Effects of Postprocessing on Topology of FMRI Connectivity in Spatial and Temporal Domains

Keri Anderson
University of Utah

UUCS-18-002

School of Computing
University of Utah
Salt Lake City, UT 84112 USA

14 April 2018

Abstract

Functional MRI connectivity is a technique that uses the synchrony of functional magnetic resonance imaging (MRI) signal over time to infer a ”wiring diagram” between brain regions, or a brain network graph. Recent advances have suggested that topological data analysis may be used to obtain novel information about the structure and function of brain networks using functional MRI connectivity data. However, there is controversy in the field about what data should be used for constructing brain graphs. Specifically, the postprocessing steps taken to remove noise from functional MRI data may substantively affect the results obtained through topological data analysis. Moreover, it is unclear whether topological measures are more useful when applied to spatial or temporal components of functional MRI data.

A dataset from the Human Connectome Project from 1003 subjects, each with four independent high-quality functional MRI scans, was used to compute differences in graph-theoretic metrics and topological data analysis results for four distinct postprocessing pipelines that attempt to correct for different aspects of physiological noise within functional MRI data. Reproducibility of measures, as well as their ability to discriminate one subject from another (brain fingerprinting), was used to assess the relative strength or weakness of a postprocessing pipeline to yield informative data.

Further, a correlation of graph-theoretic and topological metrics was made to behavioral, demographic, and technical factors across subjects to determine which measures were most informative about aspects of human brain function. Measurements were applied to both temporal and spatial dimensions of the functional MRI signals to compare efficacy and reliability.

There were marked improvements in reproducibility, as measured by an intraclass correlation coefficient, as well as stronger correlations with behavior across subjects for the more highly processed functional MRI data. Specifically, the pipeline that used independent component analysis to remove many possible noise sources was most reproducible and yielded the most useful information about brain function and behavior, both for graph-theoretic and topological measurements. When considering graph-theoretic and topological metrics applied to spatial and temporal fMRI signals, both approaches had distinct advantages with complementary information. Graph-theoretic metrics for both time and space domains suggested information about different aspects of brain function, whereas topological measurements for time and space domains yielded roughly similar information. Applying results to the time domain may confer resilience of analyses to head motion artifacts. Topological analyses, whether in time or space domains, discriminate two sets of behavioral variables and may represent a novel phenotypic characterization of human brains.

EFFECTS OF POSTPROCESSING ON TOPOLOGY OF FMRI CONNECTIVITY IN SPATIAL AND TEMPORAL DOMAINS

by

Keri Anderson

A Senior Thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

B.S. Computer Science

School of Computing

The University of Utah

April 2018

Copyright © Keri Anderson 2018
All Rights Reserved

ABSTRACT

Functional MRI connectivity is a technique that uses the synchrony of functional magnetic resonance imaging (MRI) signal over time to infer a ”wiring diagram” between brain regions, or a brain network graph. Recent advances have suggested that topological data analysis may be used to obtain novel information about the structure and function of brain networks using functional MRI connectivity data. However, there is controversy in the field about what data should be used for constructing brain graphs. Specifically, the postprocessing steps taken to remove noise from functional MRI data may substantively affect the results obtained through topological data analysis. Moreover, it is unclear whether topological measures are more useful when applied to spatial or temporal components of functional MRI data.

A dataset from the Human Connectome Project from 1003 subjects, each with four independent high-quality functional MRI scans, was used to compute differences in graph-theoretic metrics and topological data analysis results for four distinct postprocessing pipelines that attempt to correct for different aspects of physiological noise within functional MRI data. Reproducibility of measures, as well as their ability to discriminate one subject from another (brain fingerprinting), was used to assess the relative strength or weakness of a postprocessing pipeline to yield informative data.

Further, a correlation of graph-theoretic and topological metrics was made to behavioral, demographic, and technical factors across subjects to determine which measures were most informative about aspects of human brain function. Measurements were applied to both temporal and spatial dimensions of the functional MRI signals to compare efficacy and reliability.

There were marked improvements in reproducibility, as measured by an intraclass correlation coefficient, as well as stronger correlations with behavior across subjects for the more highly processed functional MRI data. Specifically, the pipeline that used independent component analysis to remove many possible noise sources was most repro-

ducible and yielded the most useful information about brain function and behavior, both for graph-theoretic and topological measurements. When considering graph-theoretic and topological metrics applied to spatial and temporal fMRI signals, both approaches had distinct advantages with complementary information. Graph-theoretic metrics for both time and space domains suggested information about different aspects of brain function, whereas topological measurements for time and space domains yielded roughly similar information. Applying results to the time domain may confer resilience of analyses to head motion artifacts. Topological analyses, whether in time or space domains, discriminate two sets of behavioral variables and may represent a novel phenotypic characterization of human brains.

CONTENTS

ABSTRACT . . . . . **ii**

LIST OF FIGURES . . . . . **vii**

ACKNOWLEDGEMENTS . . . . . **ix**

CHAPTERS

1. INTRODUCTION **1**

1.1 Motivation 1
1.2 Aim 1: Analyze Differences in Postprocessing Methods 1
1.3 Aim 2: Analyze Information Contained in Temporal Components vs. Spatial Components 2
1.4 Overview 2

2. BACKGROUND AND RELATED WORK **3**

2.1 fMRI Data 3
2.1.1 BOLD Data 3
2.1.2 Sources of Artifact 3
2.1.3 fMRI Connectivity Correlation 4
2.2 Post-Processing Pipeline Methods 5
2.2.1 Noise Sources To Be Removed 6
2.2.1.1 Head Motion Correction 6
2.2.1.2 Normalization 6
2.2.1.3 Nuisance Regression 6
2.2.2 General Linear Models for Data Cleaning 7
2.3 Information Contained in Temporal Components vs. Spatial Components . . 9
2.4 Metrics Used for Evaluating Both Temporal and Spatial Data 11
2.4.1 Graph Theoretical Methods 11
2.4.1.1 Modularity 11
2.4.1.2 Characteristic Path Length 11
2.4.1.3 Global Efficiency 12
2.4.1.4 Betweenness Centrality 13
2.4.1.5 Eigenvector Centrality 13
2.4.1.6 Clustering Coefficient 14
2.4.2 Reproducibility Methods 14
2.4.3 Topological Methods 16
2.4.3.1 Explanation of Topological Methods 16
2.4.3.2 FMRI Topological Methods Application 17

3. METHODS **20**

3.1 Data Sources . . . 20
3.1.1 Human Connectome Project Funded by NIH . . . 20
3.1.2 Measures Obtained . . . 20
3.1.2.1 Brain Parcellation. . . . 20
3.1.2.2 Behavioral Measures. . . . 21
3.2 Postprocessing Methods . . . 25
3.2.1 Cleaning Methods . . . 25
3.2.2 Connectivity Matrix . . . 27
3.3 Temporal Domain Connectivity Matrix . . . 27
3.4 Evaluation Metrics for Both Temporal and Spatial Domains . . . 27
3.4.1 Mean Connectivity . . . 27
3.4.2 Graph Theoretical Methods . . . 28
3.4.3 Intraclass Correlation Coefficient . . . 29
3.4.4 Topological Barcodes . . . 30
3.4.5 Behavior Correlation . . . 30

4. RESULTS . . . 31

4.1 Postprocessing Strategy Results . . . 31
4.1.1 Effects of Postprocessing Strategy on Functional Connectivity . . . 31
4.1.1.1 Mean Difference . . . 31
4.1.1.2 ICC Reproducibility Metric . . . 33
4.1.2 Effects of Postprocessing Strategy on Graph Theoretical Measures . . . 34
4.1.3 Effects of Postprocessing Strategy on Topological Analysis . . . 46
4.1.3.1 ROI Barcode Creation . . . 46
4.1.3.2 ROI Population Barcodes . . . 46
4.1.4 ROI Behavior Correlation . . . 47
4.2 Comparing Temporal and Spatial Connectivity Graphs . . . 51
4.2.1 Functional Connectivity Results . . . 51
4.2.2 Graph Theoretical Measures . . . 52
4.2.3 Topological Analysis . . . 64
4.2.3.1 ROI and Time Domain Barcodes . . . 64
4.2.3.2 ICC for Time Domain. . . . 65
4.2.3.3 Time Domain Behavior Correlation. . . . 65

5. CONCLUSIONS . . . 67

5.1 Postprocessing Pipeline Findings . . . 67
5.1.1 Summary of Core Postprocessing Findings . . . 67
5.1.2 Graph Measure Results by Pipeline . . . 67
5.1.3 ICC Results by Pipeline . . . 68
5.1.4 Topological Results by Pipeline . . . 68
5.1.5 Behavior Results by Pipeline . . . 68
5.1.6 Other Pipelines To Consider . . . 69
5.1.7 Limitations . . . 69
5.2 Time vs Spatial Domain Findings . . . 69
5.2.1 Summary of Core Time vs Spatial Findings . . . 69
5.2.2 Graph Measure Differences Time vs Space . . . 70
5.2.3 Topology Differences Time vs Space . . . 70

5.2.4 Future Directions for Analysis using Time-Based Correlation Matrices . 71
REFERENCES **73**

LIST OF FIGURES

2.1 Time series with scan parcellated into 361 ROIs. 4
2.2 Edge $e_{ij} = 1 \Rightarrow node_i$ and $node_j$ are connected. 5
2.3 Covariate file X provided with the Human Connectome Project data 7
2.4 Calculating the "cleaned" data. 8
2.5 Calculating $(X^T * X)^{-1}$ using Gauss-Jordan elimination 8
2.6 Transposing a vector of values for a node across time 9
2.7 Sliding window approach to measuring connectivity 10
2.8 Similar activity across time points 11
2.9 Modularity: Nodes clustered into dense connections with sparse connections in between. 12
2.10 Characteristic path length: average path length between nodes. 12
2.11 Global efficiency: a measure of how efficiently information is exchanged. 13
2.12 Betweenness centrality measures the degree to which a node behaves as a hub. 13
2.13 Eigenvector centrality is the degree a node is connected to highly influential nodes. 14
2.14 Clustering coefficient is the fraction of possible triangles a node participates in. 15
2.15 Pairwise Euclidean distances between two points. 16
2.16 As the radius size of discs around points increases, points merge into clusters. 17
2.17 Dimension 0 barcodes. 17
2.18 Dimension 1 topological feature. 18
2.19 Barcodes reordered to create a sigmoid curve. 18
4.1 Mean difference in connectivity values between post-processing methods. . . . 32
4.2 ICC measures for four cleaning pipelines. 33
4.3 Modularity results by processing type for ROI. 34
4.4 Characteristic path length results by processing type for ROI. 35
4.5 Global efficiency results by processing type for ROI. 36
4.6 Mean and median betweenness centrality values averaged across subjects. . . . 37
4.7 ICC values by ROI for betweenness centrality across cleaning pipelines. 38
4.8 Mean betweenness centrality values shown superimposed on brain ROIs. 39

4.9 Median and mean eigenvector centrality values averaged across subjects. 40
4.10 ICC values for eigenvector centrality graph measures. 41
4.11 Mean eigenvector centrality values shown superimposed on brain ROIs. 42
4.12 Median and mean clustering coefficient Values averaged across subjects. 43
4.13 ICC values for clustering coefficient graph measures 44
4.14 Mean clustering coefficient values shown superimposed on brain ROIs. 45
4.15 Reordering barcodes to create population sigmoid curves. 46
4.16 Population barcodes by processing type. 47
4.17 Average population barcodes and ROI ICC values by cleaning pipeline. 48
4.18 Correlation with behavior: minProc, Hm/Wm/Gs. 49
4.19 Correlation with behavior: Hm/Wm/Csf/Gs, ICA 50
4.20 ICC values for ROI barcodes. 51
4.21 Modularity results for ROI and time domain correlation matrices. 52
4.22 Modularity correlation results across ROI and time domains. 53
4.23 Characteristic path length results for ROI and time domain correlation matrices. 54
4.24 Characteristic path length correlation results across ROI and time domains. . . 55
4.25 Global efficiency results for ROI and time domain correlation matrices. 56
4.26 Global efficiency correlation results across ROI and time domains. 57
4.27 Betweenness centrality results for ROI and time domain correlation matrices. . 58
4.28 Betweenness centrality mean and median correlation across ROI and time domains. 59
4.29 Eigenvector centrality results for ROI and time domain correlation matrices. . . 60
4.30 Eigenvector centrality mean and median correlation across ROI and time domains. 61
4.31 Clustering coefficient results for ROI and time domain correlation matrices. . . 62
4.32 Clustering coefficient mean and median correlation across ROI and time domains. 63
4.33 Comparing ROI and time domain ICA barcodes. 64
4.34 ICC values for time domain barcodes. 65
4.35 Correlation with behavior: ROI vs time domain. 66
5.1 Behaviors Associated with Faster and Slower Barcode Convergence 71

ACKNOWLEDGEMENTS

Special thanks to Professor Bei Wang and Dr. Jeffrey Anderson for their guidance and patient help in producing these results. Thank you to Sourabh Palande for introducing and explaining dimension 0 persistent homology barcodes, and the R-TDA package to produce them.

CHAPTER 1

INTRODUCTION

1.1 Motivation

Functional magnetic resonance imaging (fMRI) uses a strong magnetic field and radio waves to measure the blood flow in the brain to detect areas of activity and has been a critical tool in discovering function of individual brain regions. By measuring synchrony of activity over time between regions, it is possible to measure *functional MRI connectivity (fcMRI)*. Measurements of functional connectivity have led to the discovery of intrinsic connectivity networks in the brain. However, many sources in brain images of shared variance or noise has nothing to do with brain function. In raw images, several artifacts such as head motion, heart rate, respiration, etc. are captured in the readings, resulting in "noisy" images that obscure the desired brain activity registration.

