propose two types of solutions for the optimization (i) a closed-form solution using generalized eigendecomposition, and (ii) an iterative solution.

Closed-form solution: To find the optimal parameter for objective function (3.2) under the constraint (3.4), we use a Lagrangian multiplier:

$$L(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_M, \mathbf{b}, \lambda) = \sum_{i=1}^{M} \mathbf{a}_i^\top \mathbf{R}_{X_i Y} \mathbf{b} - \lambda(\sum_{i=1}^{M} \mathbf{a}_i^\top \mathbf{\Sigma}_{X_i} \mathbf{a}_i + \mathbf{b}^\top \mathbf{\Sigma}_Y \mathbf{b} - 1) \tag{3.5}$$

The optimal projection can then be found as the solution of $\delta L / \delta \mathbf{a_i} = \mathbf{0}$ and $\delta L / \delta \mathbf{b} = \mathbf{0}$, which yields the following equations:

$$\begin{aligned} \mathbf{R}_{X_i Y} \mathbf{b} &= 2\lambda \mathbf{\Sigma}_{X_i} \mathbf{a}_i, \\ \sum_{i=1}^{M} \mathbf{R}_{X_i Y}^\top \mathbf{a} &= 2\lambda \mathbf{\Sigma}_Y \mathbf{b}. \end{aligned} \tag{3.6}$$

By concatenating all the $\mathbf{a_i}$ and $\mathbf{b}$ together into $\mathbf{v} = [\mathbf{a_1}^\top \cdots \mathbf{a_M}^\top \mathbf{b}^\top]$, we can rearrange equations (3.6) into a matrix form: $\mathbf{Rv} = \lambda \mathbf{Dv}$ where,

$$\mathbf{R} = \begin{pmatrix} 0 & 0 & \cdots & 0 & \mathbf{R}_{X_1 Y} \\ 0 & 0 & & & \mathbf{R}_{X_2 Y} \\ \vdots & & 0 & & \vdots \\ & & & 0 & \mathbf{R}_{X_M Y} \\ \mathbf{R}_{X_1 Y}^\top & \mathbf{R}_{X_2 Y}^\top & \cdots & \mathbf{R}_{X_M Y}^\top & 0 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} \mathbf{\Sigma}_{X_1} & 0 & \cdots & & 0 \\ 0 & \mathbf{\Sigma}_{X_2} & & & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ & & & \mathbf{\Sigma}_{X_M} & \\ 0 & 0 & \cdots & 0 & \mathbf{\Sigma}_Y \end{pmatrix}$$

When $\mathbf{D}$ is non-singular, the solution to (3.6) is the eigenvectors of

$\mathbf{D}^{-1}\mathbf{R} \in \mathbb{R}^{(M\times T+B)\times(M\times T+B)}$, a sparse matrix with $(2 \times M \times T \times B)$ entries. The solution to this sparse eigendecomposition problem is fast and accurate when the edge number M is large (Stewart [2002]).

Iterative solution: To find the iterative solution for the objective function (3.2) under constraint (3.4), combining the two equations produces a single objective function:

$$J(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_M, \mathbf{b}) = \frac{\sum_{i=1}^{M} \mathbf{a}_i^\top \mathbf{R}_{X_iY} \mathbf{b}}{(\sum_{i=1}^{M} \mathbf{a}_i^\top \mathbf{\Sigma}_{X_i} \mathbf{a}_i) + \mathbf{b}^\top \mathbf{\Sigma}_Y \mathbf{b}}. \tag{3.7}$$

This objective function J's value remains unchanged when the scale of $\mathbf{a}_i$ and $\mathbf{b}$ is multiplied by the same amount. When maximizing J, this prevents the function from only increasing the scale of $\mathbf{a}_i$ and $\mathbf{b}$ without actually increasing the correlation between the projected variables. To optimize (3.7), we use the first-order condition, yielding equations similar to (3.6). We are able to use fixed-point iteration method by moving $\mathbf{a}_i$ and $\mathbf{b}$ alone to one side. The iterative solution is summarized in Algorithm 1.

Algorithm 1 Iterative method for mCCA

Input: $\mathbf{R}_{X_iY}$, $\mathbf{\Sigma}_{X_i}$, $\mathbf{\Sigma}_Y$ - cross-covariance/covariance matrix, $i = 1, 2, ..., M$
ϵ - convergence threshold.

Output: $\mathbf{a}_i, \mathbf{b}$ - projection variable, $i = 1, 2, ..., M$.

Initialize $\mathbf{a}_i^{(0)}$, $\mathbf{b}^{(0)}$ with uniformly distributed random vectors

while $|\mathbf{b}^{(t+1)} - \mathbf{b}^{(t)}| > \epsilon$ **do**

$\mathbf{a}_i^{(t+1)} = \mathbf{\Sigma}_{X_i}^{-1} \mathbf{R}_{X_iY} \mathbf{b}^{(t)}$;

$\mathbf{b}^{(t+1)} = \mathbf{\Sigma}_Y^{-1} \sum_{i=1}^{M} \mathbf{R}_{X_iY}^\top \mathbf{a}_i^{(t)}$;

$\mathbf{a}_i^{(t+1)} = \frac{\mathbf{a}_i^{(t+1)}}{(\sum_{i=1}^{M} \mathbf{a}_i^{(t)\top} \mathbf{\Sigma}_{X_i} \mathbf{a}^{(t)}) + \mathbf{b}^{(t)\top} \mathbf{\Sigma}_Y \mathbf{b}^{(t)}}$; ▷ rescaling

$\mathbf{b}^{(t+1)} = \frac{\mathbf{b}^{(t+1)}}{(\sum_{i=1}^{M} \mathbf{a}_i^{(t)\top} \mathbf{\Sigma}_{X_i} \mathbf{a}^{(t)}) + \mathbf{b}^{(t)\top} \mathbf{\Sigma}_Y \mathbf{b}^{(t)}}$; ▷ rescaling

The projection variable $\mathbf{a}_i$ and $\mathbf{b}$ are initialized with uniform random vectors and the objective function J is calculated based on them. $\mathbf{a}_i$, $\mathbf{b}$ and J are then iteratively updated. Within each iteration, $\mathbf{a}_i$ and $\mathbf{b}$ are rescaled to keep the sum of variances of the transformed variables close to 1, *i.e.,* concatenated vector of all $\mathbf{a}_i$ and $\mathbf{b}$ is always on the ellipsoid. As mentioned above, this prevents $\mathbf{a}_i$ and $\mathbf{b}$ from increasing unboundedly and J's value is not affected by the rescaling.

mCCA-CPM: Based on the proposed mCCA method, we designed a prediction framework that takes in multiple connectomes and predicts latent phenotypes from a battery of behavioral measures. The framework works in two steps:

Step 1: **Combine multiple connectomes and multiple behavior measures.** On the training data, apply mCCA to find the optimal projection $\mathbf{a}_i$ for edge $\mathbf{X}_i$ and $\mathbf{b}$ for behaviors $\mathbf{Y}$. Then use $\mathbf{a}_i$ and $\mathbf{b}$ to combine the connectomes and behaviors. This reduces the 3D connectome input $\mathbf{X} = [\mathbf{X}_1^\top, \ldots, \mathbf{X}_M^\top] \in \mathbb{R}^{N \times T \times M}$ to 2D matrix $\tilde{\mathbf{X}} \in \mathbb{R}^{N \times M}$ and 2D behavioral matrix $\mathbf{Y} \in \mathbb{R}^{N \times B}$ to composite behavior score vector $\tilde{\mathbf{Y}} \in \mathbb{R}^N$.

Step 2: **Univariate feature selection.** For each column in $\tilde{\mathbf{X}}$, representing an individual edge, compute its Pearson correlation coefficient and the corresponding *p*-value with the behavior score. Select those informative edges by an predefined selection threshold (i.e. $p < 0.1$) to choose edges that are more correlated with the behavior. This reduces the edge number to K.

Step 3: **Use ridge regression (RR) to train the prediction model.** Now for each individual we have K edges as independent variables and one behavioral score as the dependent variable. We then use RR to train the prediction model on the training data. $y_s = \beta_0 + \beta_1 \tilde{\mathbf{x}}_s + \varepsilon_s, \quad s = 1, \ldots, N.$, where $\beta_1 \in \mathbb{R}^{K \times 1}$

3.2.3 Experiment setup

Two datasets were used: the UCLA Consortium for Neuropsychiatric Phenomics (CNP) dataset (Poldrack et al. [2016a]) and the Human Connectome Project (HCP) dataset (Van Essen et al. [2013a]).

CNP participants: The CNP data has been described in detail elsewhere (Poldrack et al. [2016a]). From the CNP data, we selected behavioral and functional imaging data from 172 total de-identified and anonymized participants (HC=73, SCZ=33, BPAD=34, ADHD=32). From the UCLA neu-

ropsychological battery, working memory was measured using the Weschler Memory Scale (WMS) symbol span, WMS digit span, and Weschler Adult Intelligence Scale (WAIS) letter-number sequencing; short-term memory was measured using the Verbal recall I, and California Verbal Learning Task (CVLT) short-delay free recall; and long-term memory was measured using Verbal recall II, CVLT long delay free recall, and CVLT scene recognition overall accuracy. For predictive modeling, fMRI acquisitions during the balloon analog risk task (BART), Paired Associative Memory encoding (PAM-E), Paired Associative Memory retrieval (PAM-R), Spatial Working Memory Capacity (SCAP), Stop Signal (SS), and Task Switching (TS) were used.

HCP Dataset: In this dataset, each individual performed 7 tasks in the scanner: gambling (GAM), language (LAN), motor (MOT), relational (REL), social (SOC), working memory (WM), and emotion (EMO). Cognitive ability was assessed by tasks from the NIH toolbox and Penn computerized neurocognitive battery (CNB). We used 9 cognitive tasks that are related with intelligence, divided into 4 aspects of intelligence, as previously defined (Dubois et al. [2018]) (see Fig. 3.2). Among the 9 cognitive measures, 3 memory-related ones were used for the general memory measure's generalization test. Unadjusted scores were used for all measures. We restricted our analyses to those individuals who participated in all 9 fMRI conditions (7 task, 2 rest), whose mean frame-to-frame displacement was less than 0.1 mm, whose maximum frame-to-frame displacement was less than 0.15 mm, and for whom the 9 behavioral measures used were available (N = 514; 240 males; ages 22-36+). This conservative threshold for exclusion due to motion was used to mitigate the substantial effects of motion on functional connectivity.

FMRI processing: For both the two datasets, fMRI data were processed with standard methods and parcellated into 268 nodes using a whole-brain, functional atlas defined previously in a separate sample. Next, the mean timecourses of each node pair were correlated and Fisher transformed, generating seven 268×268 connectomes per individual. Task connectomes were calculated based on the "raw" task timecourses, with no regression of task-evoked activity. These matrices were used to generate cross-validated predictive models of general intelligence.

Evaluation: All the experiments used 10-fold CV. PCA and mCCA were only performed on the training fold to avoid data leakage. Model performance was evaluated by the cross-validated R^2, $R^2_{CV} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})}{\sum_{i=1}^{n}(y_i - \bar{y})}$ (Alexander et al. [2015]). $\sqrt{R^2_{CV}}$ was reported for comparability to the normally-used Pearson correlation coefficient. $\sqrt{R^2_{CV}}$ was calculated within each fold separately and averaged cross folds. The 10-fold CV was repeated 100 times to avoid cherry picking. Significance is assessed at $p < 0.05$ calculated as by paired *t*-test.

Comparison of the iterative and the closed-form solutions: We tested both the speed and accuracy of the two solutions with simulated and the HCP dataset. The test is ran on a machine with an Intel Xeon Gold 6128 CPU @ 3.40GHz and 96GB of RAM, running Ubuntu 16.04.4 LTS. The algorithm is implemented and tested on MATLAB R2018a 64-bit version. For the iterative solution, a convergence threshold $\epsilon = 10^{-5}$ and a minimum number of steps $n = 10$ were used. For the simulated data, we generated random data matrix $\mathbf{X}_i \in \mathbb{R}^{100\times1000}, i = 1, 2, ..., 100$ and $\mathbf{Y} \in \mathbb{R}^{20\times1000}$. Each item in $\mathbf{X}_i$ and $\mathbf{Y}$ is independently drawn from a standard normal distribution. For the HCP dataset, different numbers of edges are tested to investigate if the preferred solution is dependent on the number of edges. To choose the sets of edges, for each task, we first calculate its average correlation with the set of behaviors and then for each edge the correlations are averaged again over all the tasks to get an overall description. Edges are then ranked by this overall correlation and the ones with higher correlation are chosen.

3.3 PCA-CPM predicts general memory measure

3.3.1 Transdiagnostic prediction of memory constructs

We were able to predict working, short-, and long-term memory constructs across diagnosis. The 6 task-based connectomes predict working memory (median $q^2 = 0.16$, $p < 0.001$, permutation testing, 1,000 iterations, 1-tailed), short-term (median $q^2 = 0.22$, $p < 0.001$, permutation testing, 1,000 iterations, 1-tailed), and long-term (median $q^2 = 0.20$, $p =< 0.001$, permutation

testing, 1,000 iterations, 1-tailed). Similar prediction accuracy is observed if all memory measures—regardless of category—are included (median $q^2 = 0.27$, $p < 0.001$, permutation testing, 1,000 iterations).

In line with previous CPM results, our models are complex with contributions from each task and distributed across multiple brain areas. In general, each task-based connectome contributes to prediction performance. For short-term and long-term memory, the PAM-RET and BART tasks contributed the most to overall prediction. For working memory, task contributions are more uniform. For the short and long-term memory models, the top three contributing nodes to prediction were located in the right prefrontal cortex, cerebellum (left crus I), and the right motor strip (Fig. 3.1C). For the working memory model, the top three contributing nodes in the left medial prefrontal, right temporal-parietal junction, and right temporal lobe (Fig. 3.1C).

3.3.2 Model validation on external datasets

The general memory model (including working, short- and long-term memory measures) trained on the CNP dataset ($n = 172$) successfully generalized to the HCP dataset ($r = 0.17$, $p < 0.01$, $d.f. = 513$). At the same time, a summary memory model trained on the HCP dataset ($n = 514$) generalized back to the CNP dataset ($r = 0.40$, $p < 0.01$, $d.f. = 170$). We observe differences in prediction performance when training with the CNP and HCP datasets. We suspect that this is due to the fact that the sample size used to train our models is three times large for the HCP compare to the CNP (i.e., 514 vs 172), allowing us to achieve higher prediction performance when training with the HCP.

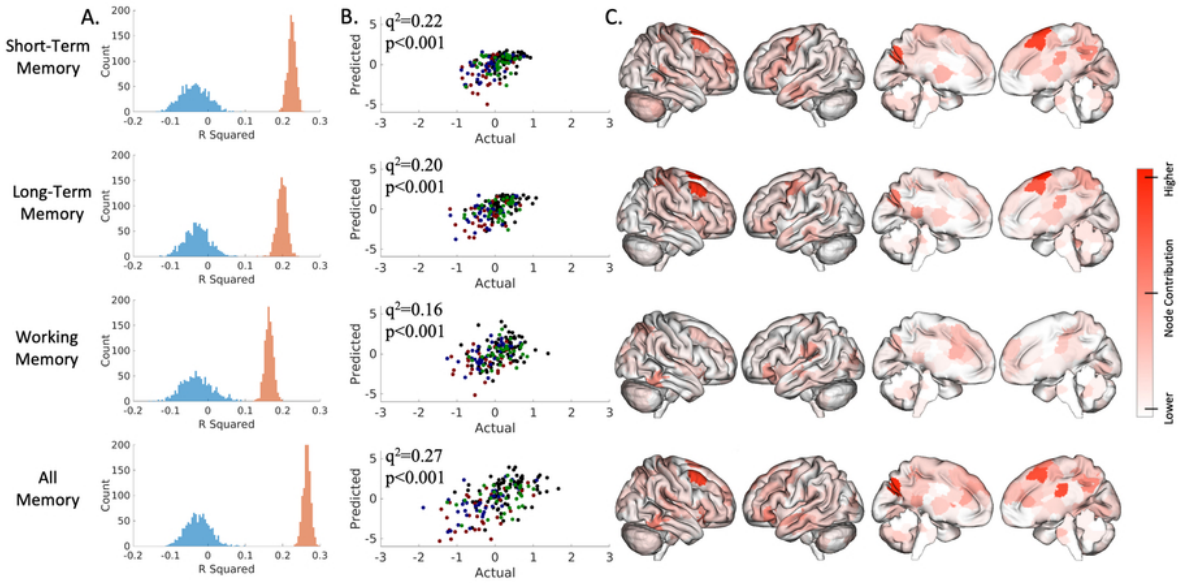


Figure 3.1: **Connectome-based predictive model performance for transdiagnostic 10-fold cross-validation.** The left column (A) shows a histogram of the model performance across 1,000 iterations of the actual (red) and randomly permuted (blue) data. The middle column (B) shows how actual and predicted values compare for the median-performing model (green, SCZ; blue, BPAD; red, ADHD). The right columns (C) show surface plots of each node's degree, which is defined as the number of edges per node that were weighted in 95% of iterations (the short-term memory model includes 289 consistently weighted edges; long-term, 276 edges; working, 174; all, 362).

3.4 mCCA-CPM predicts general intelligence

3.4.1 mCCA-CPM better predicts intelligence

mCCA-CPM generates significantly better predictions of intelligence than the best performing model for a single measure and the PCA model (Fig. 3.2a). While measures from the same category are predicted with similar accuracy, each category from the battery of behavioral measure, except visuospatial ability, has at least one score that contributes to the latent intelligence phenotype found by mCCA-CPM, suggesting that the latent phenotype spans the whole battery of behavioral measures.

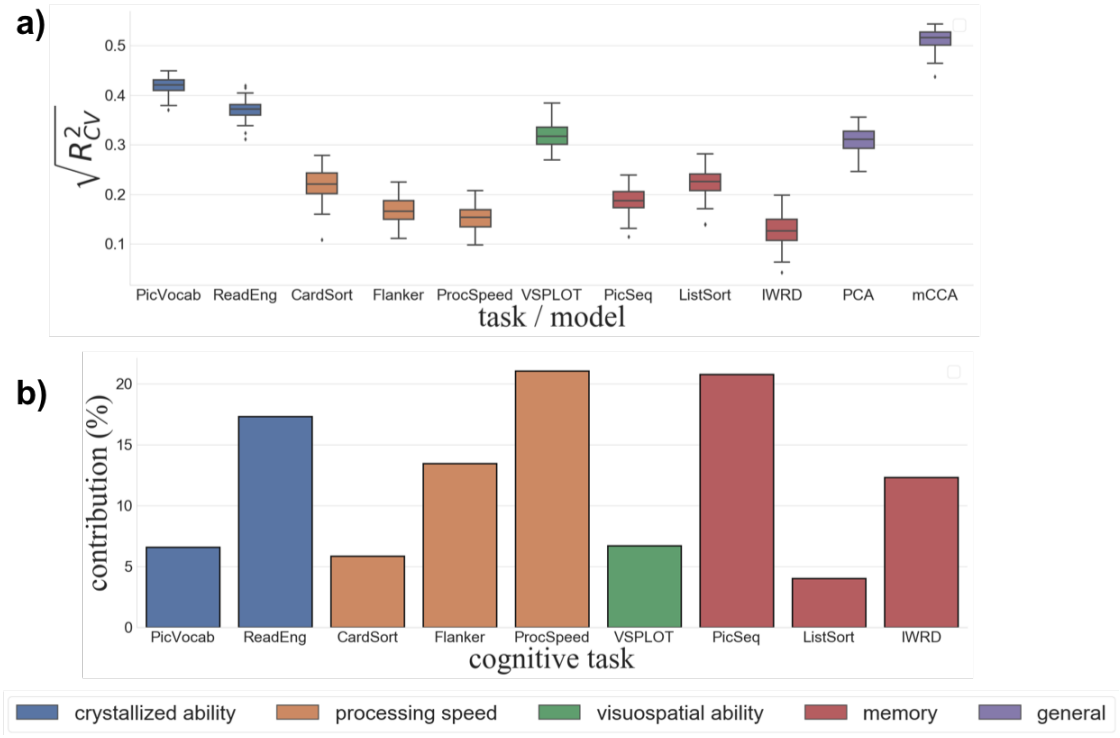


Figure 3.2: **Prediction of behavioral measures. a)** Predict single measure with ridge regression and combined measure with PCA or mCCA. Boxes show 25^{th} to 75^{th} percentile and whiskers show min and max values. **b)** Contribution of each measure in the mCCA-CPM model as defined by its fraction in the latent factor. Cognitive tasks are colored by the factor analysis result from (Dubois et al. [2018]).

3.4.2 Predictive edges are widely distributed across the brain

When we further look at the prediction model, we see that all the connectomes have significant contribution in the final prediction (Fig. 3.3a). While WM is the most influential connectome ($> 20\%$), all the other connectomes have relatively similar contribution ($\sim$10-15$\%$). The edge contribution (Fig. 3.3b-c) also shows a distributed pattern of the whole brain. Edge contribution is defined as the weighted sum of the ridge regression coefficients multiplied by the edge strength standard deviation and weighted by the task weights learned from mCCA-CPM.