1.2 Aim 1: Analyze Differences in Postprocessing Methods

In order to obtain biologically meaningful results, it is critical to identify which aspects of the fMRI signal are related to neural activity, and which are noise. There is a general consensus in the scientific community that the fMRI images do need to be cleaned, but the overall pipeline may vary, and the optimal processing strategy is still in debate. Most pipelines include an aggressive head motion correction, normalization to a common space, and regression of physiological noise sources. However, it is not clear to what degree different cleaning pipelines may render consequential correlation and topological analysis moot: will the analysis results change if different pipelines are used?

1.3 Aim 2: Analyze Information Contained in Temporal Components vs. Spatial Components

Most of the research done analyzing functional MRI connectivity has focused on measuring average synchronization (activity) over time. Although this approach has led to a better understanding of brain network correlation, average correlation analysis will miss dynamically changing connections throughout the scan. Instead, we could flip the time and space dimensions and study how correlated time points are over nodes: at a time point t in a scan, how many nodes were coactivated, and how long did the correlation last? This approach shifts the focus to correlating ***duration*** and ***changes in connectivity*** and may allow asking more biologically relevant questions about how the brain functions in time.

1.4 Overview

The ultimate goal of this research is to find new ways to measure individual differences in human behavior in the brain. The temporal approach is particularly attractive because while the architecture in the brain is similar even in individuals with severe disabilities such as down syndrome, the timing of brain activation can vary dramatically from individual to individual and condition to condition. Establishing the robustness of postprocessing methods will increase confidence that these results represent meaningful differences in behaviors rather than artifacts of processing.

CHAPTER 2

BACKGROUND AND RELATED WORK

2.1 fMRI Data

Functional MRI images are acquired in the same way and with the same scanners as traditional MRI images, but they use subtle changes in signal intensity to measure shifting patterns of brain function over time. fMRI uses a particular type of pulse sequence called BOLD.

2.1.1 BOLD Data

BOLD (blood oxygen level dependent) data is recorded as a 3D brain image over time points, producing a series of 3D volumes. A typical patient scan will produce 1200 such volumes; for a series of 1200 sequential time periods, a 3D volume of the patient's brain will be recorded. A complete scan through the brain is performed every 0.7 to 3 seconds, repeated up to a thousand times or more, depending on technique.

2.1.2 Sources of Artifact

Many sources in brain images of shared variance or noise have nothing to do with brain function. In raw fMRI images, several artifacts such as head motion, heart rate, respiration, etc. are captured in the readings, resulting in "noisy" images that obscure the desired brain activity registration. In the raw images, synchronized noise components contribute more to the overall synchrony between brain regions than does the neural activity itself. These components include drifts in signal over time related to thermal properties of the MRI scanner, changes in blood oxygenation from variability in heart rate, swallowing, and breathing, head motion including micro-movements, and noise related to magnetic artifacts, such as from a truck driving by the building with the scanner or someone turning on a microwave [7]. If the brain moves even a millimeter, it can create large artifacts,

especially when these regions are on the edges of the brain or next to an interface with bone or fluid.

2.1.3 fMRI Connectivity Correlation

The traditional approach to brain connectivity is to analyze average brain activity over the time span of a scan. A brain image contains roughly 90 thousand or more voxels per scan, each of which has limited signal to noise, so it is common practice to group nearby voxels in regions for analysis. There are many parcellations of the brain, and grouping can be arbitrary. The grouping used for this analysis consists of 361 gray matter regions of interest (ROIs) across the brain. For each of the 361 ROIs (which can loosely be thought of as "nodes"), a time series vector of 1200 time points is created, as illustrated in Figure 2.1.

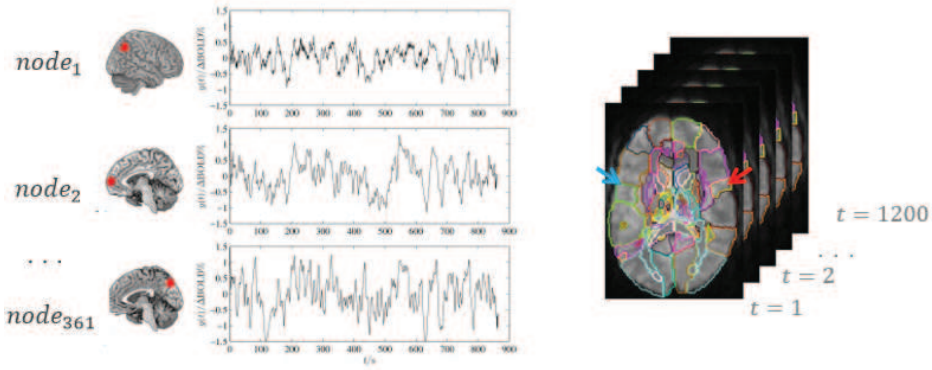

Figure 2.1. Time series with scan parcellated into 361 ROIs.

With the cleaned and normalized times series, we can represent connectivity between nodes as a graph G={V, E}, V={node$_i$}, E = {e$_{ij}$} where an edge exists between two nodes if the average activation over time is correlated greater than a given threshold. Traditional connectivity between two nodes is measured by a correlation coefficient between nodes:

- A threshold is chosen (which can be varied - strongly/weakly connected regions)
- A correlation coefficient c_{ij} between two time series is measured by the dot product of two time series, $c_{ij} = Y_i \cdot Y_j$, where Y_i is the time series for node$_i$ and Y_j is the time series for node$_j$

- If $c_{ij} >$ threshold $\Rightarrow$ there exists an edge between node$_i$ and node$_j$

Connectivity between nodes can be represented as a binary symmetric matrix, demonstrated in Figure 2.2.

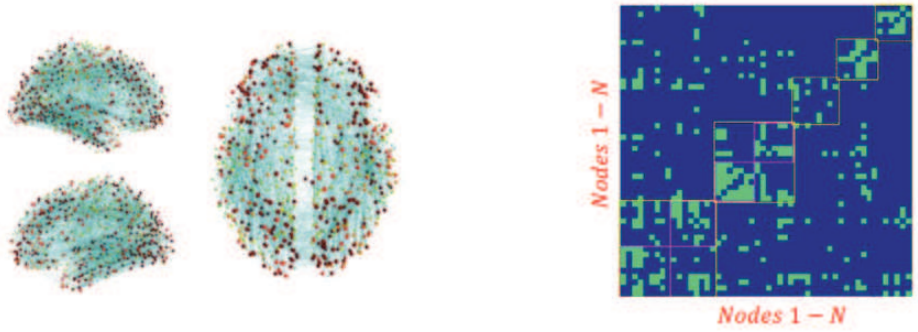

Figure 2.2. Edge $e_{ij} = 1 \Rightarrow$ node$_i$ and node$_j$ are connected.

For a fixed subject scan and a chosen postprocessing technique, a Pearson correlation coefficient is typically computed for each pair of rows in the 361 $\times$ 1200 time series, yielding 361 $\times$ 361 matrices of weighted connectivity.

2.2 Post-Processing Pipeline Methods

Specific postprocessing pipelines have been proposed, but there is no standardly implemented approach. There is controversy about specific steps, the order in which they should be performed, and why. Researches in this field agree that postprocessing is critical, and any of these schemes is better than none at all [11, 22]. Different postprocessing strategies will affect the results of functional connectivity, but any of them will likely be much more accurate than with no correction [32, 35].

The general consensus is to include some form of aggressive head motion correction [30], normalization to a common space, and regression of physiological noise sources (either by independent component analysis or general linear regression methods). Although there is an extensive literature discussing relative merits of different postprocessing strategies, there is no consensus on optimal strategies, although initial steps of normalizing brain images to a common template steps are well established and validated [18].

2.2.1 Noise Sources To Be Removed

2.2.1.1 Head Motion Correction

A complete scan through the brain is performed every 0.7 to 3 seconds, repeated up to a thousand times or more, depending on the technique. If the brain moves even a millimeter (which it does, all the time), then some regions of the brain can change signal intensity dramatically, especially when these regions are on the edges of the brain or next to an interface with bone or fluid. These large changes in signal can swamp the physiological signal, and are meaningless to analyze. Motion is particularly problematic for functional connectivity, because big spikes from motion dominate the time series. With motion, functional connectivity is ultimately a measurement of head motion, not brain function, and imaging science corrects for motion as thoroughly and completely as possible [29, 31].

2.2.1.2 Normalization

Everyone's brain has a different shape and size, but we want to compare findings in similar regions across brains. To do this, we have to "squish" or stretch a brain to be roughly the same size and shape as all the other brains we want to measure. There are two decades of literature with thousands of papers discussing methods for registration and normalization, including both linear and nonlinear methods. This remains an active area of research, but basic techniques are well-established and validated [18].

2.2.1.3 Nuisance Regression

Even with careful normalization and motion correction, many artifacts persist in the data. The remaining artifacts require some attempt to further clean the data. Covariate time series are identified that are thought to be noise, and these components are removed from the BOLD data at each point in the brain by linear regression.

However, the question is: which covariates should be used? There are two basic approaches to address the question. One explicitly identifies covariates of interest, and the other uses independent component analysis to identify components that are likely noise-related and regresses these out. In this project, we consider the following:

- Heart rate [9, 33]
- Breathing [3, 10, 19]

- Head motion (parameters obtained from motion correction step)
- White matter and CSF (Cerebrospinal fluid) [2]
- Global signal [2, 17, 25, 26]

An alternate approach is to use ICA-based methods to find regressors, identify which components are likely noise, and regress out these components [4, 21].

2.2.2 General Linear Models for Data Cleaning

When a scan is obtained, measurements of physiologic parameters can be simultaneously recorded. An fMRI study often includes a covariate file consisting of a matrix file X produced from the fMRI data itself, which estimates effects from extraneous factors (heart rate, etc.), as illustrated in Figure 2.3.

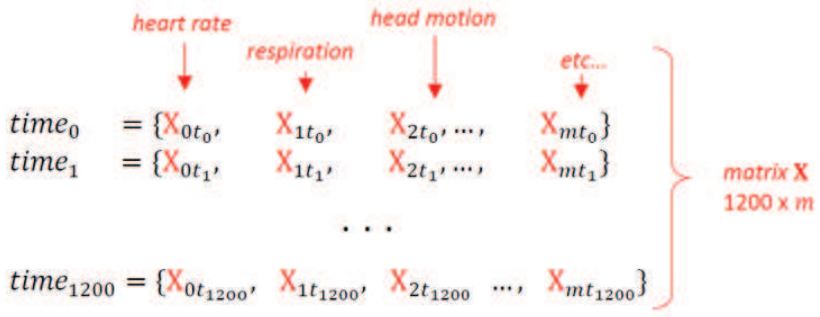

Figure 2.3. Covariate file X provided with the Human Connectome Project data

Each of the 361 ROIs is represented by a vector time series, Y_t, containing the measured values for that region across time. We are trying to find a best fit model for the time series using the covariates: how much of the time series can be explained by a linear combination of the covariates? The linear equation we wish to solve is

$$Y_t = \beta X + U_t \tag{2.1}$$

This is illustrated by Figure 2.4, where U_t represents the cleaned data matrix.

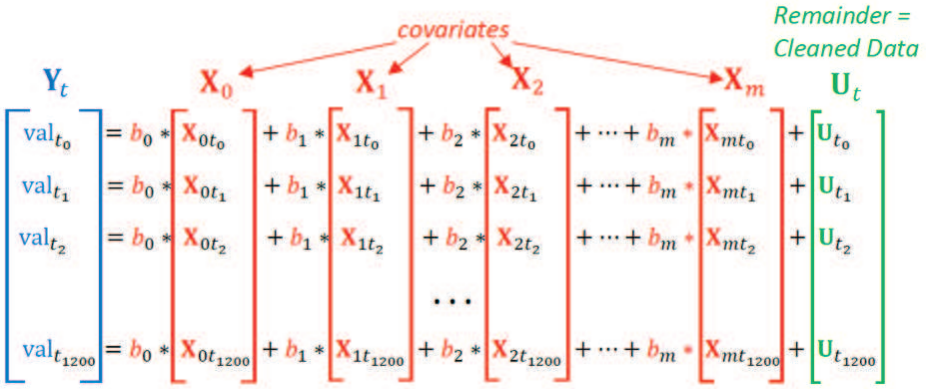

Figure 2.4. Calculating the "cleaned" data.

If we could calculate the value of the β matrix for each time series, we could extract the cleaned data U_t for that series. We use the following closed-form solution to solve for β:

$$\beta = (X^T * X)^{-1} * X^T * Y_t \tag{2.2}$$

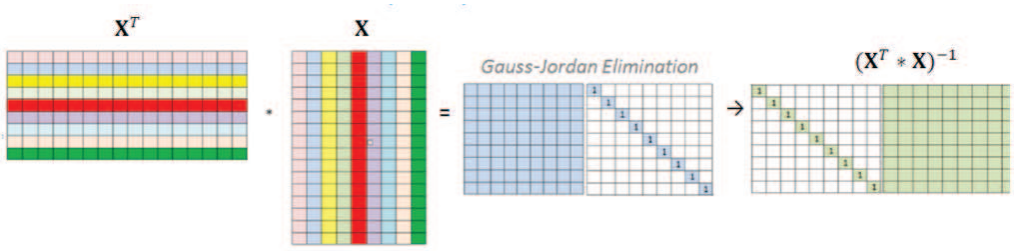

Figure 2.5. Calculating $(X^T * X)^{-1}$ using Gauss-Jordan elimination

The following steps are used to clean the data:

1. Calculate $(X^T * X)^{-1}$, as demonstrated in Figure 2.5.

2. Transpose fMRI data into "point vectors" and calculate $(X^T * Y_t)$. See Figure 2.6.

3. For ***each*** point vector Y_t, calculate a set of beta values $\beta = (X^T * X)^{-1} * X^T * Y_t$.

Figure 2.6. Transposing a vector of values for a node across time

4. Plug the calculated β values into the original equation $Y_t = \beta X + U_t$. The remainder U_t is the cleaned data for this time series.

5. Normalize the cleaned data to center around 0 by subtracting the mean and dividing by the standard deviation.

2.3 Information Contained in Temporal Components vs. Spatial Components

Traditionally, functional MRI connectivity has been performed as a measurement of connectivity between brain regions. Nevertheless, there is a symmetry where a time series can arbitrarily be seen as connectivity between time points across the brain. Although less intuitive, this perspective may offer some advantages. At a given time point in a scan, how many ROI nodes are registering similar activity? How repeatable is this pattern in the time points? Measuring connectivity in this manner may also be less sensitive to head motion noise, as the head motion will be uniquely contained in the time node itself and not likely correlated with other time nodes. Temporal information in brain connectivity may also be approached by using finer grained spatial connectivity. Several approaches have been proposed. One such approach is the "sliding window" to capture connectivity in smaller subsets of time [1]. This approach, however is limited by the fact that each window uses a small number of data points, producing noisy estimates of connectivity, as illustrated in Figure 2.7.

By inverting time and space, we are effectively asking which *time points* show similar patterns of relative brain activity. In other words, how frequently does a given pattern of brain network activity appear in the instantaneous brain activation time series? This data would allow comparison of dwell times in a given network, relative order in which

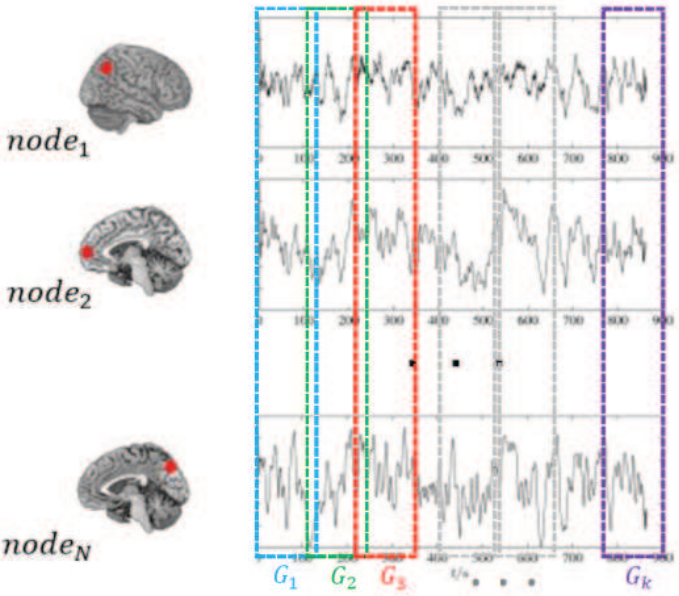

Figure 2.7. Sliding window approach to measuring connectivity

different networks are activated, and the relative duration a network persists. This type of information may speak to the dynamical stability of brain networks, rather than simply the architecture across the space of brain networks. Such dynamical information is likely to be of interest in studying brain development, mental illness, and brain function. Figure 2.8 shows two measured time points with locations of similar activity.

After connectivity is measured (using the same traditional approach of calculating correlation coefficients), the next step involves creating a graph over time points and noting how long different networks remain connected, what patterns see repetition, and topological relationships between clusters, which requires a look into topological analysis techniques currently applied to the traditional connectivity-over-time pipeline. If the time and space dimensions of fMRI data is flipped, what interesting graphical and topological results can be abstracted? Can we note how long brain networks are correlated, pick out specific patterns of correlation, and possibly note how separated brain networks are (hypothesized to be necessary for healthy brain function).

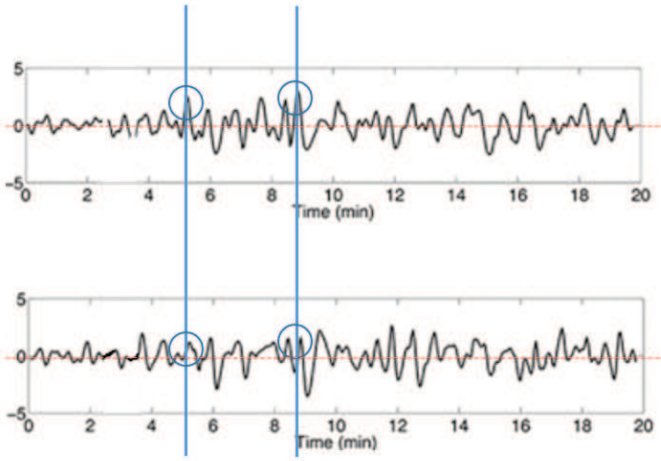

Figure 2.8. Similar activity across time points

2.4 Metrics Used for Evaluating Both Temporal and Spatial Data

2.4.1 Graph Theoretical Methods

2.4.1.1 Modularity

Modularity, shown in Figure 2.9, is a measure of the degree to which a graph's components may be separated into modules (also called groups, clusters, or communities). Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules. The optimal community structure is a subdivision of the network into nonoverlapping groups of nodes in a way that maximizes the number of within-group edges, and minimizes the number of between-group edges. Biological networks are thought to exhibit high degrees of modularity [27].

2.4.1.2 Characteristic Path Length

Characteristic path length, shown in Figure 2.10, is the average shortest path length in the network, or the number of steps along the shortest paths for all possible pairs of network nodes. It is a measure of the efficiency of information on a network, and is considered one of the more robust graph theoretical measures, particularly useful in neuroscience. The average path length distinguishes an easily negotiable network from one that is complicated and inefficient, with a shorter average path length being more desirable.

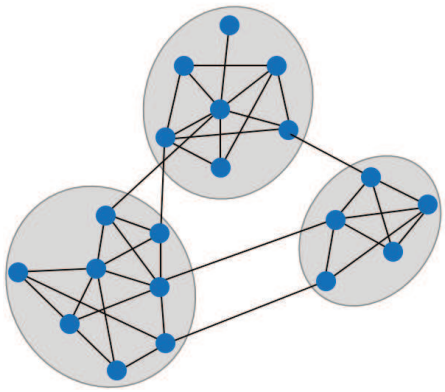

Figure 2.9. Modularity: Nodes clustered into dense connections with sparse connections in between.

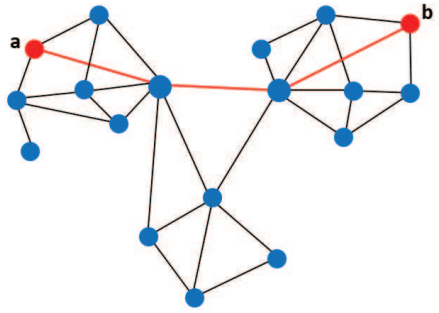

Figure 2.10. Characteristic path length: average path length between nodes.

2.4.1.3 Global Efficiency

Global efficiency, shown in Figure 2.10, is the average inverse shortest path length in the network, and is inversely related to the characteristic path length. For a specific node, local efficiency is the global efficiency computed on the neighborhood of the node, and it characterizes how efficiently information is exchanged by its neighbors if the node is removed. On the global scale, global efficiency quantifies the exchange of information across the whole network and how resistant to failure the network is.

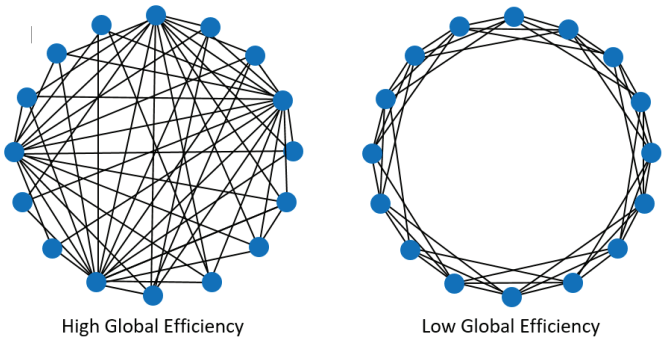

Figure 2.11. Global efficiency: a measure of how efficiently information is exchanged.

2.4.1.4 Betweenness Centrality

Node betweenness centrality is the fraction of all shortest paths in the network that contain a given node. Nodes with high values of betweenness centrality participate in a large number of shortest paths. Nodes with high values of betweenness centrality are considered "hubs" for the graph, because more information will pass through that node and therefore the node has a higher degree of influence. The red node in Figure 2.12 has a high betweenness centrality value.

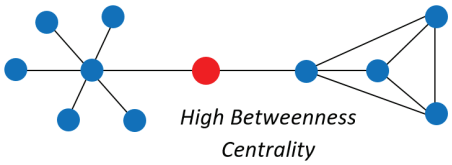

Figure 2.12. Betweenness centrality measures the degree to which a node behaves as a hub.

2.4.1.5 Eigenvector Centrality

Eigenvector centrality is also a measure of the degree to which a node behaves as a hub for the graph, and ultimately is a measure of how influential that node is. It is a self-referential measure of centrality. A high eigenvector score means that a node is connected to many nodes who themselves have high scores. Relative scores are assigned to all nodes

in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node than do equal connections to low-scoring nodes, as illustrated in Figure 2.13.

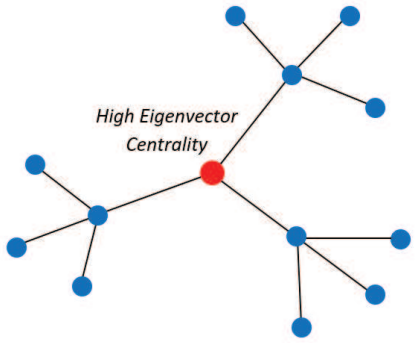

Figure 2.13. Eigenvector centrality is the degree a node is connected to highly influential nodes.

2.4.1.6 Clustering Coefficient

The clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. The clustering coefficient of an individual node is the degree to which neighbors of the node are also connected to each other, and can be visualized as the fraction of possible triangles the node participates in. A graph with an average high clustering coefficient is thought to be more robust and resilient to failure, as demonstrated in Figure 2.14.

2.4.2 Reproducibility Methods

There are two aspects of reliability of functional connectivity measurements. One is how consistently a certain measurement can be produced in the same subject, and the other is how well one subject can be discriminated from others. The intraclass correlation coefficient (ICC) is a well-accepted measurement that addresses both aspects of reliability. ICC is particularly convenient for studying the Human Connectome data, because this data set provided four scans for each subject. Although ICC is viewed as a type of correlation, unlike most other correlation measures it operates on data structured as groups,

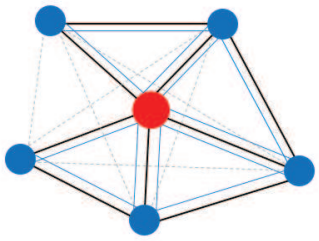

Figure 2.14. Clustering coefficient is the fraction of possible triangles a node participates in.

rather than paired observations. With the four scans, ICC can provide a sense of how strongly the four scans from a particular subject resemble each other as opposed to scans from other subjects in the group.

Several statistical methods have been proposed for calculating ICC, leading to some debate about which method should be used since different methods can lead to different results for the same data [24]. Modern ICC definitions resemble the random effect model,

$$Y_{ij} = \mu + \alpha_j + \epsilon_{ij}, \tag{2.3}$$

where Y_{ij} is the i^{th} observation in the j^{th} group, μ is an overall mean, α_j is the random effect share by all values in group j, and ϵ_{ij} is an unobserved noise term.

Matlab specifically uses the following: for a matrix M of n subjects and k observations,

$$MSR = var(mean(M, 2)) * k; \text{ (Across subject variance)} \tag{2.4}$$

$$MSW = sum(var(M, 0, 2))/n; \text{ (Average within subject variance)} \tag{2.5}$$

$$r = (MSR - MSW)/MSR; \tag{2.6}$$

To evaluate the "goodness" of an ICC score, the following guidelines are considered standard [12]:

- **Poor** Less than 0.40
- **Fair** 0.40 - 0.59
- **Good** 0.60 - 0.74
- **Excellent** 0.74 - 1.00

2.4.3 Topological Methods

2.4.3.1 Explanation of Topological Methods

A new approach to the analysis of graph representations of functional imaging data has been to extend concepts from topological data analysis (TDA) of time series to fMRI data [34, 36]. Persistent homology is a method for computing topological features of a space at different spatial metrics. To find the persistent homology of a space, the space must first be represented as a simplicial complex (a set of points and line segments). Given a set of points or nodes, we can calculate pairwise distances, often using straight forward Euclidean distances between two points, shown in Figure 2.15.

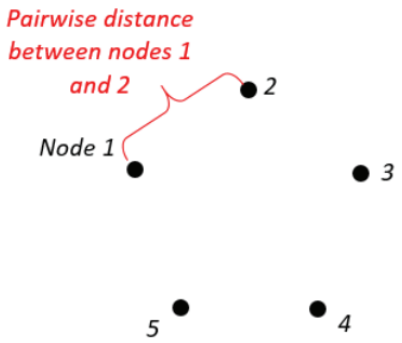

Figure 2.15. Pairwise Euclidean distances between two points.