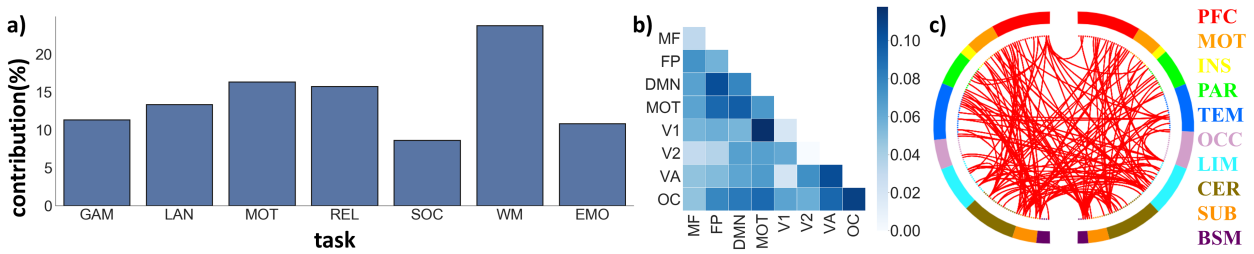


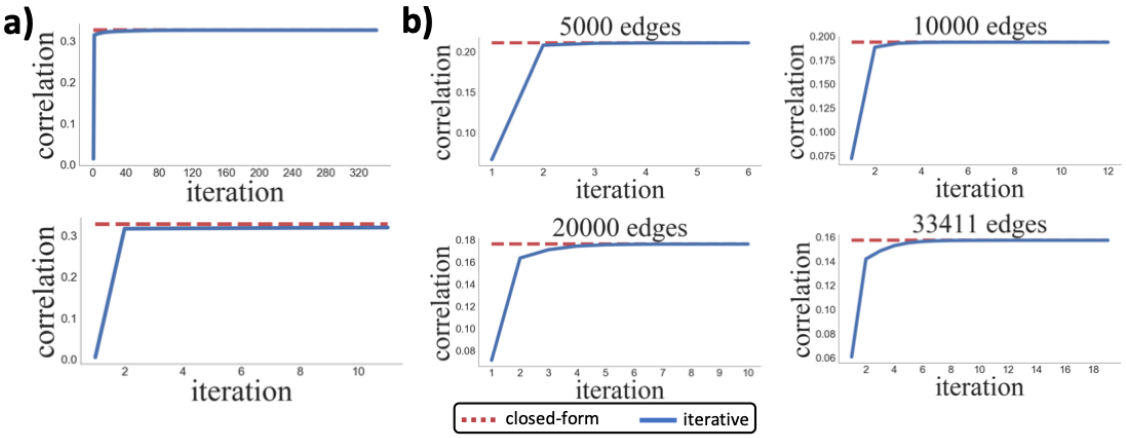
Figure 3.3: **Visualization of the task and edge contribution in the general intelligence prediction. a)** Each task contributes—defined as the sum of coefficients of all the edges selected—uniquely to the final prediction. **b)** Network and **c)** circle plot of the most influential edges.

3.5 Comparison of the mCCA's iterative solution to the closed-form solution

From the simulated data (Fig. 3.4), the closed-form solution performs better in both accuracy and computational speed. The iterative solution suffers from the variable steps to converge due to the randomness in the simulated data. However, for the result of real data, an opposite trend is observed. Because of the higher correlation between edges, the iterative solution converges to the ground truth in about 10 steps, which greatly reduces the total computation time compared with the closed-form solution. Further, when the number of edges is low, the iterative solution converges with the similar speed as solving for the closed-form solution. As the number of edges increases, the iterative solution begins to converge faster than the closed-form solution. This results from the more correlated edge structure (*i.e.* different $\mathbf{X_i}$s are more correlated) in the real data. From Fig. 3.4b, the average step to reach convergence is also relatively small (< 20 steps).

3.6 Summary

We proposed two frameworks, labeled PCA-CPM and mCCA-CPM, for both generating and predicting latent phenotypes from a battery of behavioral measures. The PCA-CPM framework used PCA to combine multiple behavioral measures and rCPM to predict the combined general mea-



c)

Test	*Closed-form solution*	*Iterative solution (<1e-5)*
Simulation	0.59±0.01s	18.26±16.04s
HCP (5000edges)	9.90±0.15s	7.91±1.18s
HCP (10000edges)	38.39±0.07s	18.59±1.57s
HCP (20000edges)	175.25±0.20s	44.22±5.85s
HCP (33411edges)	558.50±10.22s	94.43±7.05s

Figure 3.4: **Comparison of mCCA's closed-form solution to the iterative solution a)** top row shows within the 100 repeated experiments with different random data matrix, the one that iterative solution generates the most accurate solution. Bottom row shows when iterative solution generates the most inaccurate result. **b)** Each subplot shows the one fold that iterative solution gives the most inaccurate result. **c)** Timing of the two algorithms. Shown in the format of mean±standard deviation.

sure. The mCCA-CPM framework utilized a novel solution that generalizes the traditional CCA problem and can serve as a special condition of the more general multiset CCA problem. A closed-form and an iterative solution were proposed to solve the mCCA problem. The iterative solution is fast and accurate when the number of edges are large and correlated. The PCA-CPM framework was applied on the CNP to build and predict a general memory measure from a battery of memory-related cognitive scores. Transdiagonstic prediction of this general memory can be performed with 4 different cohorts, suggesting that the same macroscale brain networks subserve memory across diagnostic groups and that individual differences in memory performance are related to individual differences within this brain circuit. The score can be predicted, By predicting a composite memory score, we had better accuracy compared with predicting the single scores. Moreover, the

general memory prediction model was also generalized across datasets (between CNP and HCP datasets), suggesting that the general memory measure was not limited within a specfic dataset and can be generalized across datasets given the availability of multiple measures in the same cognitive domain. As a supervised substitute of the PCA-CPM framework that aims to maximize correlation between composite measures with the FCs, the mCCA-CPm framework was applied on the HCP dataset and a latent intelligence phenotype was estimated from a battery of 9 behavioral measures. This latent phenotype was predicted with greater accuracy than any single measure or a latent phenotype by PCA. Together, these results suggest that combining multiple measures of behaviors and connectomes in a principled manner leads to more accurate predictive models. Overall, PCA-CPM provides a unsupervised but easy-to-generalize tool for connectome-based predictions in this scenario while mCCA-CPM is supervised and more accurate within the single dataset.

Chapter 4

Non-linear manifold learning in fMRI uncovers a low-dimensional space of brain dynamics

4.1 Introduction

Understanding large-scale brain dynamics is a major goal of modern neuroscience (Jorgenson et al. [2015]). However, due to the high-dimensional nature of brain patterns, how to best operationalize and tackle this problem remains an open question. Nevertheless, the number of dimensions that explain the observed dynamics is small compared with the number of measurements. Thus, there is growing evidence to suggest that a low-dimensional space—hidden from direct observation, learned from the data, and derived from many brain regions—may be a suitable model for studying brain dynamics (Gao and Ganguli [2015]).

These low-dimensional spaces, also called brain or neural manifolds, have been observed using a variety of neural recordings and animal models (Ahrens et al. [2012], Churchland et al. [2012], Kobak et al. [2016], Mishne et al. [2016], Santhanam et al. [2009]). Research suggests that linear methods, such as principal component analysis (PCA), are appropriate when recorded neural data comes from simple stimuli that project onto a limited area within a manifold (Cunningham and Byron [2014]). However, data from richer tasks often project onto a larger portion of the

manifold, violating linear approximations (Cunningham and Byron [2014], Gallego et al. [2017]). Nonlinear dimensionality reduction methods, like diffusion maps (Coifman and Lafon [2006a]), can overcome this limitation by integrating local similarities into a global representation, which better reflect the underlying brain manifold.

Similar concepts have emerged in human functional magnetic resonance imaging (fMRI) studies to quantify moment-to-moment changes in activity and connectivity (Hutchison et al. [2013], Preti et al. [2017]). As with related research on brain manifolds, dimensionality reduction methods are used to project the fMRI time series onto a low-dimensional space (Allen et al. [2014a], Monti et al. [2017a], Shine et al. [2016]). From the low-dimensional space, characteristic brain states—or distinct, repeatable patterns of brain activity—are used quantify brain dynamics. Predominantly, these studies have relied on linear methods (Allen et al. [2014a], Monti et al. [2017a], Shine et al. [2016]). However, given the rich repertoire of tasks available in human fMRI, a manifold derived from nonlinear methods may better capture the underlying geometry of the low-dimensional space.

To address this, we recently introduced 2-step Diffusion Maps (Gao et al. [2019b]), which are a novel extension of diffusion maps. 2sDM extracts common variability between individuals by performing dimensionality reduction of a 3rd-order tensor in a two-stage manner. In the first stage, timeseries data from each individual are embedded into a low-dimensional Euclidean space. In the second stage, embedding coordinates for the same time point from different individuals are concatenated for use in another embedding. The second stage embeds similar time points across subjects to obtain a low-dimensional group-wise representation of those time points. This two-stage manner avoids directly comparing brain activation across subjects, which can be imprecise without proper alignment (Haxby et al. [2011]). As 2sDM is an unsupervised learning method, no a priori knowledge is needed to handcraft features, which are less robust, computationally intensive, and generalize poorly when compared to learned features from unsupervised methods (Bengio et al. [2013]). The data-driven nature of 2sDM offers complementary information, not only to confirm previous results using handcrafted features, but also to generate new empirical results.

We used 2sDM to embed timeseries from a rich repertoire of tasks onto a single low-dimensional manifold in two fMRI datasets: the Human Connectome Project and the UCLA Consortium for Neuropsychiatric Phenomics. By using multiple tasks spanning a range of cognitive functions and loads, we obtain a more even sampling of the original high-dimensional space of recurring patterns of brain dynamics (Cunningham and Byron [2014], Gallego et al. [2017]) to better project individual time points onto a low-dimensional manifold. Additionally, we embed resting-state data into the same task embedding to investigate differences in brain dynamics between resting-state and task performance. These results suggest that manifold learning can uncover an interpretable low-dimensional embedding for the study of brain dynamics in fMRI data.

4.2 Methods

4.2.1 Diffusion maps

Diffusion maps Coifman and Lafon [2006b] is part of a broad class of manifold learning algorithms. Specifically, diffusion maps provides a global description of the data by considering only local similarities and is robust to noise perturbation. The new nonlinear representation provided by diffusion maps reveals underlying intrinsic parameters governing the data Nadler et al. [2006]. Here we develop a new framework utilizing diffusion maps to detect repeatable brain states in fMRI data.

The diffusion maps algorithm is as follows. The input is a pairwise similarity matrix $\mathbf{S}$, which can be computed using the Gaussian kernel $w_\epsilon(x, y) = \exp(-||x - y||^2/\epsilon)$ between pairs of data points x and y. Then the rows of the similarity matrix are normalized by $\mathbf{P} = \mathbf{D}^{-1}\mathbf{S}$, where $D_{ii} = \sum_j S_{ij}$. This creates a random walk matrix on the data with entries set to $p(x, y) = w_\epsilon(x, y)/d(x)$. Taking powers of the matrix is equivalent to running the Markov chain forward. The kernel $p_t(\cdot, \cdot)$ can be interpreted as the transition probability between two points in t time steps. The matrix $\mathbf{P}$ has a complete sequence of bi-orthogonal left and right eigenvectors $\boldsymbol{\phi}_i$, $\boldsymbol{\psi}_i$, respectively, and

Algorithm 2 Diffusion Maps

Input: $\mathbf{X} \in \mathbb{R}^{N \times P}$ - N instances with P features
d - number of dimensions to keep in the embedding
t - diffusion time parameter
Output: $\mathbf{\Psi} \in \mathbb{R}^{N \times d}$ - d-dimensional embedding
function DM($\mathbf{X}, d, t$)
Step 1: Build similarity matrix $\mathbf{L}$ using Gaussian kernel $w_\epsilon(x, y) = e^{-||x-y||^2/\epsilon}$
Step 2: Normalize the matrix $\mathbf{L}$ to approximate the Laplace–Beltrami operator $\tilde{\mathbf{L}} = \mathbf{D^{-1}LD^{-1}}$, where D is a diagonal matrix and $D_{ii} = \sum_j L_{ij}$
Step 3: Form the normalized random walk matrix $\mathbf{M} = \tilde{\mathbf{D}}^{-1}\tilde{\mathbf{L}}$, where $\tilde{\mathbf{D}}$ is a diagonal matrix and $\tilde{D}_{ii} = \sum_j \tilde{L}_{ij}$
Step 4: Compute the largest d eigenvalues λ_i of $\mathbf{M}$ and the corresponding eigenvectors $\boldsymbol{\psi}_i$, $\mathbf{\Psi}(\mathbf{X}) = (\lambda_1^t\boldsymbol{\psi}_1, \lambda_2^t\boldsymbol{\psi}_2, \ldots, \lambda_d^t\boldsymbol{\psi}_d)$
=0

a corresponding sequence of eigenvalues $1 = \lambda_0 \geq |\lambda_1| \geq |\lambda_2| \geq \ldots$. Diffusion maps is a nonlinear embedding of the data points into a low-dimensional space, where the mapping is defined as $\mathbf{\Psi}(x) = (\lambda_1^t\boldsymbol{\psi}_1(x), \lambda_2^t\boldsymbol{\psi}_2(x), \ldots, \lambda_k^t\boldsymbol{\psi}_k(x))$, where t is the diffusion time. Note that $\boldsymbol{\psi}_0$ is neglected because it is a constant vector.

A diffusion distance $D_t^2(x, y)$ between two data points x, y is defined as:

$$D_t^2(x, y) = \sum_z \frac{(p_t(x, z) - p_t(y, z))^2}{\phi_0(z)}$$

where $\boldsymbol{\phi}_0$ represents the stationary distribution. This measures the similarity of two points by the evolution in the Markov chain. Two points are closer if there are more short paths connecting them. It is thus robust to noise as it considers all the possible paths between two points.

Proposition 1 (Coifman & Lafon)**.** *Diffusion maps* $\mathbf{\Psi}$ *embeds data points into a Euclidean space* $\mathbb{R}^k$ *where the Euclidean distance approximates the diffusion distance:*

$$D_t^2(x, y) = ||\mathbf{\Psi}(x) - \mathbf{\Psi}(y)||_2^2$$

A detailed proof using the spectral theorem in Hilbert space can be found in Coifman and Lafon [2006b]. In practice, eigenvalues of $\mathbf{P}$ typically exhibit a spectral gap such that the first

few eigenvalues are close to one with all additional eigenvalues much smaller than one. Then the diffusion distance can be well approximated by only these first few eigenvectors (Nadler et al. [2006]). Thus, we obtain a low-dimensional representation of the data by considering only the first few eigenvectors of the diffusion maps (See Algorithm 2). Intuitively, diffusion maps embeds data points closer when it is hard for the points to escape the local region within time t.

To remove dependence on the density of the data, the Gaussian similarity weights $w_\epsilon(\cdot,\cdot)$ are renormalized by the estimated density. This renormalization step leads to an anisotropic diffusion process and enables the algorithm to better recover the manifold structure so that it does not depend on the distribution of the points. The eigenvectors of the new random walk matrix now approximate the Laplace-Beltrami operator (Nadler et al. [2006]).

Diffusion maps is similar to the normalized cuts algorithm (Shi and Malik [2000]) which has previously been used in fMRI analysis (Shen et al. [2013b]). Normalized cuts aims to find the eigendecomposition of $D^{-1}L$ where L is the Laplacian matrix $L = D - S$. The eigendecomposition of $D^{-1}L$ yields the same eigenvectors ψ as for diffusion maps, with corresponding eigenvalues $1 - \lambda$. Thus, performing *k*-means clustering on diffusion maps coordinates as we do below is mathematically similar to spectral clustering. The key difference is that in diffusion maps the coordinates are weighted by the corresponding eigenvalues.

4.2.2 2-step Diffusion maps

Based on diffusion maps, we propose a hierarchical manifold learning framework for multi-individual fMRI BOLD time series. The framework is illustrated in Figure 4.1a. Under the assumption that individuals' fMRI responses are time-synchronized, we represent each individual's fMRI data as $\mathbf{X}_{i,:,:} \in \mathbb{R}^{T\times V}, i = 1,\ldots,M$. Here T is the number of repetition time (TR) in the scan, V is the number of voxels or brain regions, and M is the total number of individuals. We label this framework 2-step diffusion maps (2sDM). Note that 2sDM can be applied to either domains of the data, resulting in a lower-dimensional representation of either time, individuals or brain regions.

Here we illustrate the framework by embedding time into a lower-dimensional space. Reducing the other two domains just requires trivial adaptation.

First we apply diffusion maps to the fMRI time series of every single individual i to obtain the embedding $\mathbf{\Psi}_i \in \mathbb{R}^{T \times d_1}$, thus reducing each individual's V voxels or brain regions to a d_1-dimensional Euclidean space. Then, we concatenate the new representations of all individuals to a single matrix $\mathbf{\Psi}^{(1)} \in \mathbb{R}^{T \times (Md_1)}$. From this concatenated matrix, we can perform a second-step diffusion maps to further reduce the dimensionality of every time-frame to d_2. The final time-frame representation matrix with multi-individual similarity is $\mathbf{\Psi}^{(2)} \in \mathbb{R}^{T \times d_2}$. Our framework is presented in Algorithm 3. As the first-step diffusion maps produces a cleaner representation of the fMRI data, the reasoning of performing an embedding based on the results of the first-step embedding can be seen from the following proposition,

Proposition 2. *The distance between two frames u, v in $\mathbf{\Psi}^{(1)}$ equals the total diffusion distance for all individuals.*

Proof. By Proposition 1, $||\mathbf{\Psi}(x) - \mathbf{\Psi}(y)||_2 = D_t(x, y)$. Therefore

$$||\mathbf{\Psi}^{(1)}(u) - \mathbf{\Psi}^{(1)}(v)||_2^2 = \sum_{i=1}^{M} ||\mathbf{\Psi}_i(u) - \mathbf{\Psi}_i(v)||_2^2 = \sum_{i=1}^{M} D_t^2(\mathbf{X}_{i,u,\cdot}, \mathbf{X}_{i,v,\cdot}),$$

□

where $D_t^2(\mathbf{X}_{i,u,\cdot}, \mathbf{X}_{i,v,\cdot})$ is the diffusion distance of time points u and v for individual i with diffusion time t.

As such, if two concatenated vectors have relatively small Euclidean distance, it suggests that, on average, for all of the individuals there is small diffusion distance between the two time points. Additionally, no functional alignment between individuals is needed as the similarity between time points is calculated in each individual's own embedding space separately and aggregated through the sum of diffusion distances.

Algorithm 3 2-step Diffusion Maps

Input: $\mathbf{X} \in \mathbb{R}^{M \times T \times V}$ - M individuals' fMRI time series with T TRs and V regions
d_1 - number of dimensions to keep for each individual
d_2 - number of dimensions to keep in the final embedding
t - diffusion time parameter
Output: $\mathbf{\Psi}^{(2)} \in \mathbb{R}^{T \times d_2}$ - lower-dimensional embedding for the second dimension
function 2sDM($\mathbf{X}, d_1, d_2, t$)
for each individual $\mathbf{X}_{i,\cdot,\cdot}$ **do**
$\mathbf{\Psi}_i$=DM($\mathbf{X}_{i,\cdot,\cdot}, d_1, t$)
$\mathbf{\Psi}^{(1)}$=($\mathbf{\Psi}_1, \mathbf{\Psi}_2, \ldots, \boldsymbol{\psi}_M$)
$\mathbf{\Psi}^{(2)}$=DM($\mathbf{\Psi}^{(1)}, d_2, t$)

4.2.3 Out-of-sample extension framework

To embed new time points to the existing temporal manifold, we use out-of-sample extension (OOSE) for new time series data. The reason to use OOSE here is twofold: *(i)* OOSE enables to embed new data points without reapplying the eigendecomposition on the entire dataset, and *(ii)* OOSE keeps the existing manifold structure unaffected and makes it easier to interpret new time points in an unsupervised setting. Figure 4.1b. The OOSE framework for time-synchronized fMRI data works in a similar hierarchical way, using two Nyström extension steps (Algorithm 4). Nyström extension was a nonparametric method to extend the embedding learned with training dataset to unseen data points. It was based on Nyström method to extend eigenvector computed for a set of sample points to an arbitrary point. Suppose we have a kernel function $K(a, b)$ that generated a symmetric matrix M with entries $M_{ij} = K(x_i, x_j)$ upon a training dataset $D = x_1, \ldots, x_n$. Let (v_k, λ_k) be an (eigenvector, eigenvalue) pair that solves $\mathbf{M}v_k = \lambda_k v_k$. The k-th coordinate of diffusion maps embedding was then $\lambda_k v_k$. Let $y_k(x)$ denote the k-th diffusion maps embedding associated with a new point x, then $y_k(x) = \frac{1}{\lambda_k} \sum_{i=1}^{n} v_{ki} K(x, x_i)$.

Given new time-synchronized fMRI data $\mathbf{X}'_{i,\cdot,\cdot} \in \mathbb{R}^{T' \times V}, i = 1, ..., M$ for the same group of individuals, we first approximate eigenvectors $\hat{\mathbf{\Psi}}_i^{(1)}$ for each individual. Then we concatenate all the individuals' eigenvectors $\hat{\mathbf{\Psi}}_i^{(1)}$ as the new data points and approximate its eigenvectors $\hat{\mathbf{\Psi}}^{(2)}$ as the final representation. The 2-step OOSE framework is described in Algorithm 5.