With a distance metric defined on the set of points, discs centered at each point can be metaphorically drawn with a gradually increasing radius. Initially the radius starts out at 0 and each point is considered its own cluster. As the value of the radius continues to increases, certain discs will intersect, topologically merging two points into a single cluster, shown in Figure 2.15. The clustering of points are dimension 0 topological features for the set.

The distances at which points merge into clusters can be tracked by producing a set of barcodes representing the ”birth” and ”death” of a cluster. For dimension 0 features, all nodes are their own cluster with a birth of distance 0. The barcode ends for a node when it merges with another cluster. This is displayed in Figure 2.17.

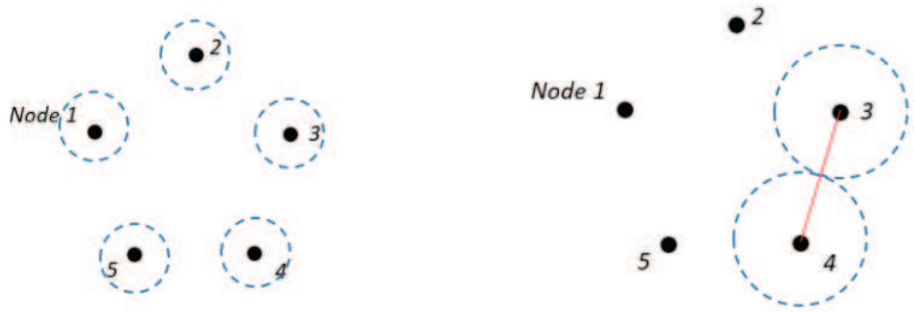

Figure 2.16. As the radius size of discs around points increases, points merge into clusters.

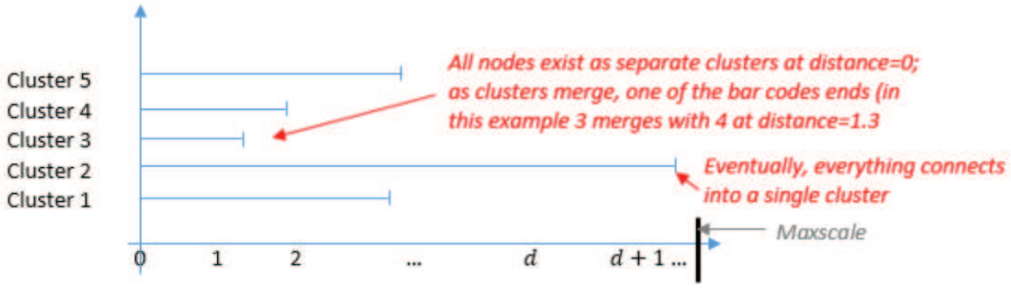

Figure 2.17. Dimension 0 barcodes.

The radius size can continue to increase, and eventually enough points can merge into clusters that form more complex structures, such as encapsulated balls in the center of the points. These are dimension 2 topological features, shown in Figure 2.18.

To get a visual sense of the dimension 0 topological behavior of a set, the barcodes can be reordered from shortest to longest. Using the endpoints of the barcodes, we can create a sigmoid curve to show how rapidly clusters are merging for the set, as shown in Figure 2.19.

2.4.3.2 FMRI Topological Methods Application

For the purposes of this study, the connectivity matrix itself is considered a simplicial complex composed of points (nodes in the graph) and line segments (connected nodes).

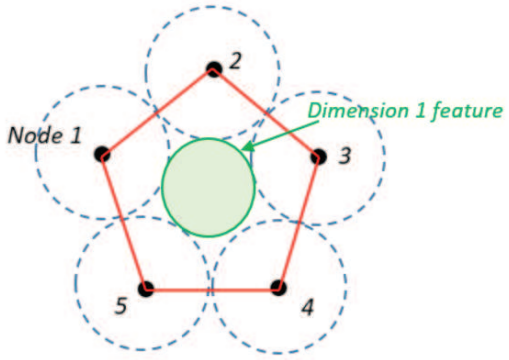

Figure 2.18. Dimension 1 topological feature.

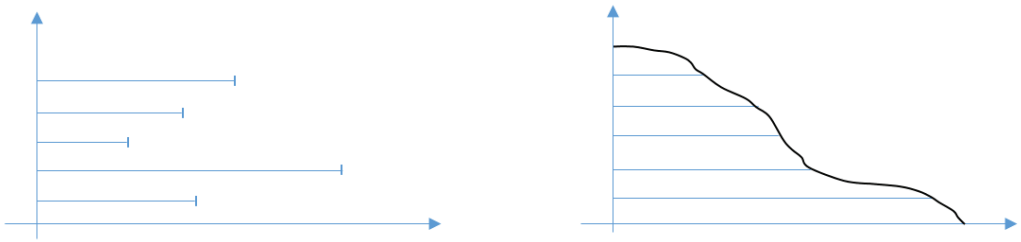

Figure 2.19. Barcodes reordered to create a sigmoid curve.

Only dimension 0 topological features (barcodes) are measured. A distance function is applied to the connectivity matrix, gradually increasing the allowed distance to measure when nodes merge into clusters.

To date, TDA methods for functional MRI connectivity are limited [14] , but an analogous approach has been described for structural MRI images [28]. More persistent features are detected over a wide range of spatial scales and are deemed more likely to represent true features of the underlying space rather than artifacts of sampling, noise, or particular choice of parameters [8, 13].

For each subject scan, a barcode representation is obtained, which can then be compared to other subjects, time points, or population average graphs. This approach assumes that the graphs representing brain connectivity may have higher order topological features

that can inform brain connectivity beyond simple graph-theoretical properties that have already been studied [6].

CHAPTER 3

METHODS

3.1 Data Sources

3.1.1 Human Connectome Project Funded by NIH

The BOLD dataset used was provied by the Human Connectome Project, consisting of 4 resting-state fMRI scans (60 minutes per subject) each from 1003 subjects with 534 female and 469 male participants, ages 21-35. In this dataset, there are 1200 time points in a single scan, with each of the time points capturing a 3D image volume of the brain. To account for artifacts at the beginning of each scan, I discarded the first 20 to obtain 1180 time points per scan.

IRB approval to analyze the Human Connectome Project dataset has been obtained from the University of Utah Institutional Review Board #IRB_00087725 Analysis of Restricted Human Connectome Project Data.

3.1.2 Measures Obtained

3.1.2.1 Brain Parcellation.

Resting functional MRI data was analyzed using a brain parcellation consisting of 333 regions in the cerebral cortex [20]. Fourteen participant-specific subcortical regions were added, using Freesurfer-derived segmentation [16] of bilateral thalamus, caudate, putamen, amygdala, hippocampus, pallidum, and nucleus accumbens, segmented independently for each participant. 14 cerebellar regions were also added comprising left and right-hemispheric representations of a 7-network parcellation [5]. This combined parcellation scheme covering cortex, subcortical structures, and the cerebellum comprised a total of 361 regions [15, 32]. Average BOLD time series were extracted for each volume in each run for each subject.

3.1.2.2 Behavioral Measures.

Provided with the Human Connectome data is a set of 81 behavioral, demographic, and technical factors for each subject. Each imaging metric of interest was correlated with the factors across subject. The behaviors are as follows:

1. **Age:** Age of patient in years
2. **Sex:** Female = 1, Male = 2
3. **Height:** Height in inches
4. **Weight:** Patient weight in kilograms
5. **Systolic:** Systolic blood pressure
6. **Diastolic:** Diastolic blood pressure
7. **Hematocrit_Mean:** Average of 2 hematocrit values (blood test)
8. **Education:** Years of education
9. **HoursSleep:** Hours of sleep per night over the last month (self report)
10. **Handedness:** Patient handedness (1 = right handed, -1 = left handed)
11. **MeanHeadMotion:** Average motion per volume in mm
12. **MaxHeadMotion:** Maximum head motion per scan in mm
13. **MeanDiffHeadMotion:** Average derivative of head motion in mm
14. **MaxDiffHeadMotion:** Maximum derivative of head motion in mm
15. **PicSeq_Unadj:** Picture Sequencing (Episodic Memory)
16. **CardSort_Unadj:** Wisconsin Card Sort (Cognitive Flexibility)
17. **Flanker_Unadj:** Flanker Task (Attention and Inhibitory Control)
18. **PMAT24_A_CR:** Progressive Matrices (Fluid Intelligence)
19. **ReadEng_Unadj:** Oral Reading Recognition Test (Reading Ability)

20. **PicVocab_Unadj:** Oral Picture Vocabulary: (Receptive Language Ability)
21. **ProcSpeed_Unadj:** Processing Speed
22. **DDisc_AUC_200:** Delay Discounting (Impulsivity)
23. **DDisc_AUC_400:** Delay Discounting (Impulsivity)
24. **VSPLOT_TC:** Variable Short Penn Line Orientation Test (Visuospatial Ability)
25. **SCPT_SEN:** Continuous Performance Test (Maintenance of Attention)
26. **SCPT_SPEC:** Continuous Performance Test (Inhibitory Control)
27. **IWRD_TOT:** Penn Word Memory Test (Verbal Episodic Memory)
28. **ListSort_Unadj:** List Sorting (Working Memory)
29. **ER40_CR:** Emotion Recognition
30. **AngAffect_Unadj:** NIH Toolbox Anger and Affect Survery (Attitudes of Anger)
31. **AngHostil_Unadj:** NIH Toolbox Anger and Affect Survery (Attitudes of Hostility)
32. **AngAggr_Unadj:** NIH Toolbox Anger and Affect Survery (Attitudes of Aggression)
33. **FearAffect_Unadj:** NIH Toolbox Fear-Somatic Arousal Survery (Attitudes of Fear)
34. **FearSomat_Unadj:** NIH Toolbox Fear-Somatic Arousal Survery (Attitudes of Somatic Arousal)
35. **Sadness_Unadj:** NIH Toolbox Sadness Survery (Attitudes of Sadness)
36. **LifeSatisf_Unadj:** NIH Toolbox Life Satisfaction Survey (Life Satisfaction)
37. **MeanPurp_Unadj:** NIH Toolbox Meaning and Purpose (Attitudes of Life Meaning/Purpose)
38. **PosAffect_Unadj:** NIH Toolbox Positive Affect (Psychological Well-Being)
39. **Friendship_Unadj:** NIH Toolbox Friendship Survery (Social Relationships)
40. **Loneliness_Unadj:** NIH Toolbox Loneliness Survery (Loneliness)

41. **PercHostil_Unadj:** NIH Toolbox Perceived Hostility Survery (Social Distress)
42. **PercReject_Unadj:** NIH Toolbox Perceived Rejection Survery (Perceived Social Rejection)
43. **EmotSupp_Unadj:** NIH Toolbox Emotional Support Survery (Social Support Advice)
44. **InstruSupp_Unadj:** NIH Toolbox Instrumental Support Survery (Social Network Resources)
45. **PercStress_Unadj:** NIH Toolbox Perceived Stress Survery (Perceived Stress Level)
46. **SelfEff_Unadj:** NIH Toolbox Self-Efficacy Survery (Sense of Control Over Life)
47. **GaitSpeed:** Speed of Walking
48. **Agreeableness:** Personality - Agreeableness
49. **Openness:** Personality - Openness to Experience
50. **Conscientiousness:** Personality - Conscientiousness
51. **Neuroticism:** Personality - Neuroticism
52. **Extroversion:** Personality - Extraversion
53. **ASR_Anxd_Raw:** Achenbach Adult Self-Report (Anxious/ Depressed)
54. **ASR_Witd_Raw:** Achenbach Adult Self-Report (Withdrawal Symptoms)
55. **ASR_Soma_Raw:** Achenbach Adult Self-Report (Somatic Complaints)
56. **ASR_Thot_Raw:** Achenbach Adult Self-Report (Thought Problems)
57. **ASR_Attn_Raw:** Achenbach Adult Self-Report (Attention Problems)
58. **ASR_Aggr_Raw:** Achenbach Adult Self-Report (Aggressive Behavior)
59. **ASR_Rule_Raw:** Achenbach Adult Self-Report (Rulebreaking Behavior)
60. **ASR_Intr_Raw:** Achenbach Adult Self-Report (Intrusive Thoughts)

61. **ASR_Intn_Raw:** Achenbach Adult Self-Report (Internalizing Symptoms)

62. **ASR_Extn_Raw:** Achenbach Adult Self-Report (Externalizing Symptoms)

63. **Agoraphobia:** Fear of Open Spaces

64. **BloodTHC:** Marijuana Detectable in Urine Test

65. **THCTimesUsed:** How Many Times Subject has Used Cannabis

66. **Drinks_7days:** How Many Alcoholic Drinks in Last 7 Days

67. **Tobacco_7days:** How Many Times Used Tobacco in Last 7 Days

68. **DSM_Depr_Raw:** Depression Score

69. **DSM_Anxi_Raw:** Anxiety Score

70. **DSM_Somp_Raw:** Somatic Problems Score

71. **DSM_Avoid_Raw:** Avoidant Personality Score

72. **DSM_Adh_Raw:** Attention Deficit/Hyperactivity Score

73. **DSM_Inat_Raw:** Inattention Score

74. **DSM_Hype_Raw:** Hyperactivity Score

75. **DSM_Antis_Raw:** Antisocial Personality Score

76. **Income:** Financial Income

77. **Relationship:** Is Subject in a Long-Term Relationship

78. **DepressiveSymptoms:** Symptoms of Depression

79. **MMSE:** Mini-Mental Status Exam (Basic Mental Function)

80. **GripStrength:** Grip Strength

81. **Dexterity:** Dexterity Test

3.2 Postprocessing Methods

3.2.1 Cleaning Methods

Five separate postprocessing pipelines were used to process the data. The first four pipelines used the minimally preprocessed data provided with the Human Connectome Project 1200 Subjects Release, and the last pipeline (ICA) used image data that had already been subject to the FIX ICA cleaning procedure [21]. Each post-processing pipeline resulted in a 361×1180 matrix of time series for each of the 361 nodes for each scan in each subject, which results in 1003 subjects $\times$ 4 scans $\times$ 361 nodes $\times$ 1180 time points $\times$ 5 post-processing pipelines.