Algorithm 4 Nyström Out-of-sample Extension

Input: $\mathbf{X} \in \mathbb{R}^{N \times P}$ - training data
$\mathbf{\Psi} \in \mathbb{R}^{N \times d}$ - d-dimensional embedding result for $\mathbf{X}$
$\mathbf{X}' \in \mathbb{R}^{N' \times P}$ - N' new data points
Output: $\hat{\mathbf{\Psi}} \in \mathbb{R}^{N' \times d}$ - approximated low-dimensional embedding for $\mathbf{X}'$
function OOSE($\mathbf{X}', \mathbf{X}, \mathbf{\Psi}$)
$\hat{\psi}_l(x') = \frac{1}{\lambda_l} \sum_{j=1}^{m} p(x', x_j)\psi_l(x_j), l = 1, \ldots, d$
$\hat{\mathbf{\Psi}}(\mathbf{X}') = (\lambda_1^t \hat{\boldsymbol{\psi}}_1, \lambda_2^t \hat{\boldsymbol{\psi}}_2, \ldots, \lambda_d^t \hat{\boldsymbol{\psi}}_d)$

Algorithm 5 2-step Out-of-sample Extension

Input: $\mathbf{X} \in \mathbb{R}^{M \times T \times V}$ - M individuals' fMRI time series with T TRs and V regions
$\mathbf{\Psi}^{(1)} \in \mathbb{R}^{T \times M \times d_1}$ - first-step diffusion maps result for $\mathbf{X}$
$\mathbf{\Psi}^{(2)} \in \mathbb{R}^{T \times d_2}$ - second-step diffusion maps result for $\mathbf{\Psi}^{(1)}$
$\mathbf{X}' \in \mathbb{R}^{M \times T' \times V}$ - M individuals' new fMRI time series with T' TRs and V regions
Output: $\hat{\mathbf{\Psi}}^{(2)} \in \mathbb{R}^{T' \times d_2}$ - approximated low-dimensional embedding for $\mathbf{X}'$
function 2-STEP OOSE($\mathbf{X}', \mathbf{\Psi}^{(1)}, \mathbf{\Psi}^{(2)}, \mathbf{X}$)
for each individual $\mathbf{X}'_{i,\cdot,\cdot}$ **do**
$\hat{\mathbf{\Psi}}_i^{(1)}(\mathbf{X}'_{i,\cdot,\cdot}) =$OOSE$(\mathbf{X}'_{i,\cdot,\cdot}, \mathbf{X}_{i,\cdot,\cdot}, \mathbf{\Psi}^{(1)}_{\cdot,i,\cdot})$
$\hat{\mathbf{\Psi}}^{(1)}(\mathbf{X}') = (\hat{\mathbf{\Psi}}_1^{(1)}(\mathbf{X}'_{1,\cdot,\cdot}), \hat{\mathbf{\Psi}}_2^{(1)}(\mathbf{X}'_{2,\cdot,\cdot}), \ldots, \hat{\mathbf{\Psi}}_M^{(1)}(\mathbf{X}'_{M,\cdot,\cdot}))$
$\hat{\mathbf{\Psi}}^{(2)}(\mathbf{X}') =$OOSE$(\hat{\mathbf{\Psi}}^{(1)}, \mathbf{\Psi}^{(1)}, \mathbf{\Psi}^{(2)})$

fMRI data that is not time-synchronized across individuals (*e.g.*, rsfMRI) needs to be synchronized across individuals before an out-of-sample application of the existing 2sDM embedding can be used. We used Brainsync (Joshi et al. [2018]) to temporally synchronize the rsfMRI data. Brainsync synchronizes one individual's time series data $\mathbf{Y} \in \mathbb{R}^{T \times V}$ to a reference individual's time series $\mathbf{X} \in \mathbb{R}^{T \times V}$ by finding an optimal orthogonal transformation $\mathbf{O}^s$ for $\mathbf{Y}$ to minimize the squared error: $\mathbf{O}^s = \arg\min_{\mathbf{O} \in \mathbf{O}(T)} ||\mathbf{X} - \mathbf{O}\mathbf{Y}||^2$. The problem can be solved by the Kabsch algorithm (Kabsch [1976]). The $T \times T$ cross-correlation matrix $\mathbf{X}\mathbf{Y}^t$ is first formed and its singular value decomposition can be calculated as $\mathbf{X}\mathbf{Y}^t = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^t$. The optimal $\mathbf{O}^s$ can be found by $\mathbf{O}^s = \mathbf{U}\mathbf{V}^t$ and $\mathbf{Y}$ can be synchronized to $\mathbf{X}$ by $\mathbf{O}^s\mathbf{Y}$.

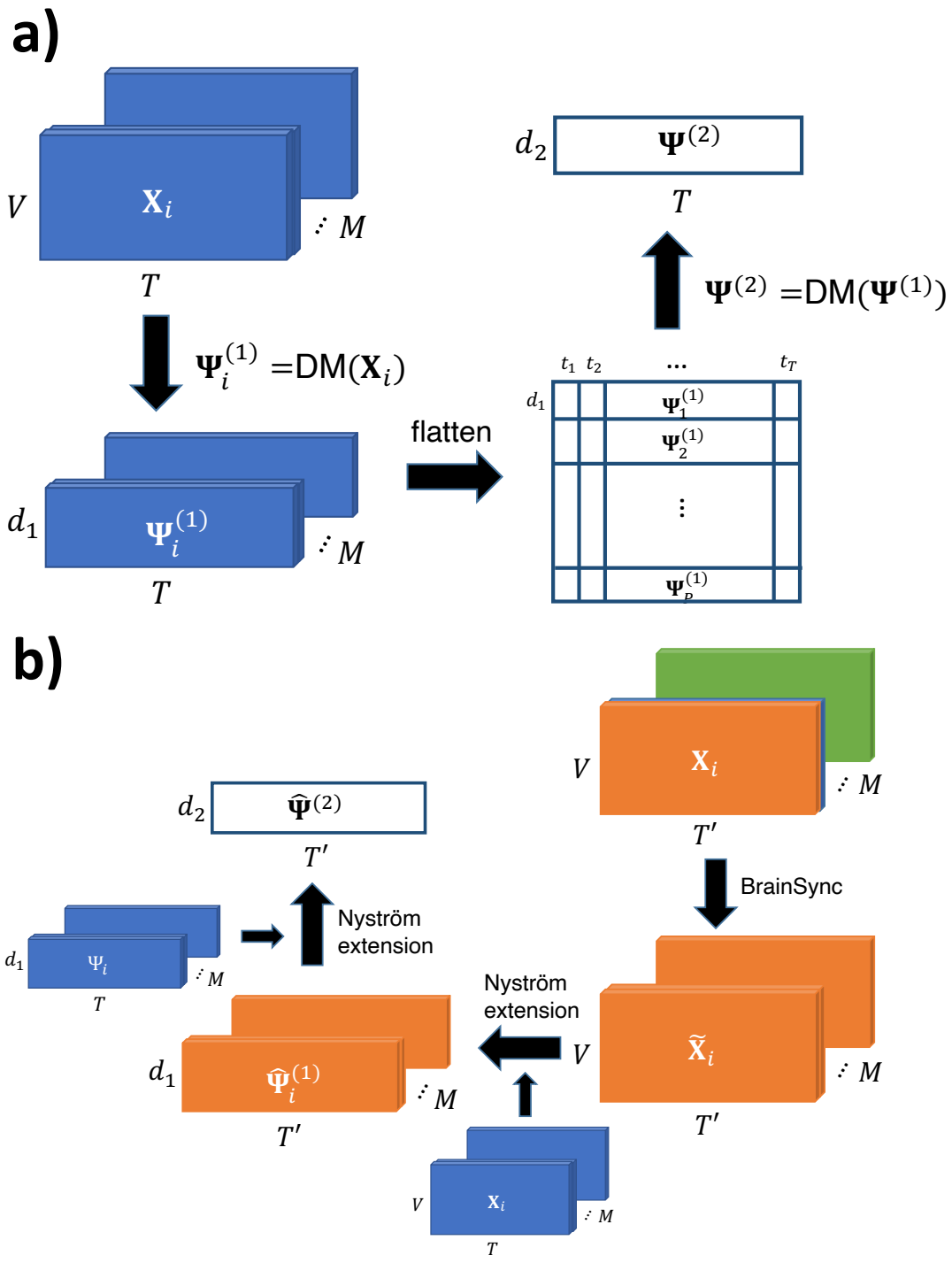


Figure 4.1: **Schematic of 2sDM manifold learning framework a)** 2sDM algorithm framework for time-synchronized multi-individual fMRI time series. **b)** 2-step out-of-sample extension framework with BrainSync for new fMRI time points. Mathematical notations in the figure are the same as those used in the corresponding Methods section.

4.2.4 Dynamic connectivity

To relate our task embedding to previously used handcrafted features (Shine et al. [2016]), we calculated B_T using sliding-window-based functional connectivity, as described in previous literature (Shine et al. [2016]). In this manuscript, handcrafted features referred to features that were chosen manually like B_T that was used here to character the integration and segregation pattern of the brain graph. B_T was averaged across all subjects. To compare our results with prior literature (Shine et al. [2016]), we calculated the dynamic functional connectivity using the multiplication of temporal derivatives (MTD; Shine et al. [2015]). MTD is calculated as the point-wise product of the temporal derivatives of paired nodes' time series:

$$MTD_{ijt} = \frac{1}{w} \sum_{t}^{(t+w)} \frac{(dt_{it} \times dt_{jt})}{\sigma_{dt_{it}} \times \sigma_{dt_{jt}}}$$

At each time point, the dynamic functional connectivity is calculated as the averaged MTD over a sliding time window in order to reduce high-frequency noise. We chose the length of the sliding window to be 15 time points, based on previous literature (Shine et al. [2016]). The participation coefficient B_T characterizes the extent to which a region connects across all modules, where modules were normally defined a priori from community detection methods that find a set of nodes as a module that were strongly connected to each other than nodes from another set. The participation coefficient for a region i at time T is calculated as:

$$B_{iT} = 1 - \sum_{s=1}^{N_M} (\frac{k_{isT}}{k_{iT}})^2$$

where k_{isT} is the number of links of node i to nodes in module s at time T, k_{iT} is the total degree of node i at time T and N_M is the number of modules, or canonical networks in our setting. The participation coefficient of a region is therefore close to 1 if its links are uniformly distributed among all the modules and 0 if all its links are within its own module. The whole brain participation

coefficient B_T represents the average of B_{iT} from each region and thus represents the integration and segregation pattern of the brain. B_T is closer to 1 if our whole brain is more integrated and closer to 0 if our whole brain is more segregated.

4.2.5 Characterizing changes in brain states

By utilizing the temporal order of time points, we characterized the brain dynamics across the four brain states by state transition probability and dwell time. State transition probabilities are calculated based on the temporally adjacent time points' brain states. From these state transition probabilities, a stochastic matrix and the dwelling times (i.e. the stationary probability distribution of the stochastic matrix) were calculated and visualized as Markov chain models. The stationary distribution of the Markov transition matrix is defined as the distribution that does not change under application of the transition matrix $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, which is the left eigenvector of $\mathbf{P}$. It represents the distribution to which the Markov process converges. It was used in our experiment to represent the dwell-time distribution of discrete brain states. As tasks putatively put a participant into certain states (as opposed to the unconstrained nature of the resting state), we investigated differences in the temporal dynamics of state switching during task and rest. We calculated entropy—a measure of the randomness—of the transition probability from one brain state to the other states. Entropy of a discrete probability distribution measures the uncertainty of the outcome. It is calculated as the negative expectation of the logarithm of the probability mass function's value $S = -\sum_i p_i \log p_i = -E_p[\log p]$ In our experiment, entropy of the brain state transition probability was used to assess the randomness of brain state transitioning with lower entropy representing more easy-to-predict brain state transition dynamics. Greater entropy indicates a less predictable transition from one state to another.

4.2.6 Dataset and imaging parameters

Data was obtained from the Human Connectome Project (HCP) 900 Subject release (Van Essen et al. [2013b]). We used fMRI data collected while 390 participants performed six tasks (gambling, motor, relational, social, working memory—WM, and emotion). We restricted our analyses to those subjects who participated in all nine fMRI conditions (seven task, two rest), whose mean frame-to-frame displacement was less than 0.1mm and whose maximum frame-to-frame displacement was less than 0.15mm, and for whom the task block order are the same as other subjects (n=390). All fMRI data were acquired on a 3T Siemens Skyra using a slice-accelerated, multiband, gradient-eco, echo planar imaging (EPI) sequence (TR=720ms, TE=33.1ms, flip angle=52°, resolution=2.0mm3, multiband factor=8). Images acquired for each subject include a structural scan and eighteen fMRI scans (working memory (WM) task, incentive processing (gambling) task, motor task, language processing task, social cognition task, relational processing task, emotion processing task, and two resting-state scans; two runs per condition (one LR phase encoding and one RL phase encoding run)) split between two sessions.

The UCLA Consortium for Neuropsychiatric Phenomics (Poldrack et al. [2016b]) dataset was used for replication. Similar to the standards for the HCP dataset, we restricted our analyses to those subjects who participated in all 5 fMRI conditions (four task, one rest), whose mean frame-to-frame displacement was less than 0.1mm and whose maximum frame-to-frame displacement was less than 0.15mm. 77 healthy controls were retained. These participants performed four tasks (paired memory retrieval task—PAMRET, paired memory encoding task—PAMENC, spatial working memory task—SCAP, task switching task—TASKSWITCH). Details of the image acquisition parameters have been published elsewhere (Poldrack et al. [2016b]). In brief, all data were acquired on one of two 3T Siemens Trio scanners at UCLA. Functional MRI data were collected using a T2*-weighted EPI sequence with the following parameters: slice thickness=4mm, 34 slices, TR=2s, TE=30ms, flip angle=90°, matrix 64×64, FOV=192mm, oblique slice orientation. Images acquired for each subject include a structural scan and seven fMRI scans (balloon

analog risk task (BART), paired-associate memory retrieval (PAMRET), paired-associate memory encoding (PAMENC), spatial capacity task (SCAP), stop signal task (SST), task-switching task (TASKSWITCH) and breath holding task).

As 2sDM requires time series to be synchronized across individuals (i.e., different individuals encounter the same task condition at the same time point), the language task from the HCP and the stop signal task, balloon analogue risk task, and breath hold task from the CNP were not included. These tasks are self-paced. Participants finished blocks at different times, causing the task block to be unsynchronized across participants.

4.2.7 fMRI processing

For the HCP dataset, the HCP minimal preprocessing pipeline was used (Glasser et al. [2013]), which includes artifact removal, motion correction, and registration to standard space. For the CNP dataset, structural scans were skull-stripped (Lutkenhoff et al. [2014]) and registered to the MNI template using a validated algorithm Scheinost et al. [2017]. Slice time and motion correction were performed in SPM8. For both datasets, all subsequent preprocessing was performed using image analysis tools available in BioImage Suite (Joshi et al. [2011]) and included standard preprocessing procedures (Finn et al. [2015a]). These procedures included removal of motion-related components of the signal, regression of mean time courses in white matter, cerebrospinal fluid, and gray matter, removal of the linear trend, and low-pass filtering. Mean frame-to-frame displacement yielded seven motion values per subject, which were used for subject exclusion and motion analyses. We restricted our analyses to subjects whose maximum frame-to-frame displacement was less than 0.15mm and mean frame-to-frame displacement was less than 0.1mm. This conservative threshold for exclusion due to motion was used to mitigate the effect of motion on the embedding. We used the Shen 268-node atlas to extract timeseries from the fMRI data for further analysis (Shen et al. [2013a]). Timeseries used for the embedding were the average of the basis of the "raw" task time courses, with no removal of task-evoked activity, for each node in the atlas. Finally, 2sDM

was applied to embed a 3rd-order tensor of fMRI data (individual $\times$ region $\times$ time) onto a single low-dimensional manifold.

4.3 Results

4.3.1 Brain dynamics during tasks embed onto a low-dimensional space

Although each task is different in many ways, individual time points in the fMRI data from all tasks mapped onto a single low-dimensional manifold (Figure 4.2a). Compared with the common goal of other low-dimensional embedding results, the advantage of our results was not in separating different task scans apart. Instead, we aim to find a global representation across multiple tasks that positioned tasks with similar cognitive loads together. By embedding multiple tasks together, rather than in isolation, the closeness of different blocks and tasks in the manifold suggest that similar, recurring patterns of brain dynamics exist across a variety of tasks. For example, in the manifold, the 2-back blocks of the WM task were significantly ($t = 201.9, p < 0.01, df = 175, 102$) closer to time points from the gambling task (Euclidean distance: 0.0258 ± 0.0096) than the 0-back blocks of the WM task (Euclidean distance: 0.0355 ± 0.0100), despite the fact that the 2-back and 0-back blocks were collected in the same fMRI run. The 2-back blocks of the WM task and the gambling task both entail a higher cognitive load. In contrast, the 0-back blocks of WM task overlap with the motor task. These tasks are simpler response tasks and less cognitively demanding. Overall, these time points are positioned based on the similarity of the cognitive load at that time point, instead of by task.

For all tasks, the average trajectories from each task are found to start near the corner where cues (task cues preceding each task block) reside and end in the other corner where fixation blocks reside (Figure 4.2b). These smooth trajectories indicate that the embedding preserves proper temporal associations between blocks when arranging time points in discrete states. As can be expected, the paths of these temporal trajectories depend on the cognitive load of the task block. For

example, the 2-back task traversed through the upper part of the manifold (higher value in terms of $\mathbf{\Psi}_3$), and, in contrast, the 0-back task traversed through the lower part of the manifold (Figure 4.2c).

Moreover, as can be seen from the top 20 eigenvalues of the diffusion matrix the spectrum decays rapidly, which suggests that the data is low-dimensional (Figure 4.3).

When projecting task fMRI time frames into 3D space using the 1st three coordinates of PCA, no clear structure is shown from the embedding (Figure 4.4). The fact that 2sDM discovered the manifold structure, while PCA could not, validates the usage of nonlinear manifold learning (more detailed comparison between 2-step PCA and 2sDM embeddings are included in the supplementary materials).

4.3.2 Task embedding captures handcrafted features in an unsupervised manner

In each time point in our task embedding is colored by its subject-averaged B_T, showing a clear pattern of decreasing B_T starting from the top left corner of the embedding; higher B_T at the top of the embedding (i.e., high cognitive load tasks such as social, 2back, relational and gambling) indicates time points of higher integration and lower B_T at the tails of the embedding (i.e., cues and fixations) indicates time points of higher segregation ($r(z, B_T) = 0.610, df = 3018, p < 0.01$, where z is the projection coordinates of points onto the diagonal of the triangular embedding; Figure 4.5b).

4.3.3 Operationalizing discrete, recurring brain states from task dynamics

When clustering the task embedding, $k = 4$ gave the largest Calinski-Harabasz score among a range, suggesting that the embedding has a clear interpretable structure (Figure S6). Based on the task contents of the temporal clusters, we labeled the four brain states as: fixation, transition, lower-level cognition, and higher-level cognition. Functionally reasonable patterns of activation

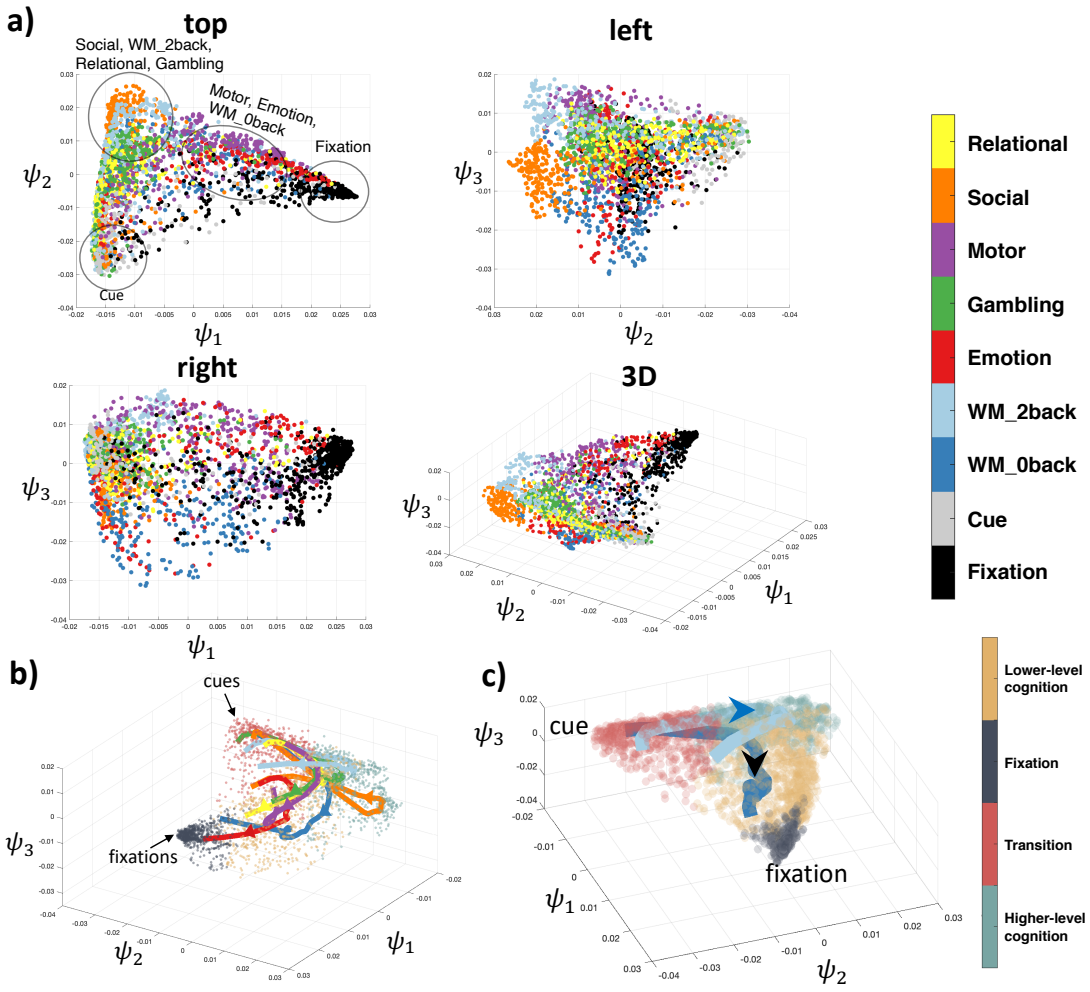


Figure 4.2: **Nonlinear embedding of fMRI time series data a)** 2sDM embedding of 6 tasks (Relational, Social, Motor, Gambling, Emotion, Working Memory 2-back, and Working-Memory 0-back) from the HCP dataset. Four different views of the manifold are shown. Each point in these subplots represents a single time point and is colored by the task type. **b)** Averaged temporal trajectory of each task with the embedding colored by the corresponding brain state as the background. **c)** WM task's 0-back and 2-back task blocks visualized separately with major cues and fixations points annotated. Arrows show the progression direction of the trajectory. Trajectory in b) and c) uses the same colormap as a).

during the different states are observed, e.g., canonical patterns of default mode network activity for the fixation state (Figure 4.6a). To relate these brain states to previous handcrafted features, we calculated the average B_T for each brain state (Figure 4.6b). The four states followed the expected patterns of integration and segregation, with the higher-level cognition state showing the greatest integration ($t = 3.01, p < 0.01, df = 1596$) and the fixation state showing the greatest segregation

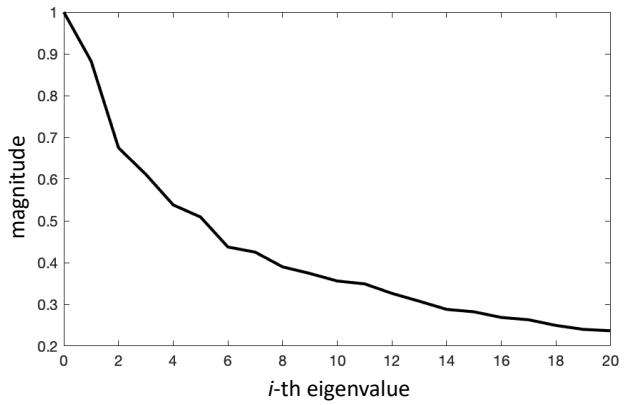


Figure 4.3: **The top 20 eigenvalues of the diffusion matrix for the fMRI data.** The spectrum decays rapidly, suggesting that the data is low-dimensional.