1. **Minimally Processed Data.** This BOLD dataset is derived from images provided by the Human Connectome Project directly in its initial release. Some minimal steps that are required for the scans to be usable and are standard practice; the dataset provided has had these steps applied. The first involves normalization to MNI space. In order for brain scans to have widely applicable meaning, a standard spatial brain template is used to manipulate images to have the same spatial architecture. After normalization, a common coordinate system allows reference of similar positions in each individual's brain. When the fMRI scan is performed, with it comes an MPRAGE structural image that captures the spatial map of the individual's brain in high resolution. This image provides the parameters necessary to perform the transformation to MNI space. The postprocessed data also contains some motion correction. fMRI scans are hypersensitive to even a fraction of a millimeter of head motion (which happens frequently during a patient scan), and without some motion correction, the important signal is lost in motion noise. From each of these scans I have extracted time series from 361 ROI???s.

2. **Head Motion, White Matter, and CSF Regressed Data.** With the BOLD dataset described above, a linear regression analysis on each ROI's time series was performed. The first processing step aggressively corrected for head motion. The minimally postprocessed data does have some correction for head motion, but there are artifacts associated with the correction. When the initial motion correction is performed, motion parameters are saved in six directions: x, y, z Euclidean space as well as

pitch, *roll*, and *yaw* rotation directions. For each of the 1200 volumes, a time series of motion parameters records how much the brain moved from one image to the next. The artifacts in the data caused by head motion will track with motion parameters, and standard practice is to regress out the motion parameters and their derivatives (the motion parameters themselves capture how much the brain has moved, and the derivatives capture how much change there is between volumes). The head motion was regressed out using a standard Ordinary-Least-Squares regression.

In addition to motion correction, a correction was applied for noisy signal generated by white matter (axons that connect all the gray cells) and CSF (*cerebrospinal fluid*, liquid around the brain and spinal cord). During a scan, some signal will come from fluid inside and around the brain as the fluid responds to motion, breath, and pressure. There are no cells in this fluid, and we wish to remove any distracting signal the fluid produces. For white matter, there is a BOLD signal resulting from the blood flow to the axons, but this is not energy being produced by neurons so we need to remove it. Care must be taken to not include signal from gray matter, as this is what we are trying to measure, so the signals from CSF and white matter are "degraded". Every voxel that is one voxel away from gray matter is discarded, and the remaining voxels are averaged to obtain CSF and white matter time series for regression.

3. **Head Motion, White Matter, CSF, and Global Signal Regressed Data.** The removal of global signal is controversial and highly debated in its effectiveness. The process involves taking an average signal from the entire image over the brain, using it as a time series, and regressing it out. With the removal of global signal, inevitably some of the BOLD signal from gray matter will be regressed out as well as noise, so the risk is that important signal is being removed or smeared across the rest of the brain. We use this dataset because of its controversial nature to see how different the numeric matrices present after an aggressive global signal removal.

4. **ICA Data.** This dataset is also released directly from the Human Connectome Project. Another commonly used approach is to bypass all of the previous steps, and instead of explicitly going after the CSF, white matter, head motion, etc., using the

ICA method. This method algorithmically finds the principal components of the 3D volume matrices and classifies input as signal or noise. This dataset represents aggressively cleaned data, and is used by many investigators prior to calculating functional connectivity.

3.2.2 Connectivity Matrix

Connectivity matrix is also called the adjacency matrix, where a Pearson correlation coefficients were computed on the data from the above five post-processing techniques. For each pair of rows in the 361×1180 time series for each subject, scan, and processing technique, this yields 361×361 matrices of weighted connectivity.

3.3 Temporal Domain Connectivity Matrix

For this part of the study, only the data from ICA cleaned time series was used (with the time and space dimensions flipped). Each column in the 361×1180 times series for a subject scan contains the values of each ROI for a given time point. For each pair of columns, a Pearson correlation coefficient was performed, resulting in an 1180×1180 weighted connectivity matrix across time nodes. For this domain, it is no longer assumed that the nodes have any analogous attachment to a physical location in space, nor do they mean the same thing across subjects and scans. The nodes have meaning within the context of a single scan only, reflecting which time points behave similarly relative to other time points in the same scan. Because the matrices are orders of magnitude larger than their spatial counterparts, computational complexity rose significantly, with several studies taking over a week to finish computations.

3.4 Evaluation Metrics for Both Temporal and Spatial Domains

3.4.1 Mean Connectivity

Mean connectivity was accomplished by calculating the L2 norm for differences between post-processing strategies for each subject. Mean connectivity provides a quantitative estimate for how much the connectivity matrix is altered in each successive step of post-processing, suggesting which steps are most influential in altering the connectivity matrix.

3.4.2 Graph Theoretical Methods

For many of the graph theoretical methods detailed below, the connectivity matrix was first normalized with a normalization function *weight_conversion.m* provided by the Brain Connectivity Toolbox for Matlab. Normalization rescales all connectivity values to the range [-1, 1] and is required for several of the Brain Connectivity Toolbox graph functions. In addition, some of the graph theoretical methods also required a distance matrix as input rather than a connectivity matrix. Nodes that are highly connected have short distances close to 0, whereas nodes that are less connected or not connected will have distances approaching 1 or greater. To accomplish this, each element of the normalized connectivity matrix was subtracted from 1 providing a measure of distance rather than connectivity for that matrix element. That is, m_{ij} = 1 - m_{ij} for each element m_{ij} of the connectivity matrix $\boldsymbol{M}$. The revised matrix was input directly as an undirected and weighted graph.

1. **Modularity** was calculated using the function *modularity_und.m* for undirected graphs from the Brain Connectivity Toolbox for Matlab. Connectivity matrices were first normalized and then input directly as an undirected and weighted graph. Output was a single modularity value for each scan's connectivity matrix (or graph).

2. **Characteristic Path Length** was calculated using the function *charpath.m* for undirected graphs from the Brain Connectivity Toolbox for Matlab. In addition to normalization, this method required a distance matrix as opposed to a connectivity matrix. The distance matrix was created from the normalized connectivity matrix as describe above and used as input to the function. The output was a single average characteristic path length value for each scan's connectivity matrix.

3. **Global Efficiency** was calculated using the function *efficiency_wei.m* for undirected weighted graphs from the Brain Connectivity Toolbox for Matlab. In addition to normalization, this method required a distance matrix as opposed to a connectivity matrix. The distance matrix was created from the normalized connectivity matrix as describe above and used as input to the function. The output was a single average global efficiency value for each scan's connectivity matrix.

4. **Betweenness Centrality** was calculated using the function *betweenness_wei.m* for undirected weighted graphs from the Brain Connectivity Toolbox for Matlab. In addition to normalization, this method required a distance matrix as opposed to a connectivity matrix. The distance matrix was created from the normalized connectivity matrix as describe above and used as input to the function. The output was a vector of betweenness centrality values for each node in a connectivity matrix. To get a sense of graph-level performance, both the mean and median values were calculated from these vectors.

5. **Eigenvector Centrality** was calculated using the function *eigencentrality_und.m* for undirected graphs from the Brain Connectivity Toolbox for Matlab. In addition to normalization, this method required a distance matrix as opposed to a connectivity matrix. The distance matrix was created from the normalized connectivity matrix as describe above and used as input to the function. The output was an eigenvector associated with the largest eignevalue of the connectivity matrix. To get a sense of graph-level performance, both the mean and median values were calculated from these vectors.

6. **Clustering Coefficient** was calculated using the function *clusteringcoef_wu.m* for weighted undirected graphs from the Brain Connectivity Toolbox for Matlab. Connectivity matrices were first normalized and then input directly as an undirected and weighted graph. Output was a clustering coefficient vector with a value for each node. To get a sense of graph-level performance, both the mean and median values were calculated from these vectors.

3.4.3 Intraclass Correlation Coefficient

The Intraclass Correlation Coefficient (ICC) was calculated using the ICC.m [23] function in MATLAB using four independent measurements by comparing results for each of a subject's 4 scans. For each ICC calculation, a matrix $\boldsymbol{M}$ of 1003 subjects by 4 observations where each row is one subject and each column is a judge or measure. When calculating ICC for metrics with more than a single value per subject, such as the minimally processed connectivity matrix, the measurement had to be calculated separately for each node in the

graph. Input for node 1 consisted of measurements for 4 scans in all subjects for node 1. This was then repeated for node 2, etc.

3.4.4 Topological Barcodes

Topological dimension **0** barcodes were calculated with the R toolbox package TDA, statistical tools for data analysis [14]. Specifically connectivity matrices were converted into normalized distance matrices and used directly as input into function *ripsDiag* from the TDA library. The distance function used was $\sqrt{\mathbf{1} - \boldsymbol{corrcoef}_{ij}}$ for each $\boldsymbol{i, j}$ element of the correlation (connectivity) matrix. This function returned a text list of births and deaths of each cluster. The list was sorted from earliest deaths to most recent deaths to create a sigmoid curve to provide a sense of the clustering rate for the nodes. The text file was also read into Matlab for further behavioral correlation processing.

3.4.5 Behavior Correlation

Behavioral correlation was accomplished with Matlab's function *corrcoef* function. Input was two vectors of measurements for the desired correlation coefficient. For example, to assess correlation between the behavior *Agreeableness* and ROI barcodes, a vector of *Agreeableness* values for all subjects and a vector of sorted values for ROI node 1 barcodes is input to the correlation function to return a single scalar. This process was repeated for all behaviors and all node barcode values.

CHAPTER 4

RESULTS

4.1 Postprocessing Strategy Results

4.1.1 Effects of Postprocessing Strategy on Functional Connectivity

4.1.1.1 Mean Difference

To get a broad sense of how different postprocessing strategies were for subjects, for every subject a mean difference was calculated between each postprocessing type. To accomplish this, each subject's four scans were first averaged per processing type, resulting in 5 averaged connectivity matrices per subject. The 5 averaged connectivity matrices were then compared to each other to create a mean difference. For example, to compare ICA connectivity matrix to minimally cleaned connectivity matrix for a subject, element-wise difference were calculated, squared and summed, divided by the number of elements, and then the square root was taken. Mathematically,

$$\boldsymbol{meanDiff(ICA, minproc)} = \sqrt{\sum ((\boldsymbol{ICA}_{ij} - \boldsymbol{minproc}_{ij})^{\boldsymbol{2}})/\boldsymbol{numElements}} \tag{4.1}$$

This resulted in 5×5 mean difference matrices for each of 1003 subjects, with the matrix being symmetrical and redundant across the diagonal (i.e. mean difference between ICA and minproc is the same as the mean difference between minproc and ICA). These mean differences between processing types were then averaged for the population, and the bottom half of the resulting difference matrix is shown in Figure 4.1.

Because the minimally postprocessed and head motion results were so similar, the head motion processing was discarded for subsequent analyses. We have noticed that progressively more aggressive cleaning methods show larger differences in measured connectivity, not just to the minimally preprocessed data, but also to each other.

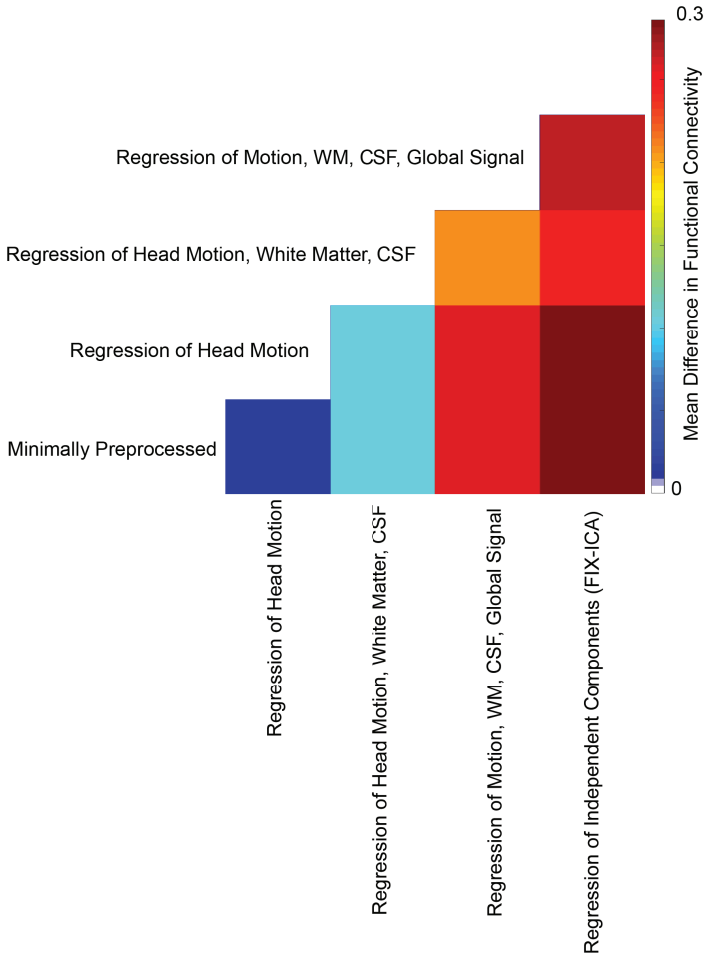

Figure 4.1. Mean difference in connectivity values between post-processing methods.