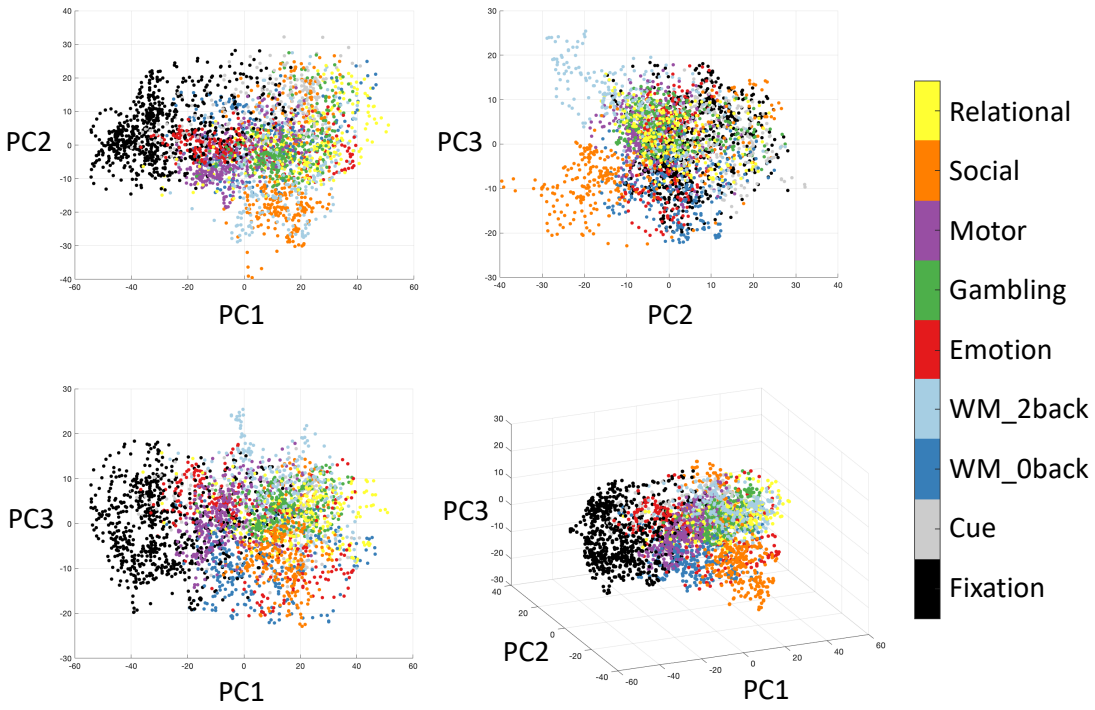


Figure 4.4: 2-step PCA embedding from the HCP dataset. Unlike the nonlinear embeddings, shown in Figure 2, no clear structure is seen for the linear embedding, which validates the usage of nonlinear manifold learning.

($t = 2.39, p < 0.01, df = 1420$). The clustering results are similar with an increased number of clusters or of embedding dimensions.

With the help of the four brain states, the dynamic trajectories can further reveal each task's cognitive process (Figure 4.6c). For example, the motor task's trajectory reveals a dynamic cogni-

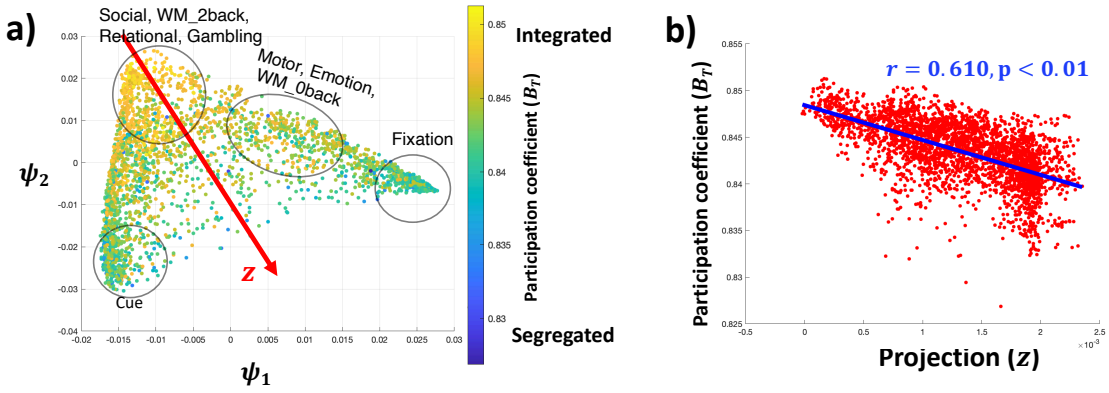


Figure 4.5: **2sDM embedding is related with global integration and segregation.** **a)** 2sDM embedding in HCP dataset colored by the time-resolved B_T. **b)** Scatter plot of the B_T with the projection onto the diagonal of the embedding structure (z). Correlation of z with B_T is shown with a line of best fit. Projection direction z was determined manually as the approximate diagonal direction of the embedding.

tive process as following: in the beginning, the individuals start from the cue state which was the common starting state across the other tasks. Then the individuals briefly enter the high-cog state, but not deep in the state and finally enter and stay in the low-cog state. Actually, it also reveals that on average, individuals wander towards the fixation state in the middle of the task block, suggesting a fatigue or practice effect. And towards the end of the task block, individuals return deep into the low-cog state and moved towards the cue state for the next task block to start.

Even for tasks like relational and social tasks that both require a certain level of high-level cognitive ability (Shine et al. [2016]), there are differences that can be revealed by the trajectories (Figure 4.6c). The relational task starts from the transition cluster, then entered the higher-level cognition cluster and ends in the low-cog state, which suggests a lack of high-level cognitive ability involvement (adaptive to the task design) in the later stage of the relational task blocks. In comparison, the social task starts near the transition cluster, goes deep into the high-cog state and returns to the transition state near the end of the task which suggests a constant requirement of higher-level cognitive ability. This trajectory view of each task enables a better understanding of the cognitive process and can also help in the future task designs.

The transitions between states were similar for all tasks except for the motor task (which had a high probability of transiting into the lower-level cognition state and out of the higher-level cognition state; Figure 4.7a). Except for the WM task, which contains an equal proportion of high (2-back) and low (0-back) cognitive loads), dwell times for the four states exhibited a non-uniform distribution ($\chi^2 > 16.3, df = 3, p < 0.001$; Figure 4.7b), indicating participants spent most of their time in certain limited states in a task-specific manner. For example, the lower-level cognition state occurred most frequently in the motor task, while the higher-level cognitive state dominated in social task time points.

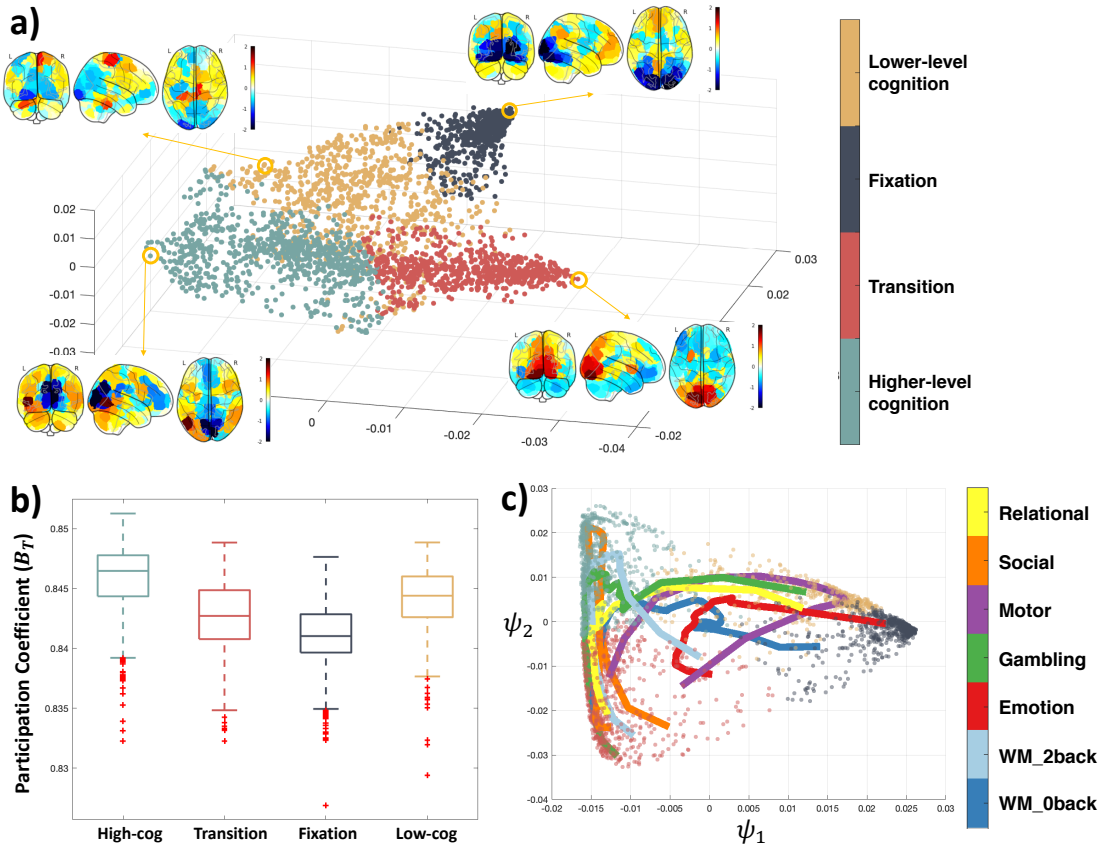


Figure 4.6: **Brain states during tasks. a)** K-means clustering of the task manifold. Averaged brain activation patterns across subjects in the circled representative time points are shown for each brain state. **b)** B_T averaged over all the time points in each brain state. **c)** Two-dimensional view of task trajectories with the embedding points. Trajectories are colored by each task and data points are colored by the brain states as in a).

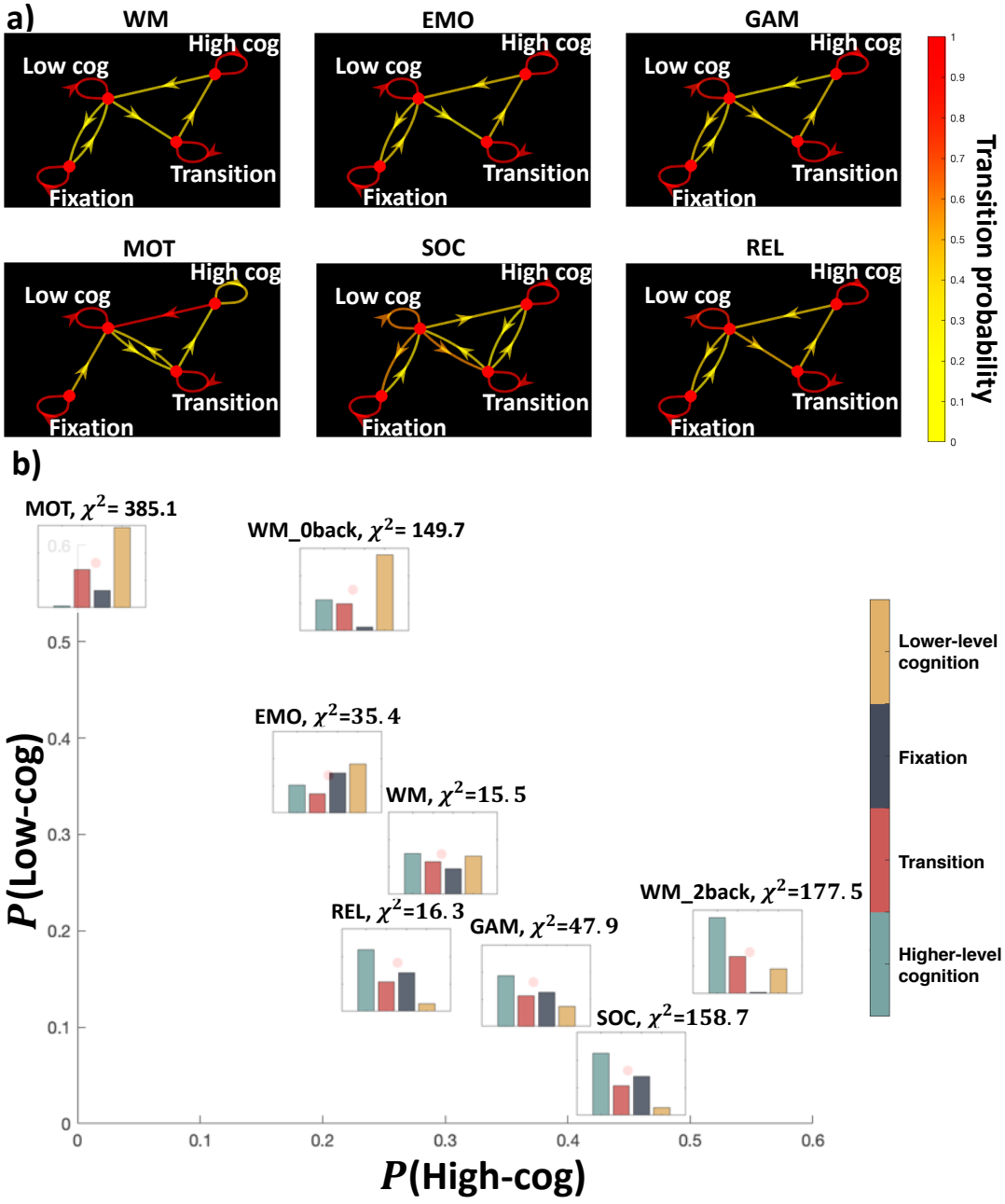


Figure 4.7: **Brain state dynamics differ between tasks. a)** Brain state dynamics visualized as the Markov chain. Transition probability is visualized by the color of the directed edges. **b)** Stationary distribution probability visualized for each task and positioned by the proportion of higher-level cognition and lower-level cognition brain states. Chi-square test result against the uniform distribution is also shown.

4.3.4 Brain dynamics during rest embed onto the same recurring brain states which appeared during tasks

Once embedded onto the task manifold, time points from the resting-state data spread across the whole manifold, including parts of the manifold corresponding to higher cognitive loads (Figure

4.8a). To quantify the distribution of states during rest, we assigned each resting-state time point to one of the four previously identified brain states based on the brain state of the nearest task time point. As with the task data, we next calculated the brain state dwell time distribution across the entire resting-state scan (Figure 4.8b). A non-uniform dwell-time distribution was discovered, with fixation and transition states having a higher proportion of time points than the cognitive states ($\chi^2 = 205, df = 3, p < 0.001$). Except for the lower-level cognition and the transition states in the social task (which have very few time points to robustly calculate entropy, see Figure 4.8c), all states exhibited higher entropy in the resting state than during a given task.

In Figure 4.9, we plot the extension of the WM task. The 2-back and 0-back task blocks go to the correct higher-level cognition or lower-level cognition state respectively, while the fixation and cue time frames are also located in the correct brain states. The correlation between the extended coordinates and the coordinates from the original embedding was highly significant ($r = 0.939, p < 0.001$). Holding out the other tasks produced similar results as the WM task.

4.3.5 Differences in brain dynamics in patients with schizophrenia

Notably, we replicated the dimensionality reduction result using participants from the CNP dataset. A similar low-dimensional structure, brain states, and association with B_T ($r(\psi_2, B_T) = 0.30, p < 0.01, df = 1007$) were found, verifying the robustness of the observed embeddings (Figure 4.10). Moreover, the same task scans from the schizophrenia cohorts were also embedded separately and found to be similar to the embedding from the HCP dataset and healthy control cohorts in the CNP dataset (Figure 4.10b). This laid foundation for the downstream brain dynamics analysis (resting-state brain dynamics) that would be based on brain states as similar brain states could be identified in both groups.

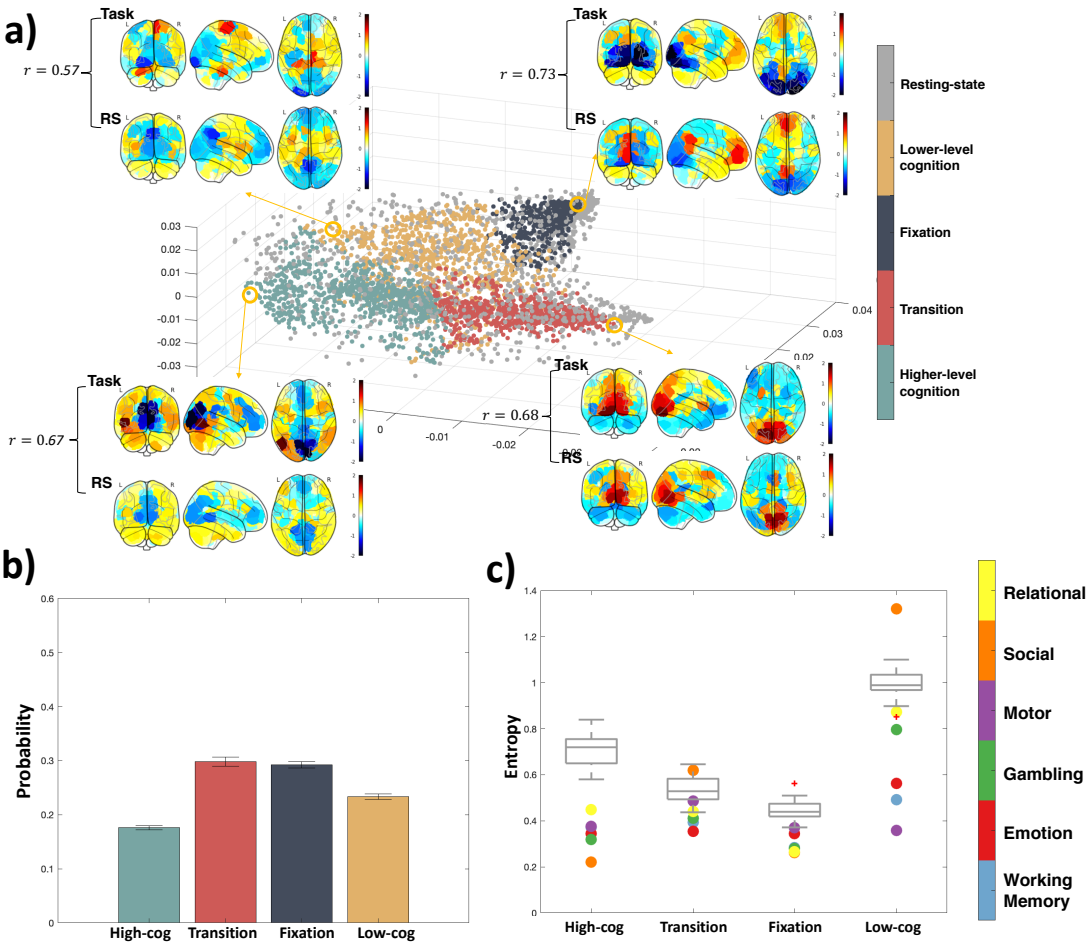


Figure 4.8: **Resting-state extended onto the task manifold.** **a)** Representative task activation patterns of each state and the neighboring resting-state activation pattern are visualized. Correlation of the activation between task and rest is calculated with higher correlation representing more accurate out-of-sample extension. **b)** Stationary probability distribution of the four brain states during resting state. **c)** Entropy of each brain state's transition probability in different tasks. Dots are colored by tasks they represent, and the grey box plot shows the entropy values of resting state with BrainSync (see Methods) referenced to different individuals.

4.3.6 Comparison of 2-step Diffusion maps and 2-step PCA

Although our 2sDM framework was not tight to only the diffusion maps algorithm, we compare the 2sDM results with its linear comparison, 2-step PCA (2sPCA) in this subsection.

Low-dimensional embedding Unlike the low-dimensional embedding works that only involved two or three tasks, in this work, as we tried to embed multiple different tasks (6 tasks in the HCP dataset) together, our goal is no longer separating different tasks. Instead, to cluster different tasks

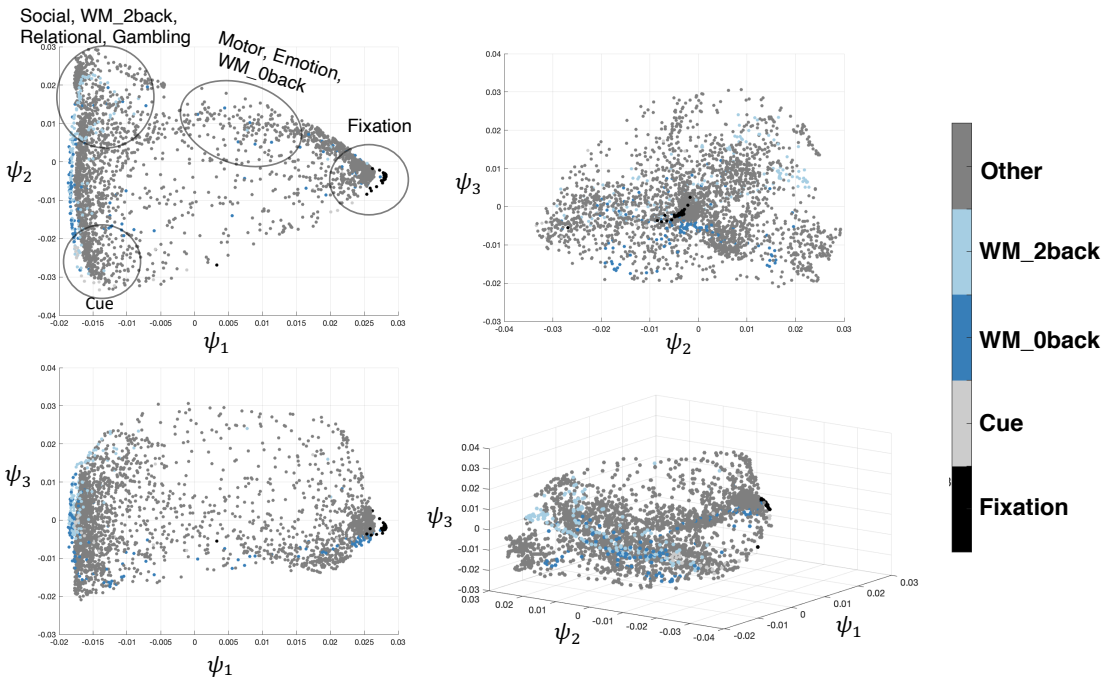


Figure 4.9: Comparison of 2-step PCA embedding with 2-step Diffusion Maps in the first 2 embedding coordinates without the points colored by task block types.

together in a meaningful way becomes more interesting as it could reveal the common factor that drives different tasks. Therefore, when we compared the embedding from 2sPCA vs 2sDM, we tried to see whether 2sPCA could also reveal a meaningful embedding structure. However, as shown in Figure 4.11, without the colors as prior, it was harder to infer the structure from the 2sPCA embedding. Although different tasks were separated (Figure 4.4), the cue and fixation time points didn't form two corners as in the 2sDM embedding, which could make the downstream analysis less interpretable (illustrated in the following trajectory analysis). Moreover, it is obvious to analyze the three corners of the 2sDM embedding at first as they are the anchors of the embeddings while harder to choose similar points in the 2sPCA embedding.

Temporal trajectory analysis As we illustrated in the main paper that the low-dimensional temporal trajectories formed with each task conditions revealed the dynamic cognitive processes, the 2sPCA trajectories was less informative. In the figure below (Figure 4.12b), we showed the average trajectory for each task. As the 2sDM embedding clustered the cues and fixations in

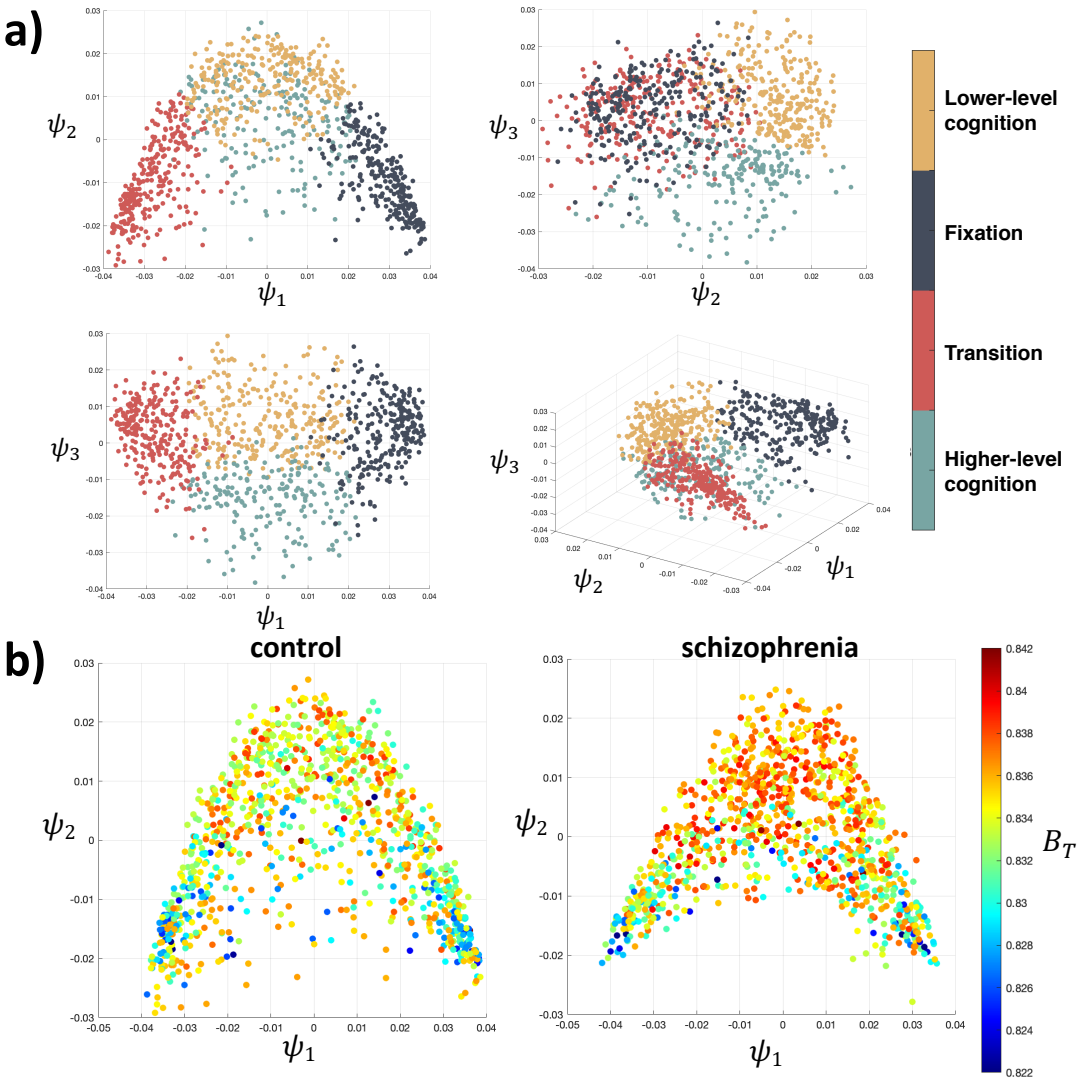


Figure 4.10: **a)** 2sDM embedding and K-Means clustering result of CNP dataset. **b)** Embedding with the first 2 dimensions of 2sDM in CNP dataset, colored by the corresponding B_T with the same colormap.

separate corners, the 2sDM trajectories were easier to interpret and compare across tasks. In comparison, as the fixation and cues were more scattered in the 2-step PCA embedding, the cross-task comparison was more difficult. For example, we can infer that for the Motor task, the 2sDM trajectory revealed the cognitive process that in the beginning, the subjects started from the cue state which was the common starting state across the other tasks. Then the individuals briefly entered the high-cog state, but not deep in the state and finally entered and stayed in the low-cog

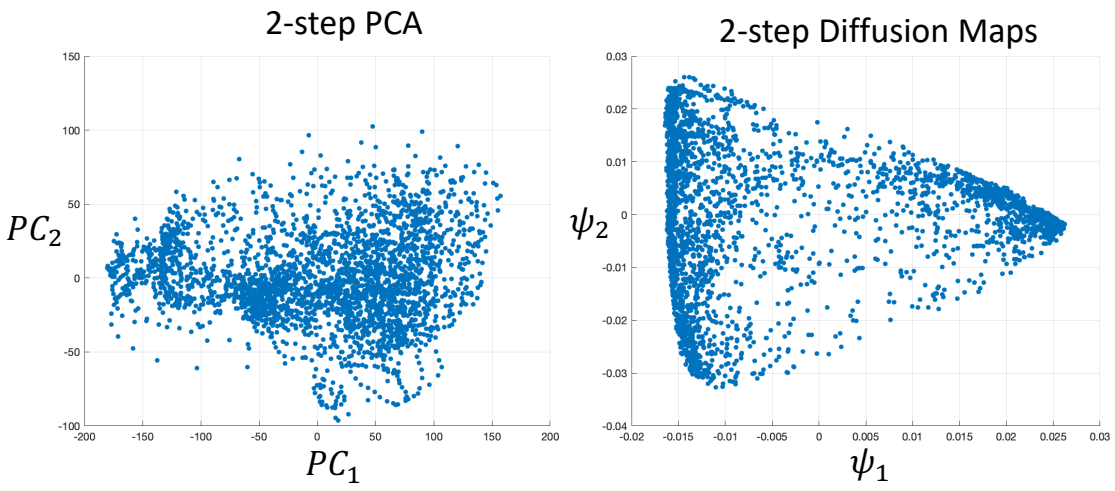


Figure 4.11: Comparison of 2-step PCA embedding with 2-step Diffusion Maps in the first 2 embedding coordinates without the points colored by task block types.

state. Actually, it also revealed that on average, individuals wandered towards the fixation state in the middle of the task block, suggesting a fatigue or practice effect. And towards the end of the task block, individuals returned deep into the low-cog state and moved towards the cue state for the next task block to start. However, none of these analyses would be obvious for the 2-step PCA trajectory as the motor task trajectory was in different location and progression pattern as other tasks.

Brain state clustering analysis As mentioned in the paper, the number of the clusters was chosen based on the Calinski-Harabasz criterion and 4 cluster was chosen as Calinski-Harabasz value was maximized (Figure 4.13b). In comparison, if we run k-mean clustering on the PCA embedding, we could also get clusters like shown below (Figure 4.13a). However, as shown below by the Calinski-Harabasz value (Figure 4.13b), there is no local maximum value as k increases, thus suggesting a lack of clear clustering structure in the low-dimensional embedding of 2sPCA. This validated that compared with our nonlinear embedding, PCA-based linear methods generates less-structured embeddings for the multi-task fMRI data.

2sPCA embedding also reveals global integration and segregation In the paper, we have shown that 2sDM embedding's first 2 coordinates are highly correlated with the participation coefficient

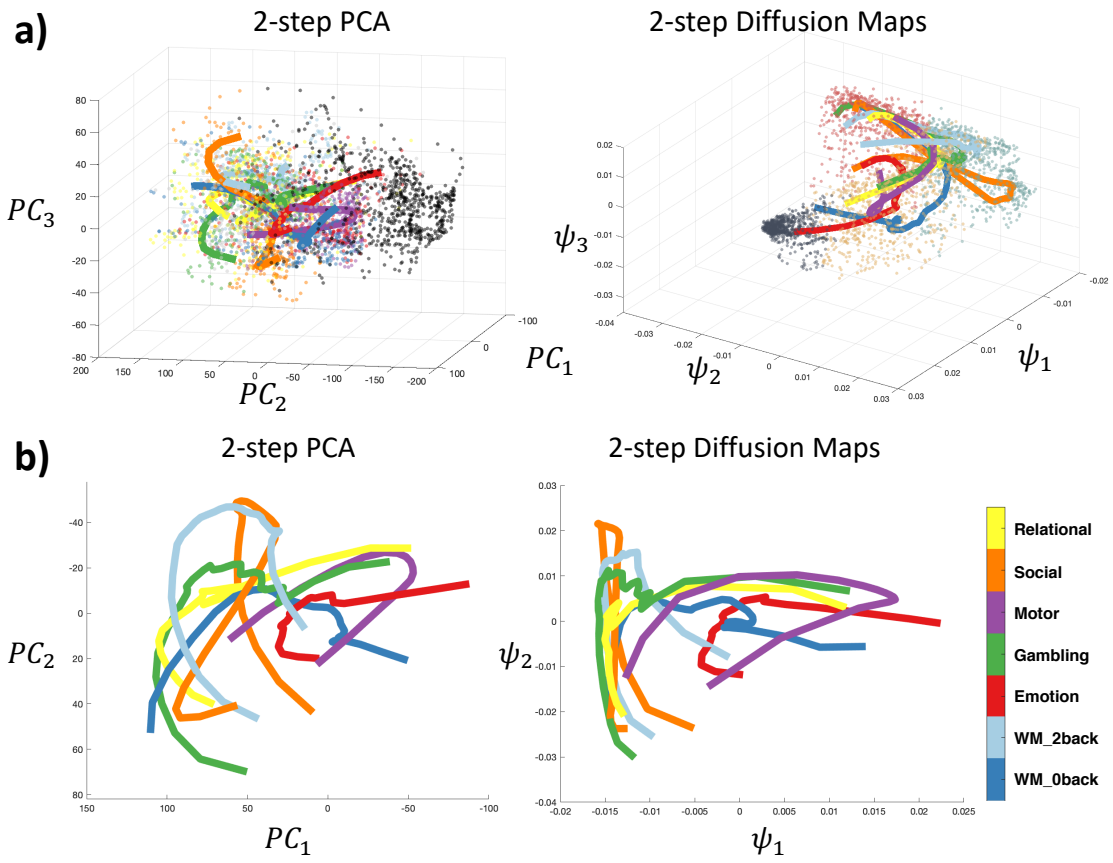


Figure 4.12: Trajectory comparison of 2sPCA and 2sDM.

(B_T), suggesting that the task embedding was able to capture handcrafted features like B_T in an unsupervised manner (Figure 4.5). It is also interesting that the 2sPCA embedding also revealed a similar relationship (Figure 4.14), suggesting that the embedding was also related with the global segregation and integration pattern. Thus, both 2sDM and 2sPCA's embedding was related with B_T.

4.4 Discussion

Using a novel manifold learning framework, we demonstrate that fMRI data from different tasks span the same low-dimensional embedding (i.e. brain states). In other words, moment-to-moment dynamics from any of these tasks group into the same small number of representative patterns that

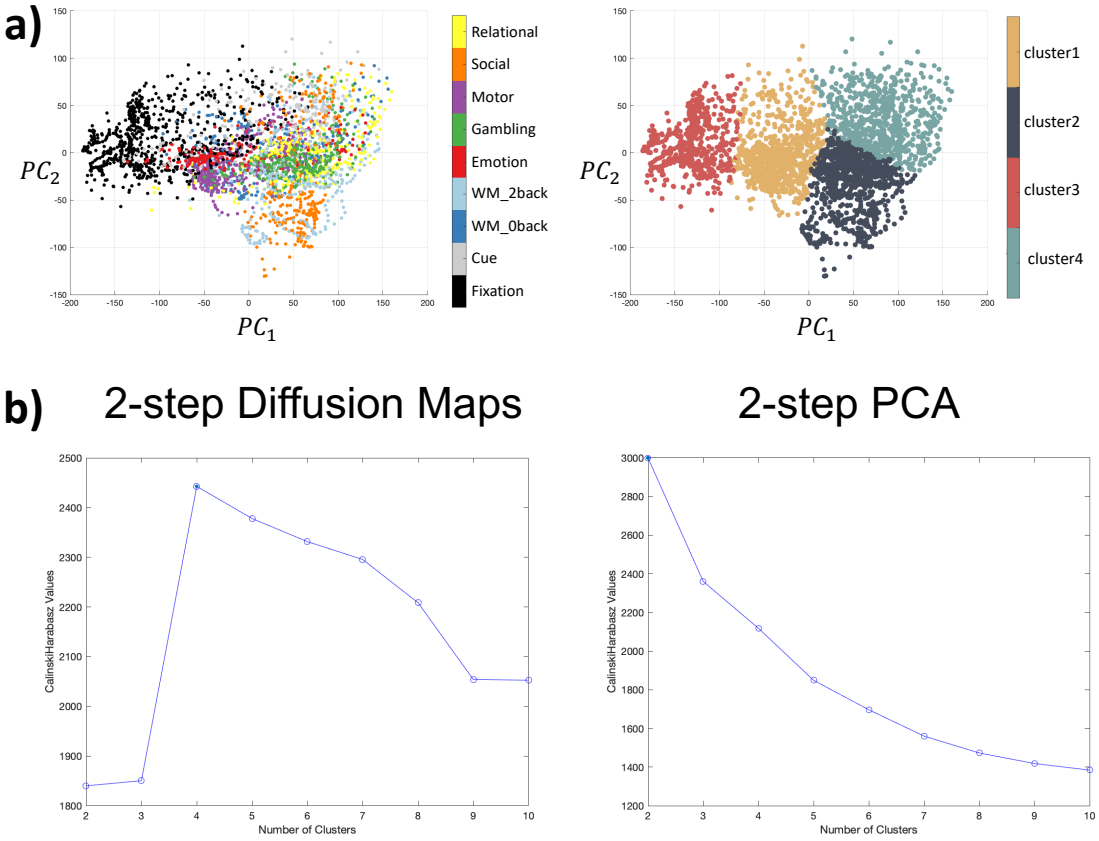


Figure 4.13: **Brain state clustering comparison.a)** K-means clustering result based on the 2-step PCA embedding. **b)** Clustering evaluation based on the Calinski-Harabasz criterion.

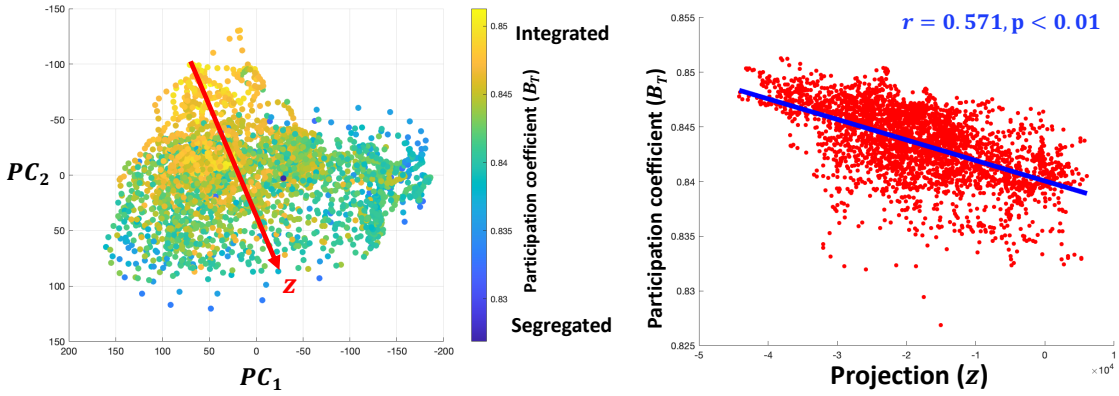


Figure 4.14: **2sPCA embedding is related with global integration and segregation. a)** 2sDM embedding in HCP dataset colored by the time-resolved B_T. **b)** Scatter plot of the B_T with the projection onto the diagonal of the embedding structure (z). Correlation of z with B_T is shown with a line of best fit.

are hidden from direct observation. To recover this embedding, we employed nonlinear methods (e.g. 2-step Diffusion Maps—2sDM) to project the fMRI data onto a larger portion of the manifold than would be possible using linear methods only. The embedding maintained proper temporal progression of the tasks, revealing brain states and temporal dynamics of changes in network integration. Further, we demonstrate that resting-state data project onto the same task embedding using a specially designed out-of-sample-extension method, indicating similar brain states are present. Finally, we validate this embedding using an independent dataset.

Several other publications have organized the temporal dynamics of the brain into a low dimension space or into distinct brain states (Allen et al. [2014b], Vidaurre et al. [2017], Saggar et al. [2018]) using data from resting-state or a single task to construct the embedding (Gallego et al. [2017], Shine et al. [2019]). Together, this work suggests that a low-dimensional structure exists; however, it is unclear how these structures adapt to diverse cognitive loads. By projecting a rich repertoire of task data into a single manifold, we show that, across different tasks, parts of the embedding (i.e. brain states) are well characterized by network segregation (i.e. communication mainly within brain networks) and integration (i.e. communication mainly across diverse brain networks) (Deco et al. [2015]). Overall, the discrete states and association with network segregation/integration suggest that our embedding finds an intrinsic, latent structure of brain dynamics.