4.1.1.2 ICC Reproducibility Metric

Element-wise ICC values were calculated over each of the four cleaning pipeline connectivity matrices. Note that ICA clearly shows the strongest ICC reproducibility, whereas global signal shows a slight drop in reproducibility from the minimally processed data. Results are shown in Figure 4.2.

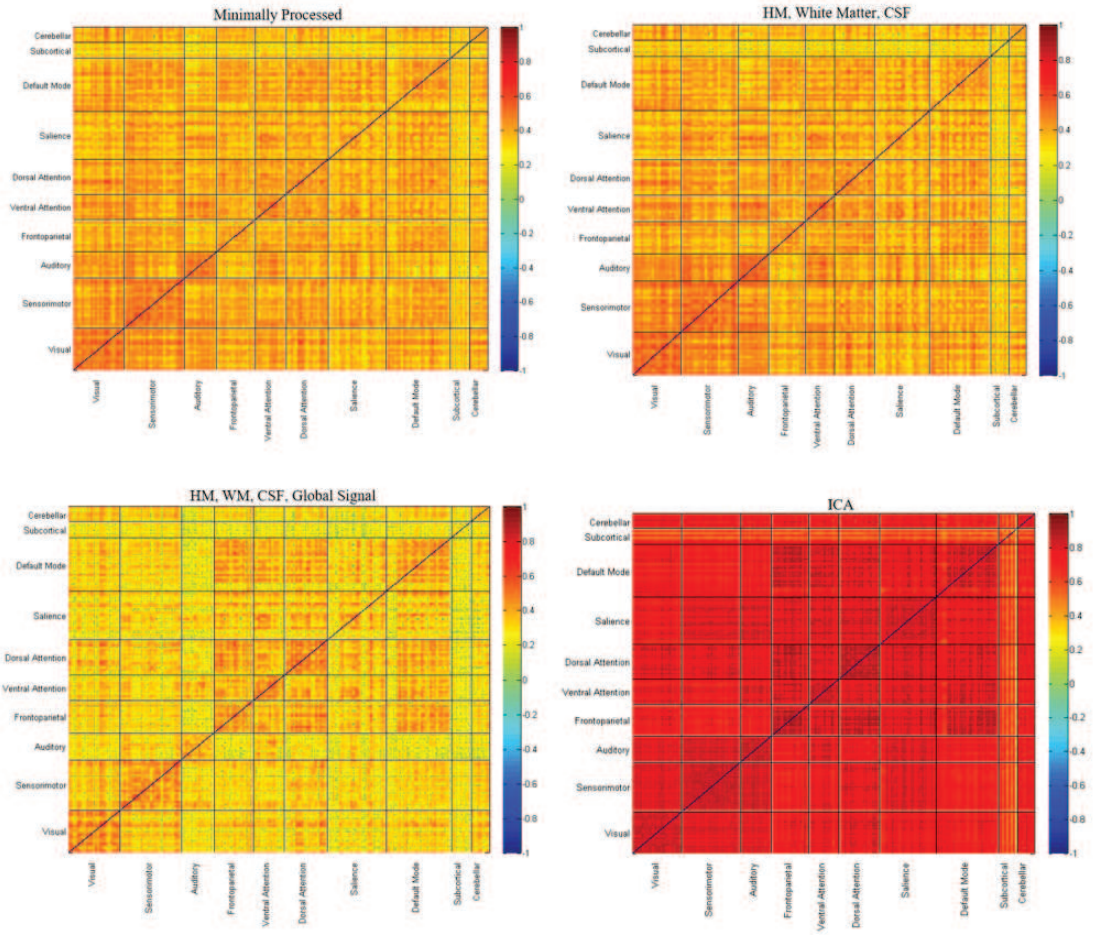

Figure 4.2. ICC measures for four cleaning pipelines.

4.1.2 Effects of Postprocessing Strategy on Graph Theoretical Measures

1. **Modularity.** Shown in Figure 4.3 are modularity and ICC results across processing type. Note that ICA processing reduces both the mean and variance of calculated modularity while simultaneously increasing ICC reproducibility. Removing the Global Signal has the opposite effect.

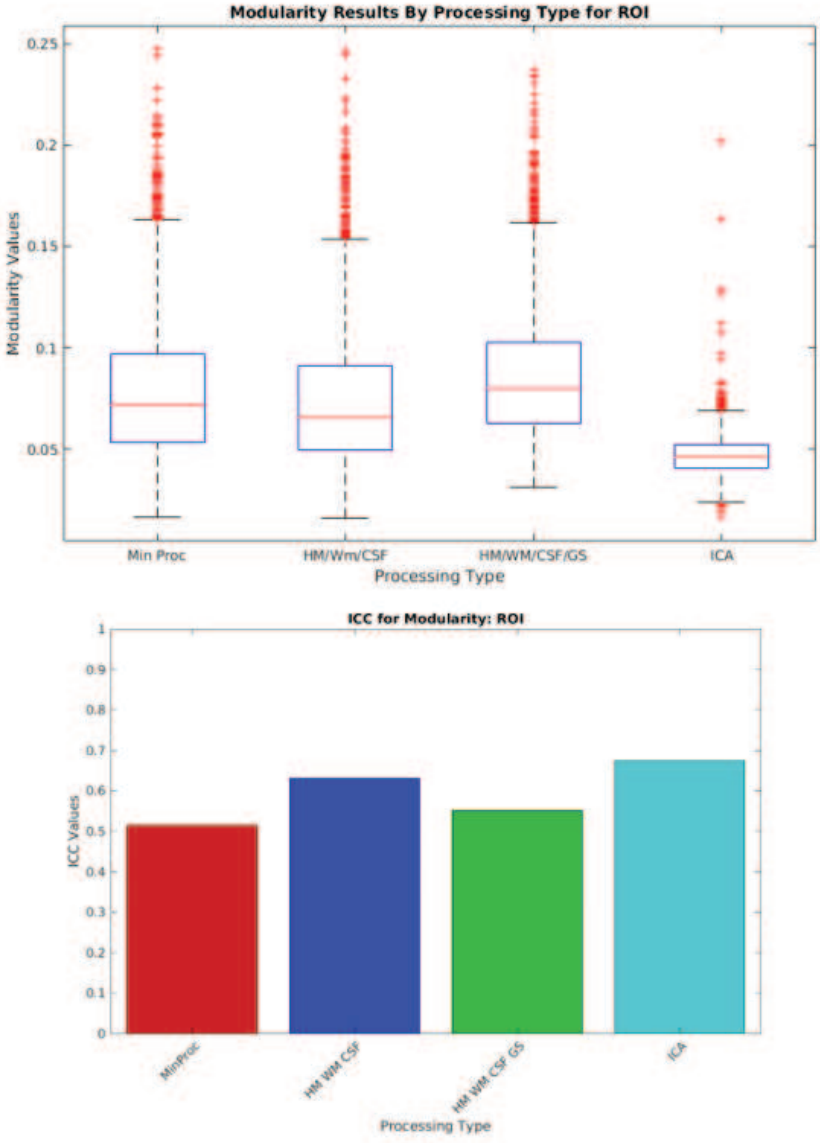

Figure 4.3. Modularity results by processing type for ROI.

2. **Characteristic Path Length.** Shown in Figure 4.4 are the results for characteristic path length and corresponding ICC measures by processing type. The more aggressive ICA cleaning process increases the average path length between nodes. ICC results remain stable across processing type except for a drop in Global Signal processing.

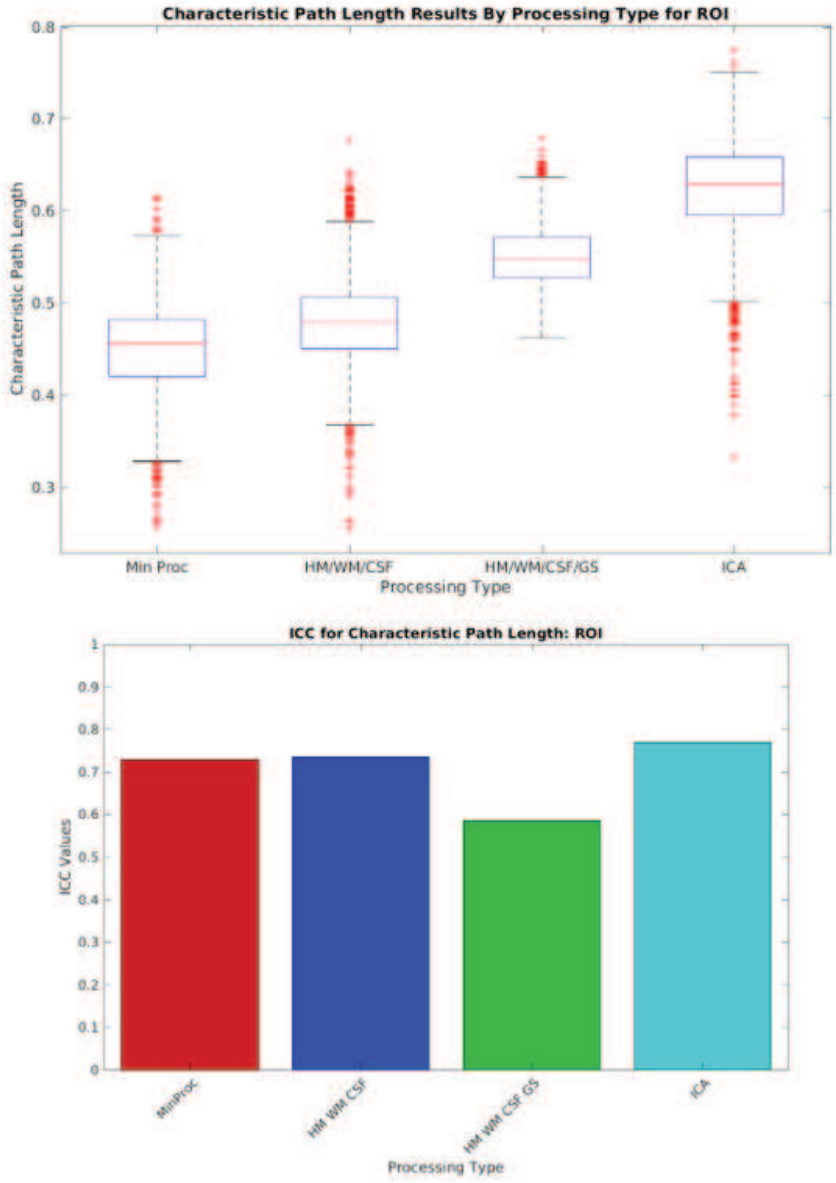

Figure 4.4. Characteristic path length results by processing type for ROI.

3. **Global Efficiency.** Shown in Figure 4.5 are the results for ROI global efficiency across cleaning pipelines. Interestingly, more aggressive cleaning pipelines suggest a higher Global Efficiency, or a more efficient exchange of information. Once again, ICC results drop with Global Signal removal.

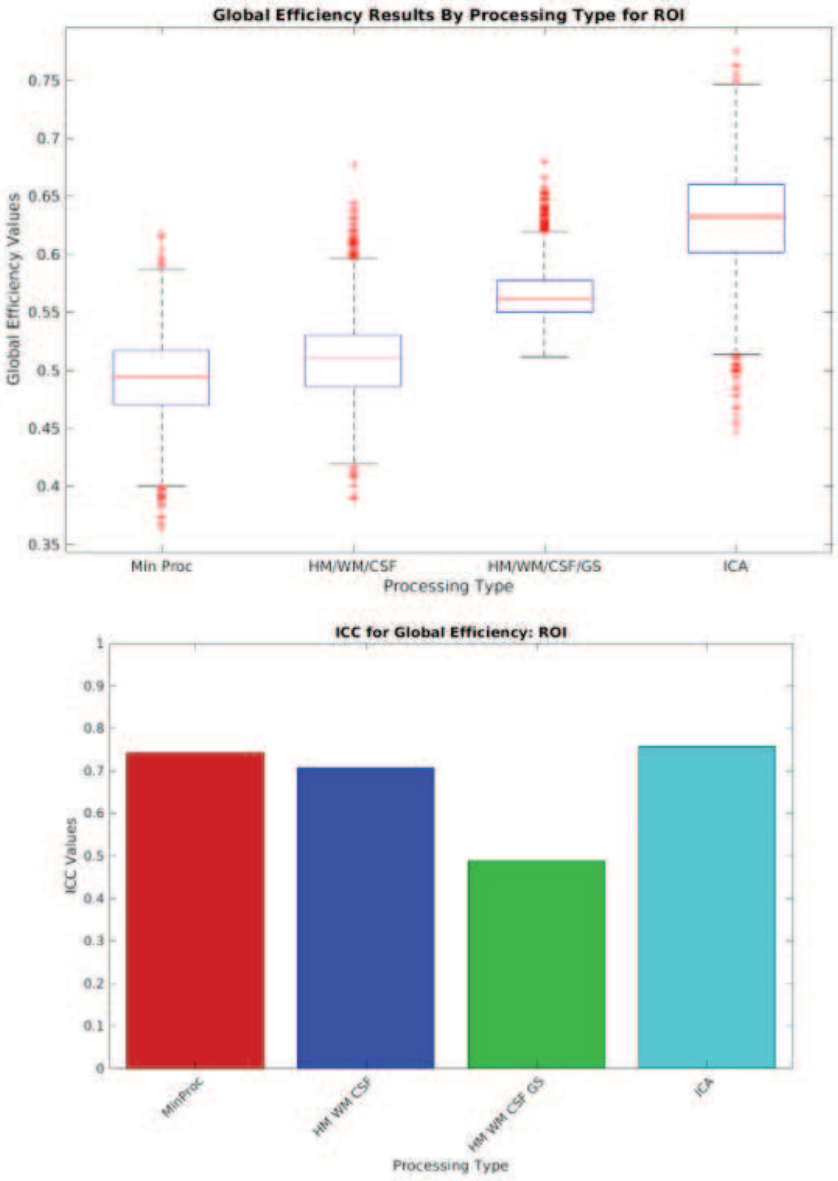

Figure 4.5. Global efficiency results by processing type for ROI.