These results are in line with the theory that the brain is able to reconfigure its large-scale organization dynamically either between different cognitive tasks or within resting-state (Cohen and D'Esposito [2016], Shine et al. [2016]). Further, they emphasize that this reconfiguration is shared across different cognitive loads and, importantly, resting-state. In other words, the same highly integrated state that characterizes a cognitively demanding task, such as a 2-back WM task, can be observed during resting-states and less cognitively demanding tasks, just with less frequency. These states can also be viewed from a dynamic system perspective (Taghia et al. [2018]). As clustering based on the eigenvectors of the normalized graph Laplacian has been used to find meta-stable state in the stochastic dynamical systems (Huisinga et al. [1999]), the four brain

states defined from the task scan can also be viewed as four different metastable states. Further, the temporal trajectories can separate different portions of tasks based on cognitive demand, suggesting a potential utility of the embedding for other downstream analyses of brain dynamics.

In line with this, the dynamics between states, rather than within brain states themselves, appear to be the key distinguishing factor between task and rest. In support of this, how the brain transitions between different states is dependent on the task being performed and is less predictable in resting-state compared to tasks. Executing a task limits the transitions between states; while, during resting-state, the brain can more liberally traverse through different states. Though speculative, these results offer an explanation as to why task connectivity data is better at identifying individuals and subsequent predicting behaviors than resting-state connectivity data (Finn et al. [2017a], Greene et al. [2018a]). Together, while the resting state may exhibit similar states as observed during task, the temporal dynamics of switching states are less predictable in resting state compared to task.

Previous work demonstrates that brain networks fluctuate between states of low and high global integration during tasks as characterized by the participation coefficient (B_T) from sliding-window functional connectivity. Tasks requiring higher cognitive loads, such as the 2-back condition in the WM task, exhibit greater integration while less cognitive load, such as the motor task, exhibits lower integration (Shine et al. [2016]). A key drawback of these results is that they rely on two intermediate steps (e.g. the method used to construct dynamic functional connectivity and topological metrics to study), rather than the learned features from unsupervised methods. Together, our results suggest that the task embedding reveals latent information about changes in network topology without the need for handcrafted features. For example, each task can be effectively characterized from the proportion of time spent in lower-level and higher-level cognition states creating a similar ordering of task (see Figure 4.7b) as in (Shine et al. [2016]).

While resting-state fMRI is a powerful tool to map the functional organization of the brain, inherent limitations exist. Resting-state is often conceptualized as a single task state. Though

emerging data, including our results, suggest that resting-state is not one single, monolithic state, but rather a collection of multiple states associated with different cognitive loads that also appear during tasks (Vidaurre et al. [2017]). For example, while the majority of resting-state time points cluster into a single part of the manifold (such as the fixation blocks, which putatively are the most like "rest"), nearly a third of the time points more closely match cognitive states. Perhaps, more importantly different groups may have differences in "performing" rest (Buckner et al. [2013]). How best to interpret changes in resting-state connectivity in the presence of group differences in dynamics is still an open question.

A key strength of our embedding framework is its data-driven nature. Although the only inputs are time-courses from task fMRI data, we demonstrated that the embedding coordinates can reveal topological information originally found using dynamic functional connectivity methods (Shine et al. [2016]). This brain topology was found without specifying common modeling choices in dynamic functional connectivity or fMRI, in general, such as how to model the functional connectivity (i.e. statistical interdependence of signals) between brain regions, an underlying graph/network, or even information about task stimuli (e.g. block lengths). As a multitude of methodological choices have been proposed to analyses (Calhoun et al. [2014], Hutchison et al. [2013]) (e.g. ways of estimating connectivity (Allen et al. [2014b], Chang and Glover [2010], Shine et al. [2015]), constructing a weighted or unweighted graph (Rubinov and Sporns [2010]), specific graph theory measures (Honey et al. [2007], Meunier et al. [2010], Shine et al. [2016], Sizemore and Bassett [2018]), our embedding framework provides an end-to-end, data-driven approach without the need for modeling choices to investigate brain dynamics. More generally, handcrafted features are being substituted by more automatic feature learning-based nonlinear methods such as deep learning and nonlinear embedding methods (Hamilton et al. [2017]). Our results show a specific scenario in which "let the data speak for itself" is an achievable option for modeling fMRI data.

A limitation of this work is that the embedding can only "look under the light." That is to say that, while a rich amount of task data was needed to create the embedding, we could not include

every possible task in creating the embedding. Indeed, it is highly likely that many more than four brain states exist and that we cannot detect every single one. A finer grade delineation of states, probably through further advancement in non-linear embedding methods, is a needed future direction of work. Moreover, although here brain states are defined based on the k-means clustering result, it doesn't rule out other ways to define brain states. For example, at each time point, the brain can also be modeled as being at different states with distinct possibilities (Vidaurre et al. [2017]), which can be achieved by a fuzzy-clustering algorithm. Moreover, the brain state can also be characterized by the temporal trajectory where trajectory clustering technique can be used to cluster trajectories into trajectory-based brain states, which takes account the temporal information of the embedding (Lee et al. [2007]). The k-means clustering way of defining brain state is only one of the ways to summarize information of the embedding and serves as a proof-of-concept that our embedding contains information that is relevant to brain dynamics. Nevertheless, the observed task embedding was similar across two different input datasets with different tasks, suggesting that embedding is general to factors such as scanner, task, processing, and sample size.

One of the assumptions of 2sDM is that the time frames from all individuals are temporally aligned so that a group-average embedding of the time frames can be obtained. However, this doesn't rule out the applicability of the task scans that has different task block lengths/orders across individuals (e.g., language task in the HCP dataset) or the resting-state scans, which we have demonstrated in the paper by applying BrainSync. So, task scans with distinct block lengths/orders can also be embedded with 2sDM by applying BrainSync first. It is worth noting that as BrainSync requires a specific individual chosen as the reference, by aligning all the other individuals to the same selected individual, the group-average embedding then will approximate a cleaner temporal embedding of the selected individual, which can be used to investigate individual-level dynamics.

The ability to use data-driven methods to clearly identify a low-dimensional space of brain dynamics, regardless of how the brain is engaged during imaging, indicates that these brain dynamics are robust and reliable across conditions in addition to being unique. Together, these advances

suggest that analysis of individual fMRI data from multiple cognitive tasks in a low-dimensional space is possible, and indeed, desirable.

Chapter 5

Poincaré embedding reveals edge-based functional networks of the brain

5.1 Introduction

Elucidating the functional organization of the human brain by grouping distinct brain regions into functional networks is a major goal of current fMRI research (Eickhoff et al. [2018]). Using measures of functional connectivity or spatiotemporal patterns of brain activity (Friston [2011]), many approaches have been used to form $\sim$10 functional networks (*e.g.*, motor and default mode networks) (Power et al. [2011a], Damoiseaux et al. [2006], Salehi et al. [2018], Thomas Yeo et al. [2011]).

An important problem with these approaches is the so-called "resolution limit" (Fortunato and Barthélemy [2007]). As a system becomes larger, the expected number of connections between regions decreases, eventually leading to situations where merging two distinct networks is better than keeping them separated. Additionally, a region's membership to a particular network is likely fuzzy, such that two networks can overlap in a particular region (Wu et al. [2011], Salehi et al. [2020]).

An alternative method for creating networks within a large system is the "link community" paradigm, where the networks are redefined as sets of links (*i.e.*, edges or connections) rather than

regions (Ahn et al. [2010], Evans and Lambiotte [2010]). This framework provides a natural way to allow regions to belong to multiple networks as edges originating from a particular node can belong to many networks (Ahn et al. [2010], Evans and Lambiotte [2010, 2009]).

However, the large number of edges in fMRI data (*e.g.*, $O(N^2)$) poses challenges in terms of computation and representation when finding link communities (Ahn et al. [2010], Evans and Lambiotte [2010, 2009]). Here, we show that, while traditional approaches for finding link communities fail to discover a valuable network representation of the human brain, a novel embedding approach—based on the Poincaré embedding (Nickel and Kiela [2017])–offers a naturalistic approach to find link communities in high-dimensional fMRI space. An overview of our approach and how it differs from region clustering and previous link community detection approaches is shown in Figure 5.1.

5.2 Methods

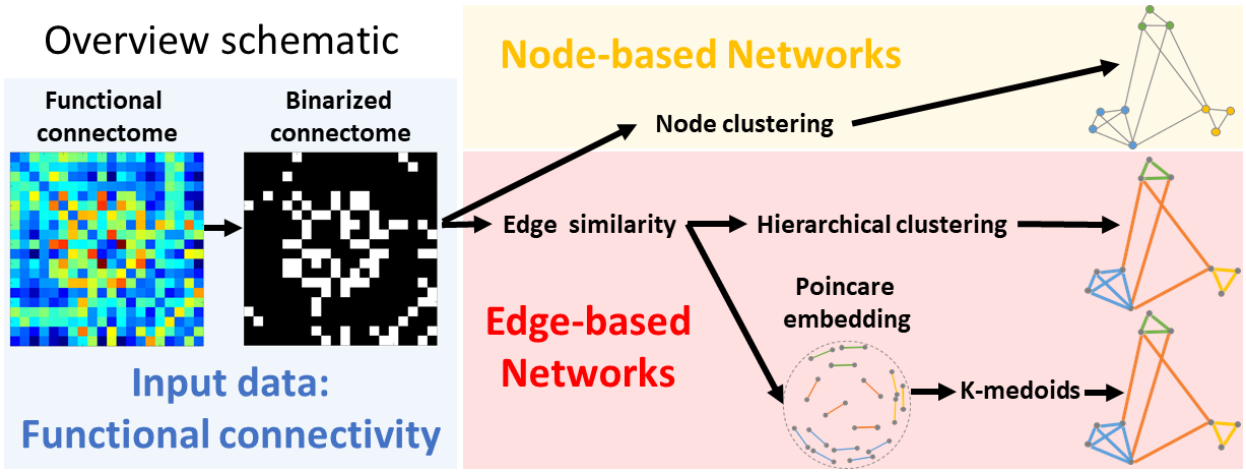


Figure 5.1: **Overview of using the Poincaré embedding to form edge-based networks.** Starting with functional connectivity matrices, or connectomes (*blue box*), functional brain networks can be formed based on clustering nodes or edges. Previous approaches using fMRI data have focused only on creating node-based networks (*yellow box*). Yet, approaches to create edge-based networks exist. Here, we show that the link community detection methods proposed by Ahn *et al.* (Ahn et al. [2010]) do not reveal interpretable functional networks (*top row in red box*) and that our novel Poincaré embedding approach does (*bottom row in red box*).

5.2.1 Link community detection

Let $G = (V, E)$ be a graph with vertex set V and edge set E. Traditional community detection methods define a community as a set of nodes that have more internal than external connections. Link community detection methods work in a similar way, but define a set of interrelated edges as the community. Each node then inherits all memberships of its edges and, thus, can belong to multiple, overlapping communities. As edges—rather than nodes—are assigned to different communities, similarities between edges need to be determined. One of the first methods, proposed in Ahn *et al.* (Ahn et al. [2010]), calculates the similarity between edges e_{ik} and e_{jk} that share a common node k. The similarity is calculated as the Jaccard index between the sets of node neighbors from the outer node i and j: $S(e_{ik}, e_{jk}) = |n(i) \cap n(j)|/|n(i) \cup n(j)|$, where $n(i)$ is the set of neighboring nodes of node i. With the similarity defined, standard hierarchical clustering is applied to group edges into link communities as it also allows to reveal hierarchy. This relies on applying a threshold to the hierarchical clustering dendrogram to create distinct communities. Partition density is a measure to determine the quality of the partitions and to find the optimal threshold to cut the dendrogram. For a network with M links, the partition density is defined as $D = \frac{2}{M} \sum_c m_c \frac{m_c - n_c - 1}{(n_c - 2)(n_c - 1)}$, where the candidate partition creates c subsets with each subset having m_c links and n_c nodes. Yet, even when determined through a systematic way (*e.g.*, partition density), this thresholding may lose information between different tree levels, leading to an uninterpretable number of communities.

5.2.2 Hyperbolic space for embedding tree structures

Embedding a tree structure, such as a dendrogram, in Euclidean space is difficult as the number of child nodes grows exponentially with their distance from the root of the tree. Thus, the dimensionality of the Euclidean embedding rapidly grows to handle these increasingly complex hierarchies. Increasing this dimensionality leads to increased computational complexity and overfitting. However, a hyperbolic space is more suitable for embedding tree structures as the area of a hyperbolic

disc grows exponentially with its radius. Specifically, a hyperbolic space is a non-Euclidean space with constant negative curvature. For a two-dimensional hyperbolic space $\mathbb{H}^2_\zeta$ with constant curvature $K = -\zeta^2 < 0$, the length of a circle and the area of a disk with hyperbolic radius r, are $L(r) = 2\pi \sinh \zeta r, A(r) = 2\pi(\cosh \zeta r - 1)$, both growing exponentially as $e^{\zeta r}$ with r (Figure. 5.2). With this property, hyperbolic spaces can be constructed as continuous versions of trees. This is not possible in $\mathbb{R}^2$. Furthermore, given the existence of hierarchical structure in complex systems Clauset et al. [2008], they have been modeled by hyperbolic spaces Krioukov et al. [2010].

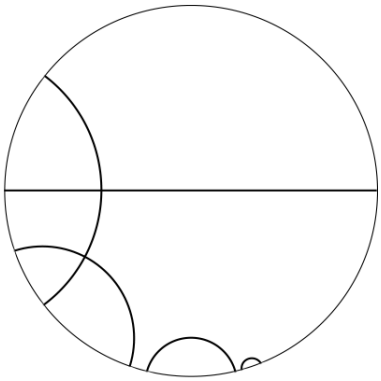


Figure 5.2: **Geodesics of Poincaré disk model.**

5.2.3 Poincaré ball model and embedding

To take advantage of a hyperbolic space for embedding edge similarities, we use the Poincaré embedding, an approach based on the Poincaré ball model (Nickel and Kiela [2017]). Let $\mathcal{B}^d = \{\boldsymbol{x} \in \mathbb{R}^d \mid ||\boldsymbol{x}|| < 1\}$ be the open d-dimensional unit ball where $||\boldsymbol{x}||$ is the Euclidean norm. Then, the Poincaré ball model corresponds to a Riemannian manifold $(\mathcal{B}^d, g_{\boldsymbol{x}})$. The Riemannian metric tensor is $g_{\boldsymbol{x}} = (\frac{2}{1-||\boldsymbol{x}||^2})^2 g^E$, where $\boldsymbol{x} \in \mathcal{B}^d$ and g^E represents the Euclidean metric tensor. From the Riemannian metric tensor, the Poincaré distance between points $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{B}^d$ is given as $d(\boldsymbol{u}, \boldsymbol{v}) = \operatorname{arcosh}(1+2\frac{||\boldsymbol{u}-\boldsymbol{v}||^2}{(1-||\boldsymbol{u}||^2)(1-||\boldsymbol{v}||^2)})$. Geodesics in $\mathcal{B}^d$ are circles perpendicular to the boundary $\delta\mathcal{B}$. Moreover, the model excludes the boundary $\delta\mathcal{B}$.

The goal of the Poincaré embedding is to find a representation $\Theta = \{\boldsymbol{\theta}_i\}_{i=1}^n$, where $\boldsymbol{\theta}_i \in \mathcal{B}^d$, that minimizes the loss function $\mathcal{L}(\Theta)$. Specifically, a soft ranking loss function is used:

$$\mathcal{L}(\Theta) = \sum_{(\boldsymbol{u},\boldsymbol{v})\in\mathcal{D}} \log \frac{e^{-d(\boldsymbol{\theta}_{\boldsymbol{u}},\boldsymbol{\theta}_{\boldsymbol{v}})}}{\sum_{\boldsymbol{v}'\in\mathcal{N}(\boldsymbol{u})} e^{-d(\boldsymbol{\theta}_{\boldsymbol{u}},\boldsymbol{\theta}_{\boldsymbol{v}'})}},$$

where the set $\mathcal{D} = \{(\boldsymbol{u},\boldsymbol{v})\}$ is the set containing input pairs that are similar, $\mathcal{N}(\boldsymbol{u}) = \{\boldsymbol{v}|(\boldsymbol{u},\boldsymbol{v}) \notin \mathcal{D})\} \cup \{\boldsymbol{u}\}$ is the set of negative examples for $\boldsymbol{u}$. Ten negative examples are chosen during training. This loss function encourages similar points (*i.e.*, edges) to be close in the hyperbolic space with regard to their Poincaré distance.

Since the Poincaré ball model has a Riemannian manifold structure, manifold optimization methods such as Riemannian stochastic gradient descent (RSGD) (Bonnabel [2013]) can be used to minimize the loss function, which requires to calculate the Riemannian gradient and apply the retraction operator to map the gradient from the tangent space onto the manifold. The Riemannian gradient can be obtained by scaling the Euclidean gradient ∇_E by the inverse of the Poincaré ball metric tensor $g_\theta^{-1} = \frac{(1-||\boldsymbol{\theta}_t||^2)^2}{4}$. The retraction operation we use is $\Re_{\boldsymbol{\theta}}(\boldsymbol{v}) = \boldsymbol{\theta} + \boldsymbol{v}$. The embedding is further restricted within the Poincaré ball via the projection.

$$\text{proj}(\boldsymbol{\theta}) = \begin{cases} \boldsymbol{\theta}/||\boldsymbol{\theta}|| - \epsilon, & \text{if } ||\boldsymbol{\theta}|| \geq 1 \\ \boldsymbol{\theta}, & \text{otherwise} \end{cases}.$$

One full update of a single embedding is thus given by

$$\boldsymbol{\theta}_{t+1} \leftarrow \text{proj}(\boldsymbol{\theta}_t - \eta_t \frac{(1-||\boldsymbol{\theta}_t||^2)^2}{4} \nabla_E),$$

where η_t is the learning rate. It is also worth mentioning that this combination of Riemannian gradient with the simple retraction operation corresponds to the natural gradient method.

5.2.4 Poincaré embedding of brain edge network

Functional connectivity represents the temporal correlation of time series between brain regions. Let $G = (V, E)$ be a functional connectivity with brain regions V and edge set E, a subset of edges $e_{ij} \in E$ is first selected as the objects to be embedded. Specifically, top x percent of edges with the strongest edge weight (correlation value) are selected, resulting in a binary connectivity matrix. Next, the similarities between edges that share a common node were calculated based on the similarity measure described above $S(e_{ik}, e_{jk}) = |n(i) \cap n(j)|/|n(i) \cup n(j)|$. Edges with no common nodes will have similarity of 0. The Poincaré embedding of those edges is then calculated based on this measure of similarities. As a result, in the embedding space, an edge will be closer to edges for which they have higher similarity values than other unrelated edges.

5.3 Results

5.3.1 Datasets and processing

We applied our algorithm to the resting-state data from the Human Connectome Project dataset (Van Essen et al. [2013a]). After excluding subjects for mean frame-to-frame displacement $> 0.15\,mm$, 514 (240 males) healthy subjects were used for analysis. This conservative threshold for exclusion due to motion was used to mitigate the substantial effects of motion on functional connectivity. fMRI data were processed with standard methods and parcellated into 268 nodes using a whole-brain functional atlas defined previously in a separate sample (Shen et al. [2013b]). Next, the mean timecourses of each node pair were correlated and Fisher transformed, generating a 268×268 functional connectivity matrix (also called a commectome) per individual. Connectomes were averaged over all individuals and binarized by taking the top 5% (~ 1800) edges based on previous work (Power et al. [2011a]). For the proposed method based on the Poincaré embedding, half of the subjects (257 subjects) were used to generate the embedding; while, the other half were used

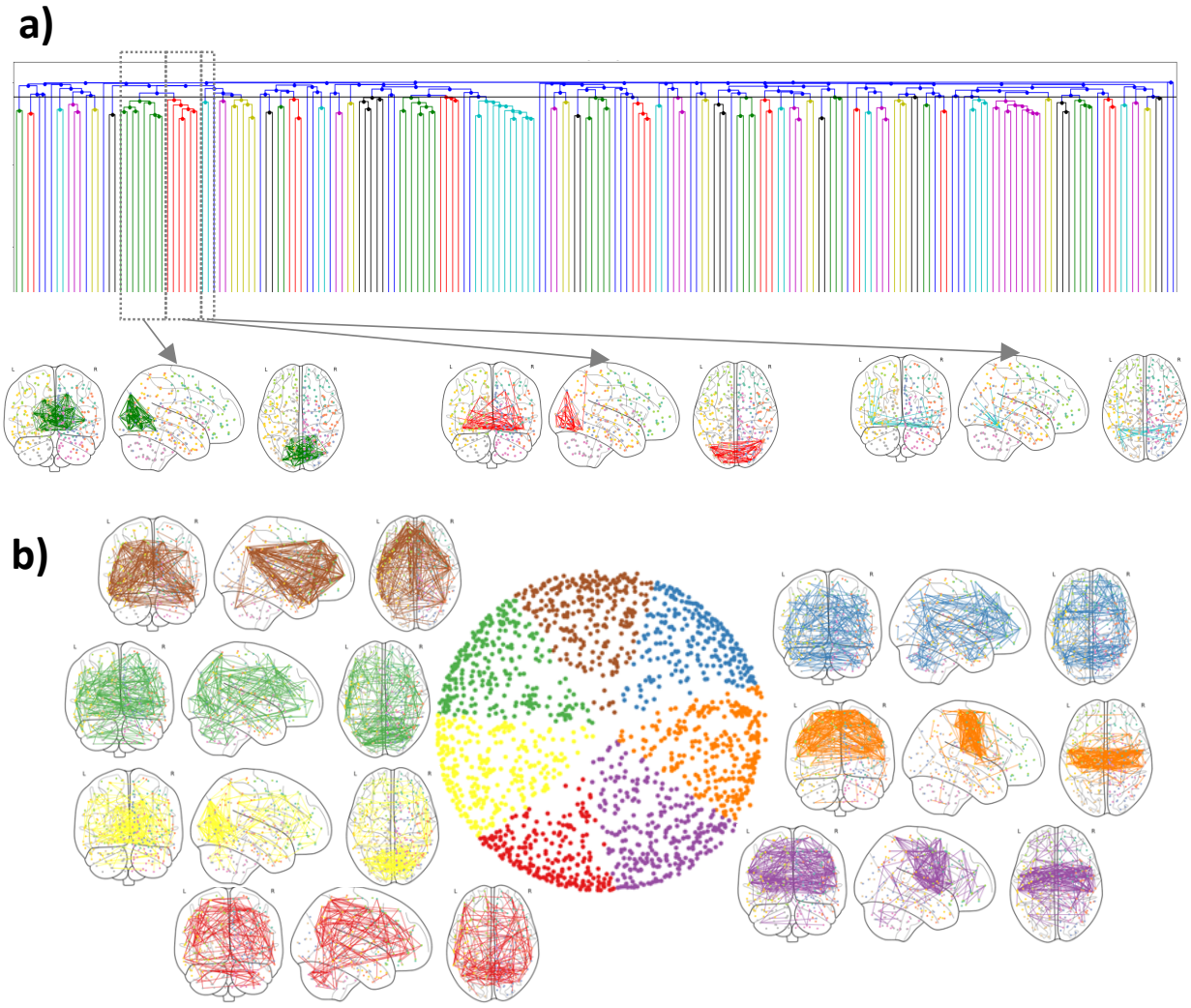


Figure 5.3: **a) Dendrogram from linkage clustering method.** Colors of adjacent links represent the same cluster under the optimal cut threshold. In total, 84 clusters were found. Edges from three disjoint clusters that together compromise the visual network are shown. **b) Multidimensional scaling embedding and K-means clustering result.** Edges are densely located within the embedding space (*center*); but, most networks—shown on the periphery—remain mostly uninterpretable

for replication of the embedding.