4. **Betweenness Centrality.** Shown in Figure 4.6 are population mean and median results for Betweenness Centrality. Betweenness results were sparse vectors for all cleaning types, suggesting only a few outliers acted as strong hubs for the graph. ICA processing removed most outliers.

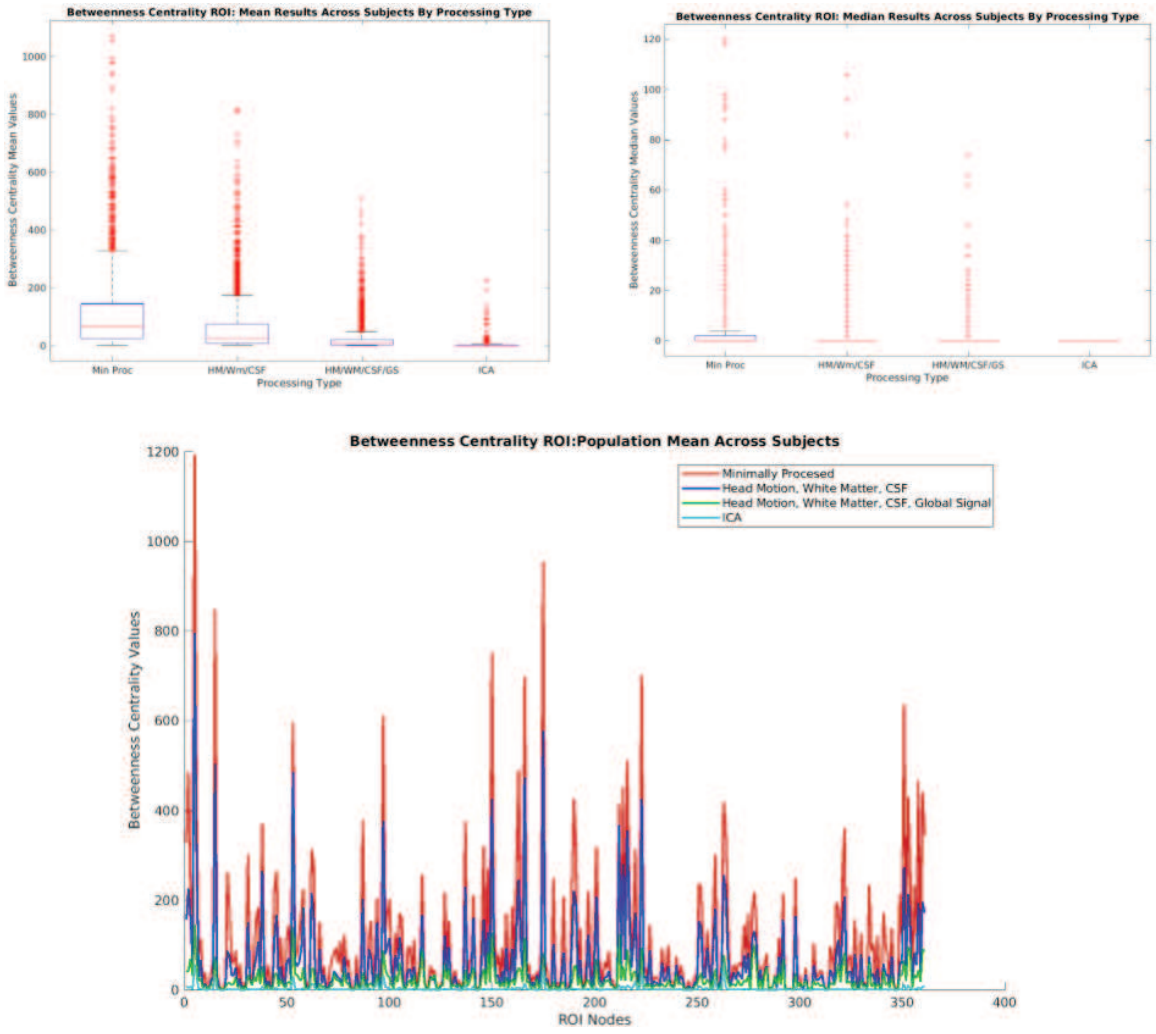

Figure 4.6. Mean and median betweenness centrality values averaged across subjects.

Figure 4.7 displays the ICC results for betweenness centrality for each ROI. Over all processing pipelines, ICC results for this graph measure were lower, perhaps because

the vectors were so sparse. Once again the higher ICC values surfaced with ICA processing, while the lowest values were reported in global signal processing.

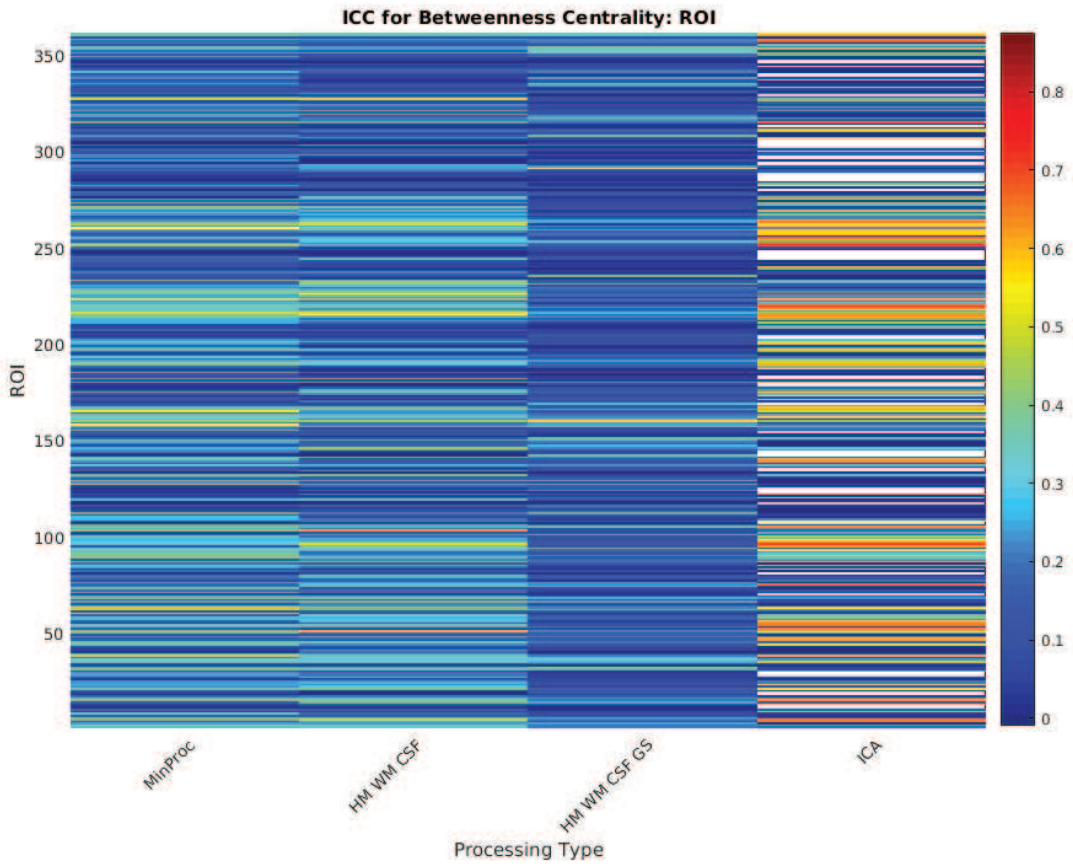

Figure 4.7. ICC values by ROI for betweenness centrality across cleaning pipelines.

Figure 4.8 shows population ROI mean betweenness centrality results mapped onto the brain surface. The relatively homogeneous blue color suggests that all regions are seen as hubs more or less equally for Betweenness Centrality results regardless of cleaning type. Aggressively cleaned ICA data more accurately localizes influential hubs within complex regions of the parietal lobes.

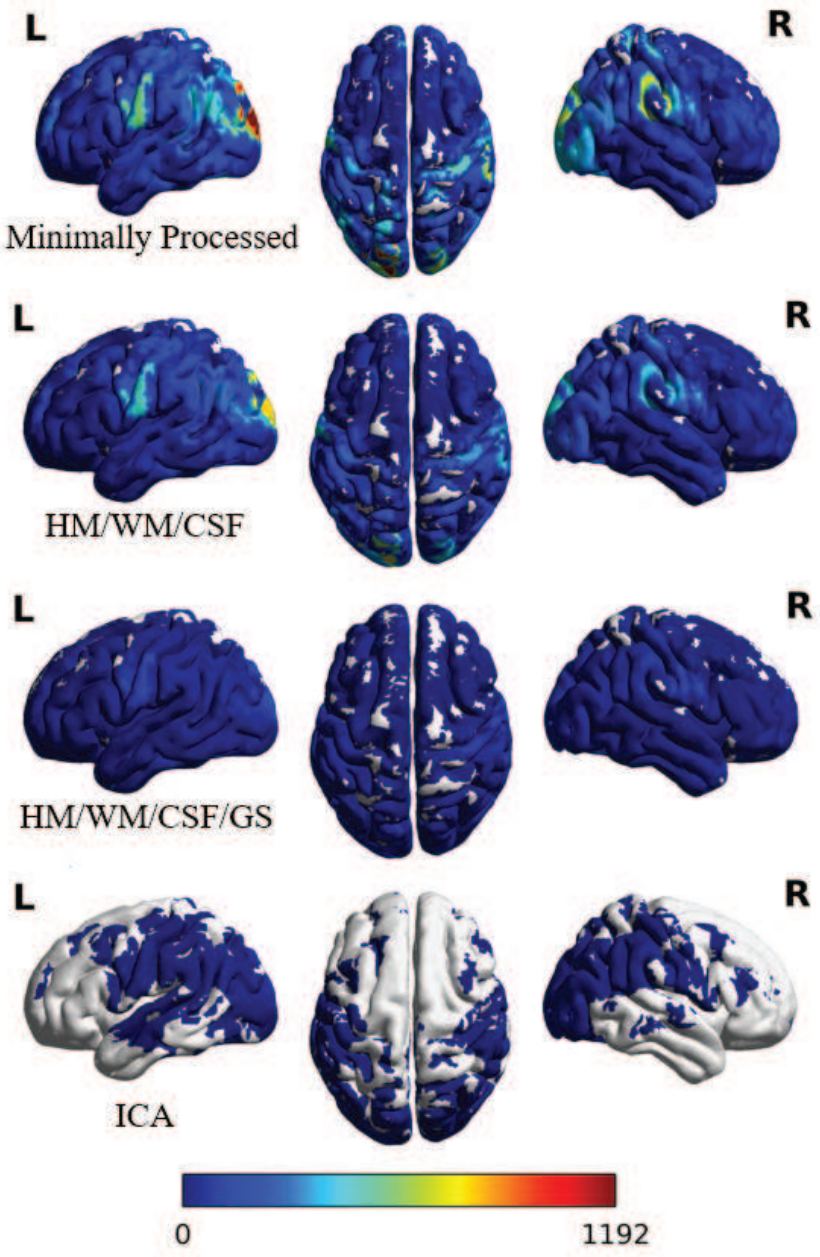

Figure 4.8. Mean betweenness centrality values shown superimposed on brain ROIs.

5. **Eigenvector Centrality.** Another capture of hub behavior is eigenvector centrality. Although the means and medians themselves remain relatively stable across cleaning types, the variance notably changes with ICA data showing the greatest mean variance, yet reduced median variance. Population mean and median eigenvector centrality are shown in Figure 4.9.