5.3.2 Traditional link community detection fails

The traditional link community detection framework, proposed by Ahn *et al.* (Ahn et al. [2010]) and described in 5.2.1, did not lead to meaningful results (Figure 5.3a). Specifically, 84 networks

were found by cutting the dendrogram based on maximizing the partition density. This number of clusters was significantly greater than the $\sim$10 networks from using node-based methods (Power et al. [2011a], Damoiseaux et al. [2006], Salehi et al. [2018], Thomas Yeo et al. [2011]). Additionally, these clusters identified disjoint functional networks that putatively should belong to the same network. For example, the green and red clusters in Figure 5.3a are both part of the visual network and should be combined into a single network. An additional post-hoc analysis is needed to merge these and other clusters (*i.e.*, the cyan cluster) to from a proper functional network.

Finally, as hierarchical clustering produced too many clusters, we attempted to generate functional networks using K-means, which groups data into a specified number of clusters. In order to perform K-means, first, we generated a 2-dimensional Multidimensional scaling (MDS) embedding based on the same edge similarity as above. Edges are densely located within the embedding space, but most networks remain uninterpretable (Figure 5.3b). For example, while the orange network represents the motor network, the blue and green clusters consists of edges from all over the brain and are difficult to interpret.

5.3.3 Poincaré embedding of edges

The Poincaré Embedding of the resting-state data is shown in Figure 5.4a. Edges that appeared closer in the embedding had denser inter-connections (*i.e.*, more common node neighbors). To form functional networks, K-medoids clustering was applied to the Poincaré distance between edges. Visualizing these edge-based networks on the brain (Figure 5.4b) verified that edges within the same cluster form dense communities in different locations. The replication (see supplementary material) showed similar embedding and clustering, demonstrating the robustness of the framework.

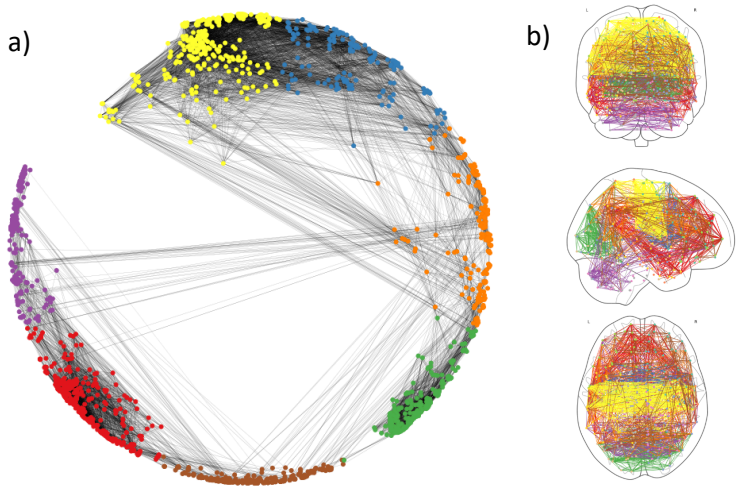


Figure 5.4: **a) Poincaré embedding of resting-state functional connectivity.** Each node in the embedding represents an edge in the function connectome, colored by the K-medoids clustering result. Width of the black lines connecting nodes represent similarities between two edges. **b) Embedded edges visualized on the brain.** Edges in the same cluster form densely inter-connected networks.

5.3.4 Functional edges show a canonical network structure

Compared with the edge-based networks in Section 5.3.2, the edge-based networks derived from clustering the Poincaré embedding showed a more interpretable structure. For example, in addition to the clear motor network observed in Section 5.3.2, the visual, default mode, auditory, language, and medial frontal networks were easily identified (Figure 5.5a). Next, we compared the nodes covered by edges in each network to previously defined canonical node-based networks (Finn et al. [2015b]) (Figure 5.5b). Visually the edge-based and the node-based networks were comprised of similar nodes. Finally, each edge's Poincaré distance from the origin $(0, 0)$ represents its hierarchy in the network structure. The closer an edge is with the origin; the higher in hierarchy that edge is (see arrow in Figure 5.5a). Although networks formed with higher hierarchy edges were sparser, the overall topological structure was preserved, suggesting the supporting role those edges had in the overall network topology (see supplementary material).

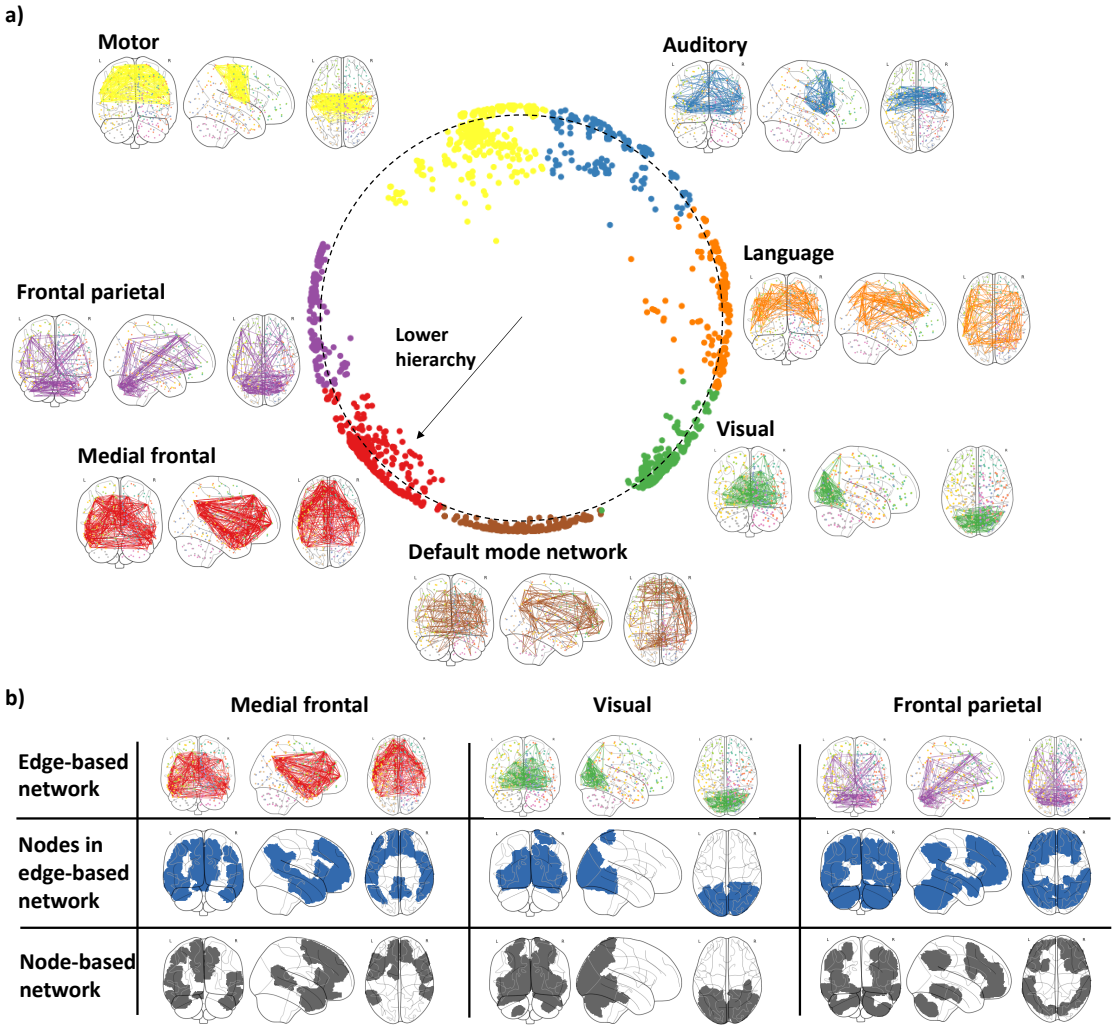


Figure 5.5: **a) Edge-based networks using Poincaré embedding.** Network labels were determined by matching the nodes in the edge-based network with predefined canonical networks. Dashed circle in the embedding represents the average distance to the center. The arrow points towards lower in the hierarchy of edges within a network. **b) Comparison with node-based networks** Nodes belonging to the edge-based networks were similar to the nodes belonging to canonical networks using node community detection approaches.

5.3.5 Overlapping edge-based networks for a node are meaningful

A major advantage of our edge-based networks is that each region of the brain can be associated with multiple canonical networks. For each region, we summed the number of edge-based networks associated with that node. As shown in Figure 5.6, regions associated with cognitive processing (posterior cingulate cortex, prefrontal cortex, and parietal lobe) showed the highest

number of overlapping networks. In contrast, regions in the motor cortex or visual lobe showed membership to the least number of networks. Overall, these results align with previous research, suggesting functional specialization in those areas (Cole et al. [2013], Salehi et al. [2020]).

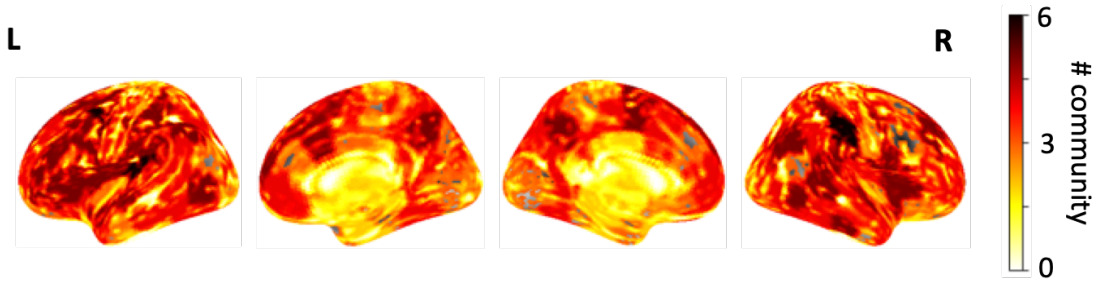


Figure 5.6: **Overlapping networks.** Nodes were color coded by the number of edge-based networks that edges from the node were part of. Regions with the highest number of overlapping networks were in prefrontal and association cortices, while regions with the lowest number of overlapping networks were in the visual and motor cortices.

5.4 Summary

In this paper, we propose a framework to find link communities from functional connectivity data. The framework consists of: first, embedding edges into a Poincaré disk model and, then, using K-medoids clustering to group the embedded edges into functional networks. These edge-based networks matched canonical brain networks, defined using conventional node-based approaches. Yet, edge-based networks allow nodes the flexibility to belong to multiple networks, a major advantage of this framework over standard community detection approaches. Although the Poincaré embedding has only 2 dimensions, it provided a parsimonious representation that was able to partition the brain into functionally meaningful networks. Future work includes using higher dimensional embedding to test if more information can be preserved to improve network detection. Overall, our framework provides a novel tool for characterizing the functional network organization of the brain.

Chapter 6

Conclusion

With the increasing size of the recent fMRI dataset, it has become more and more important to have appropriate methods for the complex data. The ideal method would be able to aggregate information from multiple available sources and extract those that are relevant to the behavior of interest, with less human in the loop (e.g., parameter tuning). In this work, the methods we propose fill some of the gaps between fMRI data and computational methods and demonstrates the advantage of using those advanced methods.

In chapter 2, we propose two extensions of the CPM framework that are multidimensional in terms of the connectomes. By incorporating more available connectomes in CPM, we are able to achieve higher prediction accuracy than using any of the single connectome alone, suggesting the advantage of having a holistic model that can utilize more information in prediction. Moreover, the proposed methods also seem to be less sensitive to hyperparameter tuning, which increases both the applicability and generalizability of the CPM framework.

In chapter 3, we propose another two multidimensional CPM framework, which are multi-dimensional in the behavior dimension. The proposed frameworks can be used to derive new composite behavioral measures (e.g., general intelligence) from multiple behavior measures (e.g., visuospatial processing score, working memory test score), and this new composite measure is shown to be more accurately predicted than any of the single behavior measure, suggesting the

increased robustness and brain-behavior relationship. Altogether with the chapter 2, we have come up with a holistic CPM framework that is able to take in multiple available connectomes and behavioral measures, and output a brain-behavioral prediction model that utilizes all the available information while providing optimized predictions.

In chapter 4, we take a step back from modeling the brain as a graph model to the time series. A nonlinear dimensionality reduction method is proposed to embed each time frame of the fMRI data. By this low-dimensional embedding, we are able to view the brain's activity change more dynamically and intuitively compared with the static FC. Moreover, we can summarize the whole fMRI dynamics across multiple tasks with four common brain states, which can be used to investigate cognitive differences across different tasks. Last but not least, we reveal brain graph topological information from the embedding even without explicitly modeling the brain as a graph, which again suggests the holistic model's advantage in revealing new insights with less human-in–the-loop.

In chapter 5, we look at the low-dimensional representation of the brain graph and propose a hyperbolic space based brain graph edge embedding framework. The framework is able to represent $o(N^2)$ edges in the more powerful hyperbolic space and cluster edges into disjoint edge networks. The formed edge networks not only provide a more accurate description of brain graphs (e.g., edges between node networks are explicitly associated with certain networks), but also provide a natural definition of overlapping brain node networks, which also reveal the functional flexibility of each brain node.

In all, these works provide solutions to some of the existing challenges due to the increasing size of fMRI data and also shine light on new discoveries that can be made with less human efforts. These results also suggest that there is room of improvement in terms of aggregating multi-source information in fMRI research and the benefits of having a holistic model in incorporating these information. The proposed methods only cover limited area of applications in fMRI data and most of the results serve as a proof-of-concept that these methods are applicable and useful. Future work will include more wide-range applications of the proposed methods (e.g., multi-modal scans

integration) and use the existing results for other downstream analyses (e.g, utilization of the edge networks).

Bibliography

Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann. Link communities reveal multiscale complexity in networks. *nature*, 466(7307):761, 2010. 7, 79, 80, 84

M. B. Ahrens, J. M. Li, M. B. Orger, D. N. Robson, A. F. Schier, F. Engert, and R. Portugues. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature*, 485(7399):471–7, 2012. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/nature11057. URL `https://www.ncbi.nlm.nih.gov/pubmed/22622571`. 46

D. L. J. Alexander, A. Tropsha, and D. A. Winkler. Beware of r2: Simple, unambiguous assessment of the prediction accuracy of qsar and qspr models. *Journal of Chemical Information and Modeling*, 55(7):1316–1322, 2015. doi: 10.1021/acs.jcim.5b00206. URL `https://doi.org/10.1021/acs.jcim.5b00206`. PMID: 26099013. 39

E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, 2014a. doi: 10.1093/cercor/bhs352. 47

E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, 24(3):663–676, 2014b. 2, 73, 75

N. A. Asendorf. Informative data fusion: Beyond canonical correlation analysis. 2015. 34

M. Belkin, D. Hsu, and J. Xu. Two models of double descent for weak features. *arXiv preprint arXiv:1903.07571*, 2019. 28

Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013. ISSN 0162-8828. 47

W. B. Bilker, J. A. Hansen, C. M. Brensinger, J. Richard, R. E. Gur, and R. C. Gur. Development of abbreviated nine-item forms of the raven's standard progressive matrices test. *Assessment*, 19 (3):354–369, 2012. 15

S. Bonnabel. Stochastic gradient descent on riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, 2013. 82

R. L. Buckner, F. M. Krienen, and B. T. Yeo. Opportunities and limitations of intrinsic functional connectivity mri. *Nature neuroscience*, 16(7):832–837, 2013. 75

V. D. Calhoun, R. Miller, G. Pearlson, and T. Adali. The chronnectome: time-varying connectivity networks as the next frontier in fmri data discovery. *Neuron*, 84(2):262–74, 2014. ISSN 1097-4199 (Electronic) 0896-6273 (Linking). doi: 10.1016/j.neuron.2014.10.015. URL https://www.ncbi.nlm.nih.gov/pubmed/25374354. 75

C. Cavada and P. S. Goldman-Rakic. Posterior parietal cortex in rhesus monkey: I. parcellation of areas based on distinctive limbic and sensory corticocortical connections. *Journal of Comparative Neurology*, 287(4):393–421, 1989. 3

C. Chang and G. H. Glover. Time–frequency dynamics of resting-state brain connectivity measured with fmri. *Neuroimage*, 50(1):81–98, 2010. ISSN 1053-8119. 75

M. M. Churchland, J. P. Cunningham, M. T. Kaufman, J. D. Foster, P. Nuyujukian, S. I. Ryu, and K. V. Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51–6, 2012. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/nature11129. URL https://www.ncbi.nlm.nih.gov/pubmed/22722855. 46

A. Clauset, C. Moore, and M. E. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008. 81

J. R. Cohen and M. D'Esposito. The segregation and integration of distinct brain networks and their relationship to cognition. *J Neurosci*, 36(48):12083–12094, 2016. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.2965-15.2016. URL https://www.ncbi.nlm.nih.gov/pubmed/27903719. 73

R. R. Coifman and S. Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21 (1):5–30, 2006a. 47

R. R. Coifman and S. Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21 (1):5–30, 2006b. 6, 48, 49

M. W. Cole, J. R. Reynolds, J. D. Power, G. Repovs, A. Anticevic, and T. S. Braver. Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, 16(9):1348–1355, 2013. ISSN 1546-1726. doi: 10.1038/nn.3470. URL https://doi.org/10.1038/nn.3470. 88

Z. Cui and G. Gong. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *Neuroimage*, 178: 622–637, 2018. 29

J. P. Cunningham and M. Y. Byron. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11):1500, 2014. ISSN 1546-1726. 3, 6, 46, 47, 48

K. Dadi, M. Rahim, A. Abraham, D. Chyzhyk, M. Milham, B. Thirion, G. Varoquaux, A. D. N. Initiative, et al. Benchmarking functional connectome-based predictive models for resting-state fmri. *Neuroimage*, 192:115–134, 2019. 9, 27

J. S. Damoiseaux, S. A. R. B. Rombouts, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, and C. F. Beckmann. Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences*, 103(37):13848–13853, 2006. ISSN 0027-8424. doi: 10.1073/pnas.0601417103. URL https://www.pnas.org/content/103/37/13848. 78, 85

G. Deco, G. Tononi, M. Boly, and M. L. Kringelbach. Rethinking segregation and integration: contributions of whole-brain modelling. *Nat Rev Neurosci*, 16(7):430–9, 2015. ISSN 1471-0048. doi: 10.1038/nrn3963. URL https://www.ncbi.nlm.nih.gov/pubmed/26081790. 73

N. U. Dosenbach, B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-Schlaggar, et al. Prediction of individual brain maturity using fmri. *Science*, 329(5997):1358–1361, 2010. 9

J. Dubois and R. Adolphs. Building a science of individual differences from fmri. *Trends in Cognitive Sciences*, 20(6):425 – 443, 2016a. ISSN 1364-6613. doi: https://doi.org/10.1016/j.tics.2016.03.014. URL http://www.sciencedirect.com/science/article/pii/S1364661316300079. 31

J. Dubois and R. Adolphs. Building a science of individual differences from fmri. *Trends in cognitive sciences*, 20(6):425–443, 2016b. 9

J. Dubois, P. Galdi, L. K. Paul, and R. Adolphs. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756):20170284, 2018. 27, 38, 42

S. B. Eickhoff, R. T. Constable, and B. T. Yeo. Topographic organization of the cerebral cortex and brain cartography. *NeuroImage*, 170:332 – 347, 2018. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2017.02.018. URL http://www.sciencedirect.com/science/article/pii/S1053811917301222. Segmenting the Brain. 78

M. L. Elliott, A. R. Knodt, M. Cooke, M. J. Kim, T. R. Melzer, R. Keenan, D. Ireland, S. Ramrakha, R. Poulton, A. Caspi, et al. General functional connectivity: Shared features of resting-state and task fmri drive reliable and heritable individual differences in functional brain networks. *NeuroImage*, 189:516–532, 2019. 10, 17, 27

T. S. Evans and R. Lambiotte. Line graphs, link partitions, and overlapping communities. *Phys. Rev. E*, 80:016105, Jul 2009. doi: 10.1103/PhysRevE.80.016105. URL https://link.aps.org/doi/10.1103/PhysRevE.80.016105. 79

T. S. Evans and R. Lambiotte. Line graphs of weighted networks for overlapping communities. *The European Physical Journal B*, 77(2):265–272, 2010. ISSN 1434-6036. doi: 10.1140/epjb/e2010-00261-8. URL https://doi.org/10.1140/epjb/e2010-00261-8. 79

J. Faskowitz, F. Z. Esfahlani, Y. Jo, O. Sporns, and R. F. Betzel. Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture. *Nature neuroscience*, 23(12):1644–1654, 2020. 6

E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11):1664–1671, 2015a. 58

E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11):1664–1671, 2015b. 9, 30, 86

E. S. Finn, D. Scheinost, D. M. Finn, X. Shen, X. Papademetris, and R. T. Constable. Can brain state be manipulated to emphasize individual differences in functional connectivity? *Neuroimage*, 2017a. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2017.03.064. URL https://www.ncbi.nlm.nih.gov/pubmed/28373122. 74

E. S. Finn, D. Scheinost, D. M. Finn, X. Shen, X. Papademetris, and R. T. Constable. Can brain state be manipulated to emphasize individual differences in functional connectivity? *Neuroimage*, 160:140–151, 2017b. 2, 9

S. Fortunato and M. Barthélemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41, 2007. ISSN 0027-8424. doi: 10.1073/pnas.0605965104. URL https://www.pnas.org/content/104/1/36. 78

L. E. Frank and J. H. Friedman. A statistical view of some chemometrics regression tools. *Technometrics*, 35(2):109–135, 1993. 27

K. J. Friston. Functional and effective connectivity: A review. *Brain Connectivity*, 1(1):13–36, 2011. doi: 10.1089/brain.2011.0008. URL https://doi.org/10.1089/brain.2011.0008. PMID: 22432952. 78

J. A. Gallego, M. G. Perich, L. E. Miller, and S. A. Solla. Neural manifolds for the control of movement. *Neuron*, 94(5):978–984, 2017. ISSN 0896-6273. 3, 47, 48, 73

P. Gao and S. Ganguli. On simplicity and complexity in the brave new world of large-scale neuroscience. *Current opinion in neurobiology*, 32:148–155, 2015. ISSN 0959-4388. 6, 46

S. Gao, A. S. Greene, R. T. Constable, and D. Scheinost. Combining multiple connectomes via canonical correlation analysis improves predictive models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 349–356. Springer, 2018a. 31, 32

S. Gao, A. S. Greene, R. T. Constable, and D. Scheinost. Task integration for connectome-based prediction via canonical correlation analysis. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 87–91. IEEE, 2018b. 12

S. Gao, A. S. Greene, R. T. Constable, and D. Scheinost. Combining multiple connectomes improves predictive modeling of phenotypic measures. *Neuroimage*, 201:116038, 2019a. 15, 33

S. Gao, G. Mishne, and D. Scheinost. A hierarchical manifold learning framework for high-dimensional neuroimaging data. In *International Conference on Information Processing in Medical Imaging*, pages 631–643. Springer, 2019b. 47

J. O. Garcia, A. Ashourvan, S. Muldoon, J. M. Vettel, and D. S. Bassett. Applications of community detection techniques to brain graphs: Algorithmic considerations and implications for neural function. *Proceedings of the IEEE*, 106(5):846–867, 2018. 3

L. Geerligs, M. Rubinov, R. N. Henson, et al. State and trait components of functional connectivity: individual differences vary with mental state. *Journal of Neuroscience*, 35(41):13949–13961, 2015. 2

M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. Van Essen, M. Jenkinson, and W.-M. H. Consortium. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–24, 2013. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2013.04.127. URL `http://www.ncbi.nlm.nih.gov/pubmed/23668970`. 58

A. S. Greene, S. Gao, D. Scheinost, and R. T. Constable. Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, 9, 2018a. ISSN 2041-1723. doi: 10.1038/s41467-018-04920-3. 74

A. S. Greene, S. Gao, D. Scheinost, and R. T. Constable. Task-induced brain state manipulation improves prediction of individual traits. *Nature communications*, 9(1):2807, 2018b. 10, 16, 17, 26, 28, 31

W. L. Hamilton, R. Ying, and J. Leskovec. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584*, 2017. 75

J. V. Haxby, J. S. Guntupalli, A. C. Connolly, Y. O. Halchenko, B. R. Conroy, M. I. Gobbini, M. Hanke, and P. J. Ramadge. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2):404–416, 2011. ISSN 0896-6273. 47

T. He, R. Kong, A. J. Holmes, M. R. Sabuncu, S. B. Eickhoff, D. Bzdok, J. Feng, and B. T. Yeo. Is deep learning better than kernel regression for functional connectivity prediction of fluid intelligence? In *2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 1–4. IEEE, 2018. 27

A. J. Holmes, M. O. Hollinshead, T. M. O'Keefe, V. I. Petrov, G. R. Fariello, L. L. Wald, B. Fischl, B. R. Rosen, R. W. Mair, J. L. Roffman, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Scientific Data, Volume 2, Issue, pp. 150031 (2011).*, 2:150031, 2011. 3

C. J. Honey, R. Kotter, M. Breakspear, and O. Sporns. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc Natl Acad Sci U S A*, 104(24):10240–5, 2007. ISSN 0027-8424 (Print) 0027-8424 (Linking). doi: 10.1073/pnas.0701519104. URL `https://www.ncbi.nlm.nih.gov/pubmed/17548818`. 75

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154, 1962. 3

W. Huisinga, C. Best, R. Roitzsch, C. Schütte, and F. Cordes. From simulation data to conformational ensembles: Structure and dynamics-based methods. *Journal of computational chemistry*, 20(16):1760–1774, 1999. 73

R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo, D. A. Handwerker, S. Keilholz, V. Kiviniemi, D. A. Leopold, F. de Pasquale, O. Sporns, M. Walter, and C. Chang. Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*, 2013. doi: 10.1016/j.neuroimage.2013.05.079. 47, 75

L. A. Jorgenson, W. Newsome, D. J. Anderson, C. I. Bargmann, E. N. Brown, K. Deisseroth, J. P. Donoghue, K. L. Hudson, G. S. Ling, P. R. Macleish, E. Marder, R. A. Normann, J. R. Sanes, M. J. Schnitzer, T. J. Sejnowski, D. W. Tank, R. Y. Tsien, K. Ugurbil, and J. C. Wingfield. The brain initiative: Developing technology to catalyse neuroscience discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2015. doi: 10.1098/rstb.2014.0164. 46

A. Joshi, D. Scheinost, H. Okuda, D. Belhachemi, I. Murphy, L. H. Staib, and X. Papademetris. Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics*, 9(1):69–84, 2011. ISSN 1559-0089 (Electronic) 1539-2791 (Linking). doi: 10.1007/s12021-010-9092-8. URL `https://www.ncbi.nlm.nih.gov/pubmed/21249532`. 58

A. A. Joshi, M. Chong, J. Li, S. Choi, and R. M. Leahy. Are you thinking what i'm thinking? synchronization of resting fmri time-series across subjects. *NeuroImage*, 172:740–752, 2018. 53

W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32(5):922–923, 1976. 53

D. Kobak, W. Brendel, C. Constantinidis, C. E. Feierstein, A. Kepecs, Z. F. Mainen, X. L. Qi, R. Romo, N. Uchida, and C. K. Machens. Demixed principal component analysis of neural population data. *Elife*, 5, 2016. ISSN 2050-084X (Electronic) 2050-084X (Linking). doi: 10.7554/eLife.10989. URL `https://www.ncbi.nlm.nih.gov/pubmed/27067378`. 46

D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguná. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, 2010. 81

A. Krishnan, L. J. Williams, A. R. McIntosh, and H. Abdi. Partial least squares (pls) methods for neuroimaging: a tutorial and review. *Neuroimage*, 56(2):455–475, 2011. 27

J.-G. Lee, J. Han, and K.-Y. Whang. Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 593–604, 2007. 76

K. Lee, J.-M. Lina, J. Gotman, and C. Grova. Spark: Sparsity-based analysis of reliable k-hubness and overlapping network structure in brain functional connectivity. *NeuroImage*, 134:434–449, 2016. 3

E. S. Lutkenhoff, M. Rosenberg, J. Chiang, K. Zhang, J. D. Pickard, A. M. Owen, and M. M. Monti. Optimized brain extraction for pathological brains (optibet). *PLoS One*, 9(12):e115551, 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0115551. URL `https://www.ncbi.nlm.nih.gov/pubmed/25514672`. 58

M.-M. Mesulam. From sensation to cognition. *Brain: a journal of neurology*, 121(6):1013–1052, 1998. 3

D. Meunier, R. Lambiotte, and E. T. Bullmore. Modular and hierarchically modular organization of brain networks. *Front Neurosci*, 4:200, 2010. ISSN 1662-453X (Electronic) 1662-453X (Linking). doi: 10.3389/fnins.2010.00200. URL `https://www.ncbi.nlm.nih.gov/pubmed/21151783`. 75

G. Mishne, R. Talmon, R. Meir, J. Schiller, M. Lavzin, U. Dubin, and R. R. Coifman. Hierarchical coupled-geometry analysis for neuronal structure and activity pattern discovery. *IEEE Journal of Selected Topics in Signal Processing*, 10(7):1238–1253, 2016. ISSN 1932-4553. 46

R. P. Monti, R. Lorenz, P. Hellyer, R. Leech, C. Anagnostopoulos, and G. Montana. Decoding time-varying functional connectivity networks via linear graph embedding methods. *Frontiers in Computational Neuroscience*, 2017a. doi: 10.3389/fncom.2017.00014. 47

R. P. Monti, R. Lorenz, P. Hellyer, R. Leech, C. Anagnostopoulos, and G. Montana. Decoding time-varying functional connectivity networks via linear graph embedding methods. *Frontiers in computational neuroscience*, 11:14, 2017b. 2

T. M. Moore, S. P. Reise, R. E. Gur, H. Hakonarson, and R. C. Gur. Psychometric properties of the penn computerized neurocognitive battery. *Neuropsychology*, 29(2):235, 2015. 15

B. Mwangi, T. S. Tian, and J. C. Soares. A review of feature reduction techniques in neuroimaging. *Neuroinformatics*, 12(2):229–244, 2014. 27

B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis*, 21 (1):113–127, 2006. 48, 50

A. L. Nichols, T. Eichler, R. Latham, and M. Zimmer. A global brain state underlies c. elegans sleep behavior. *Science*, 356(6344):eaam6851, 2017. 3

M. Nickel and D. Kiela. Poincaré embeddings for learning hierarchical representations. In *Advances in neural information processing systems*, pages 6338–6347, 2017. 79, 81

R. A. Poldrack, E. Congdon, W. Triplett, K. Gorgolewski, K. Karlsgodt, J. Mumford, F. Sabb, N. Freimer, E. London, T. Cannon, et al. A phenome-wide examination of neural and cognitive function. *Scientific data*, 3(1):1–12, 2016a. 32, 37

R. A. Poldrack, E. Congdon, W. Triplett, K. J. Gorgolewski, K. H. Karlsgodt, J. A. Mumford, F. W. Sabb, N. B. Freimer, E. D. London, T. D. Cannon, and R. M. Bilder. A phenome-wide examination of neural and cognitive function. *Scientific Data*, 2016b. doi: 10.1038/sdata.2016.110. 57

R. A. Poldrack, E. Congdon, W. Triplett, K. J. Gorgolewski, K. H. Karlsgodt, J. A. Mumford, F. W. Sabb, N. B. Freimer, E. D. London, T. D. Cannon, and R. M. Bilder. A phenome-wide examination of neural and cognitive function. *Scientific Data*, 3(160110), 2016c. 1

J. Power, A. Cohen, S. Nelson, G. Wig, K. Barnes, J. Church, A. Vogel, T. Laumann, F. Miezin, B. Schlaggar, and S. Petersen. Functional network organization of the human brain. *Neuron*, 72 (4):665–678, Nov 2011a. ISSN 0896-6273. doi: 10.1016/j.neuron.2011.09.006. URL https://doi.org/10.1016/j.neuron.2011.09.006. 78, 83, 85

J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar, et al. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011b. 3

M. G. Preti, T. A. Bolton, and D. Van De Ville. The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage*, 2017. doi: 10.1016/j.neuroimage.2016.12.061. 47

M. D. Rosenberg, E. S. Finn, D. Scheinost, X. Papademetris, X. Shen, R. T. Constable, and M. M. Chun. A neuromarker of sustained attention from whole-brain functional connectivity. *Nature neuroscience*, 19(1):165–171, 2016. 9, 10, 28

M. D. Rosenberg, B. Casey, and A. J. Holmes. Prediction complements explanation in understanding the developing brain. *Nature communications*, 9(1):1–13, 2018. 9

M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–69, 2010. ISSN 1095-9572 (Electronic) 1053-8119 (Linking). doi: 10.1016/j.neuroimage.2009.10.003. URL https://www.ncbi.nlm.nih.gov/pubmed/19819337. 75

M. Saggar, O. Sporns, J. Gonzalez-Castillo, P. A. Bandettini, G. Carlsson, G. Glover, and A. L. Reiss. Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature communications*, 9(1):1–14, 2018. 73

M. Salehi, A. Karbasi, X. Shen, D. Scheinost, and R. T. Constable. An exemplar-based approach to individualized parcellation reveals the need for sex specific functional networks. *NeuroImage*, 170:54 – 67, 2018. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2017.08.068. URL http://www.sciencedirect.com/science/article/pii/S1053811917307139. Segmenting the Brain. 78, 85

M. Salehi, A. Karbasi, D. S. Barron, D. Scheinost, and R. T. Constable. Individualized functional networks reconfigure with cognitive state. *NeuroImage*, 206:116233, 2020. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2019.116233. URL http://www.sciencedirect.com/science/article/pii/S1053811919308249. 78, 88

G. Santhanam, B. M. Yu, V. Gilja, S. I. Ryu, A. Afshar, M. Sahani, and K. V. Shenoy. Factor-analysis methods for higher-performance neural prostheses. *J Neurophysiol*, 102(2):1315–30, 2009. ISSN 0022-3077 (Print) 0022-3077 (Linking). doi: 10.1152/jn.00097.2009. URL https://www.ncbi.nlm.nih.gov/pubmed/19297518. 46

T. D. Satterthwaite, J. J. Connolly, K. Ruparel, M. E. Calkins, C. Jackson, M. A. Elliott, D. R. Roalf, R. Hopson, K. Prabhakaran, M. Behr, et al. The Philadelphia Neurodevelopmental Cohort: a publicly available resource for the study of normal and abnormal brain development in youth. *Neuroimage*, 124:1115–1119, 2016. 10

D. Scheinost, S. H. Kwon, C. Lacadie, B. R. Vohr, K. C. Schneider, X. Papademetris, R. T. Constable, and L. R. Ment. Alterations in anatomical covariance in the prematurely born. *Cerebral Cortex*, 27(1):534–543, 2017. ISSN 1047-3211;1460-2199. doi: 10.1093/cercor/bhv248. 58

D. Scheinost, S. Noble, C. Horien, A. S. Greene, E. M. Lake, M. Salehi, S. Gao, X. Shen, D. O'Connor, D. S. Barron, et al. Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage*, 193:35–45, 2019. 17, 29, 30

L. D. Selemon and P. S. Goldman-Rakic. Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. *Journal of Neuroscience*, 8(11):4049–4068, 1988. 3

X. Shen, F. Tokoglu, X. Papademetris, and R. T. Constable. Groupwise whole-brain parcellation from resting-state fmri data for network node identification. *Neuroimage*, 82:403–415, 2013a. 58

X. Shen, F. Tokoglu, X. Papademetris, and R. T. Constable. Groupwise whole-brain parcellation from resting-state fmri data for network node identification. *Neuroimage*, 82:403–415, 2013b. 50, 83

X. Shen, E. S. Finn, D. Scheinost, M. D. Rosenberg, M. M. Chun, X. Papademetris, and R. T. Constable. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *nature protocols*, 12(3):506–518, 2017. 1, 4, 10, 11, 33

J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000. 50

J. M. Shine, O. Koyejo, P. T. Bell, K. J. Gorgolewski, M. Gilat, and R. A. Poldrack. Estimation of dynamic functional connectivity using multiplication of temporal derivatives. *Neuroimage*, 122:399–407, 2015. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2015.07.064. URL <GotoISI>://WOS:000363125200038. 55, 75

J. M. Shine, P. G. Bissett, P. T. Bell, O. Koyejo, J. H. Balsters, K. J. Gorgolewski, C. A. Moodie, and R. A. Poldrack. The dynamics of functional brain networks: Integrated network states during cognitive task performance. *Neuron*, 2016. doi: 10.1016/j.neuron.2016.09.018. 47, 55, 63, 73, 74, 75

J. M. Shine, M. Breakspear, P. Bell, K. E. Martens, R. Shine, O. Koyejo, O. Sporns, and R. Poldrack. The low dimensional dynamic and integrative core of cognition in the human brain. *bioRxiv*, page 266635, 2018. 6

J. M. Shine, M. Breakspear, P. T. Bell, K. A. Ehgoetz Martens, R. Shine, O. Koyejo, O. Sporns, and R. A. Poldrack. Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nat Neurosci*, 22(2):289–296, 2019. ISSN 1546-1726 (Electronic) 1097-6256 (Linking). doi: 10.1038/s41593-018-0312-0. URL `https://www.ncbi.nlm.nih.gov/pubmed/30664771`. 2, 3, 73

A. E. Sizemore and D. S. Bassett. Dynamic graph metrics: Tutorial, toolbox, and tale. *NeuroImage*, 180:417–427, 2018. ISSN 1053-8119. 75

G. W. Stewart. A krylov–schur algorithm for large eigenproblems. *SIAM Journal on Matrix Analysis and Applications*, 23(3):601–614, 2002. 36

I. Stitt, K. J. Hollensteiner, E. Galindo-Leon, F. Pieper, E. Fiedler, T. Stieglitz, G. Engler, G. Nolte, and A. K. Engel. Dynamic reconfiguration of cortical functional connectivity across brain states. *Scientific reports*, 7(1):8797, 2017. 3

J. Taghia, W. Cai, S. Ryali, J. Kochalka, J. Nicholas, T. Chen, and V. Menon. Uncovering hidden brain state dynamics that regulate performance and decision-making during cognition. *Nature communications*, 9(1):1–19, 2018. 73

B. T. Thomas Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fischl, H. Liu, and R. L. Buckner. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125–1165, 2011. doi: 10.1152/jn.00338.2011. URL `https://doi.org/10.1152/jn.00338.2011`. PMID: 21653723. 78, 85

D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80: 62–79, 2013a. 10, 32, 37, 83

D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, K. Ugurbil, and W.-M. H. C. P. Consortium. The wu-minn human connectome project: an overview. *Neuroimage*, 80: 62–79, 2013b. 57

T. Vanderwal, J. Eilbott, E. S. Finn, R. C. Craddock, A. Turnbull, and F. X. Castellanos. Individual differences in functional connectivity during naturalistic viewing conditions. *Neuroimage*, 157: 521–530, 2017. 9

G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Idrobo, Y. Schwartz, and B. Thirion. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *NeuroImage*, 145:166–179, 2017. 29

D. Vidaurre, S. M. Smith, and M. W. Woolrich. Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, 114(48):12827–12832, 2017. ISSN 0027-8424. 73, 75, 76

R. Whelan and H. Garavan. When optimism hurts: inflated predictions in psychiatric neuroimaging. *Biological psychiatry*, 75(9):746–748, 2014. 30

K. Wu, Y. Taki, K. Sato, Y. Sassa, K. Inoue, R. Goto, K. Okada, R. Kawashima, Y. He, A. C. Evans, and H. Fukuda. The overlapping community structure of structural brain network in young healthy individuals. *PLOS ONE*, 6(5):1–14, 05 2011. doi: 10.1371/journal.pone.0019608. URL `https://doi.org/10.1371/journal.pone.0019608`. 78

B. T. Yeo, F. M. Krienen, M. W. Chee, and R. L. Buckner. Estimates of segregation and overlap of functional connectivity networks in the human cerebral cortex. *Neuroimage*, 88:212–227, 2014. 3

B. T. Yeo, F. M. Krienen, S. B. Eickhoff, S. N. Yaakub, P. T. Fox, R. L. Buckner, C. L. Asplund, and M. W. Chee. Functional specialization and flexibility in human association cortex. *Cerebral cortex*, 25(10):3654–3672, 2015. 3

K. Yoshida, Y. Shimizu, J. Yoshimoto, M. Takamura, G. Okada, Y. Okamoto, S. Yamawaki, and K. Doya. Prediction of clinical depression scores and detection of changes in whole-brain using resting-state functional mri data with partial least squares regression. *PLoS ONE*, 12(7): e0179638, 2017. 1

Y. Zhong, H. Wang, G. Lu, Z. Zhang, Q. Jiao, and Y. Liu. Detecting functional connectivity in fmri using pca and regression analysis. *Brain topography*, 22(2):134–144, 2009. 27

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005. 27, 28