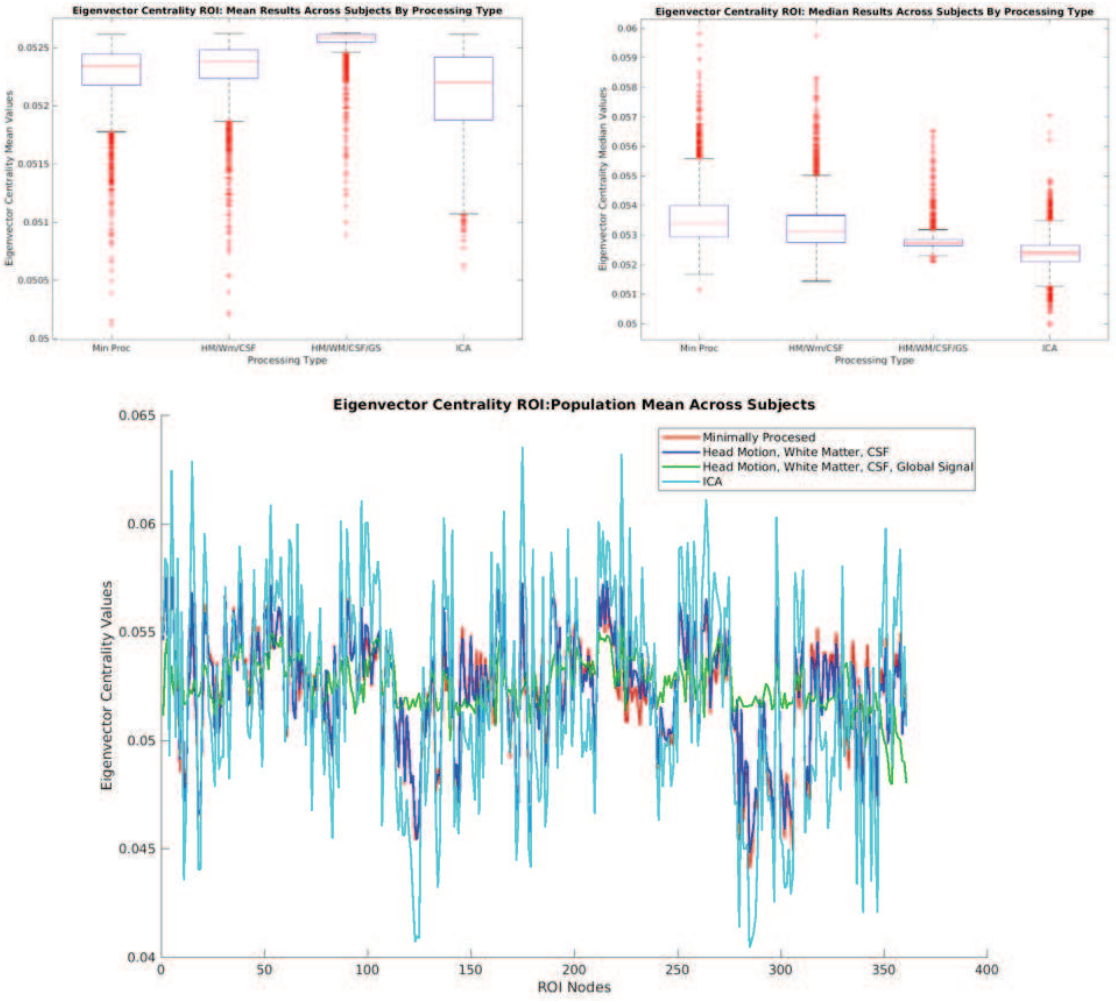

Figure 4.9. Median and mean eigenvector centrality values averaged across subjects.

Of particular note is the stark contrast in ICC results for eigenvector centrality. Minimally processed, HM/WW/CSF, HM/WW/CSF/Global Signal cleaning pipelines

all performed poorly, while ICA alone had remarkably high ICC reproducibility.

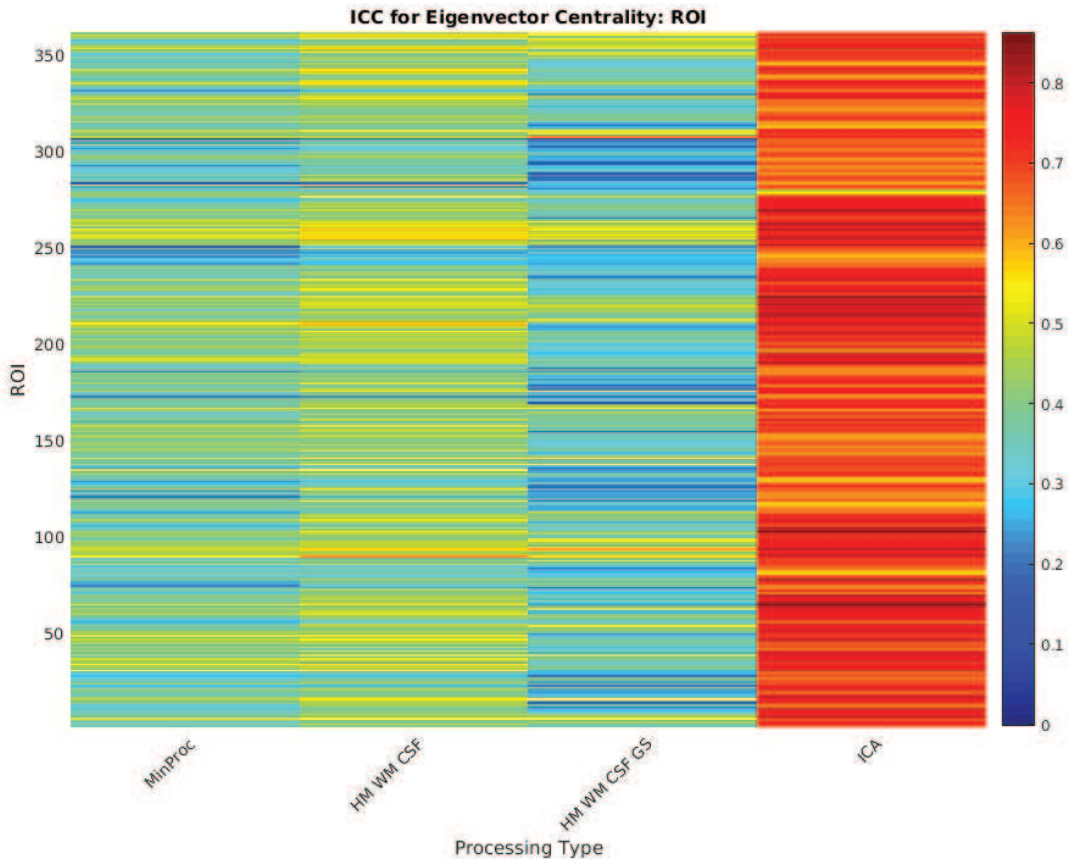

Figure 4.10. ICC values for eigenvector centrality graph measures.

Figure 4.11 shows population ROI mean eigenvector centrality results mapped onto brain models. While other graph measures were reduced by more aggressively cleaned data, in this case ICA data detects ROI hub behavior that other pipelines miss.

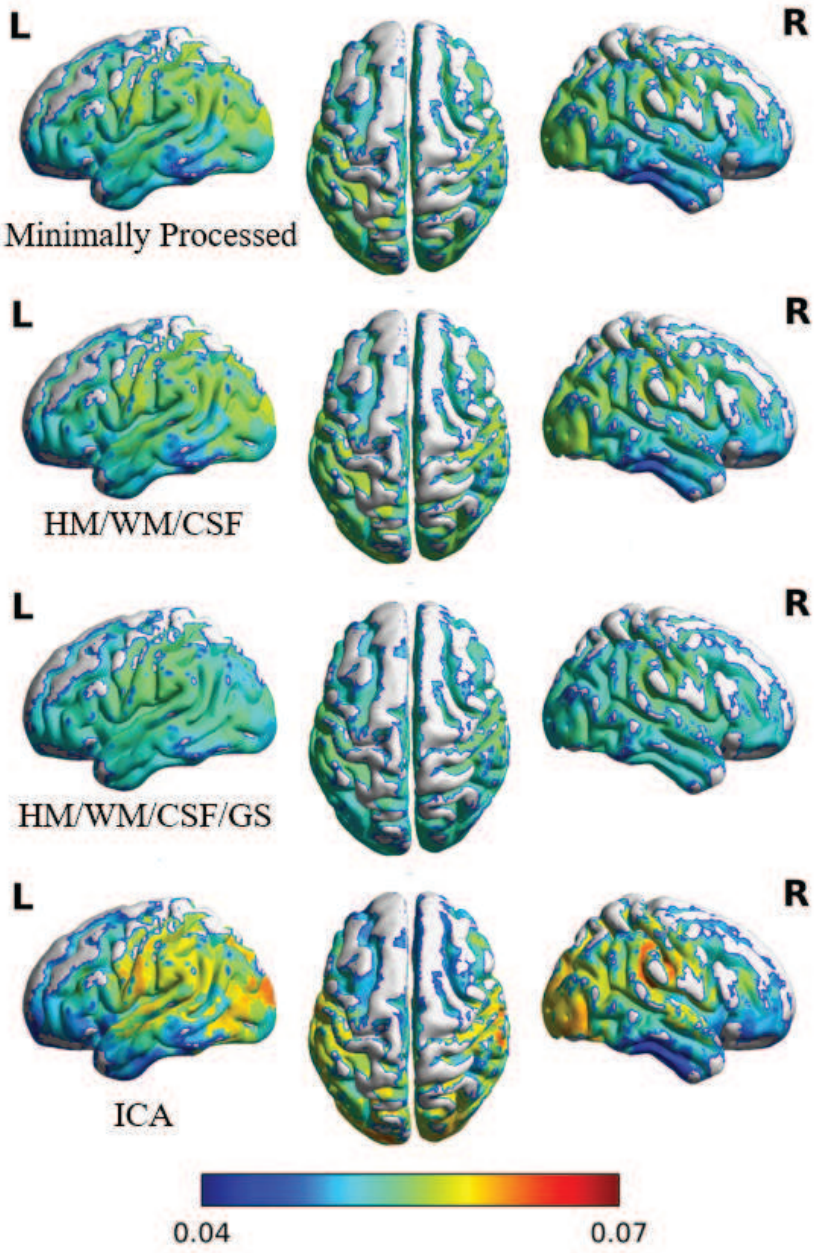

Figure 4.11. Mean eigenvector centrality values shown superimposed on brain ROIs.

6. **Clustering Coefficient.** Figure 4.12 shows the population mean and median clustering coefficient for each cleaning type. Each cleaning progression tends to reduce reported node clustering. Note that mean and median clustering coefficient results are almost identical.

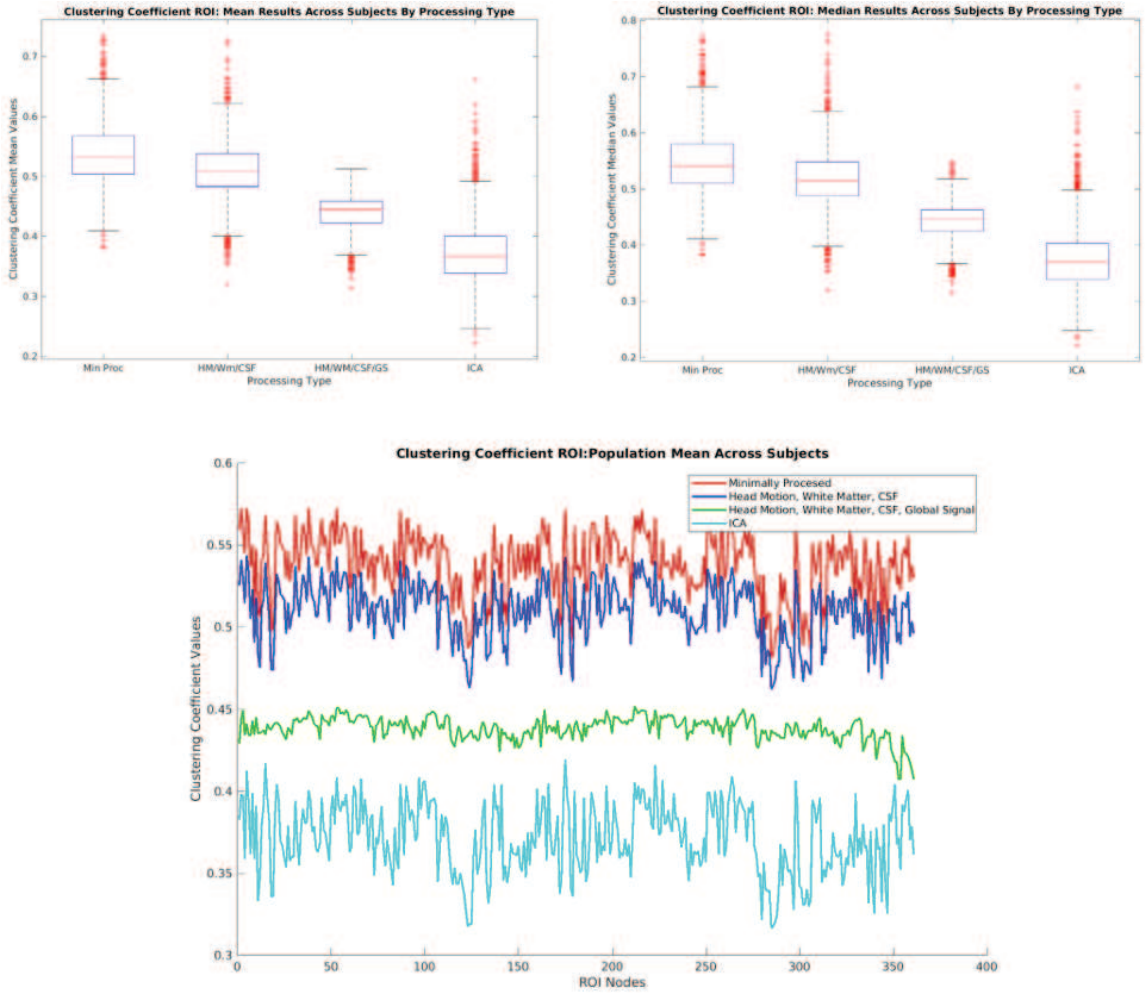

Figure 4.12. Median and mean clustering coefficient Values averaged across subjects.

Figure 4.13 shows the ICC results by ROI node across cleaning pipelines. Global signal processing reveals a sharp drop in ICC reproducibility, whereas again ICA processing returns the highest ICC scores.

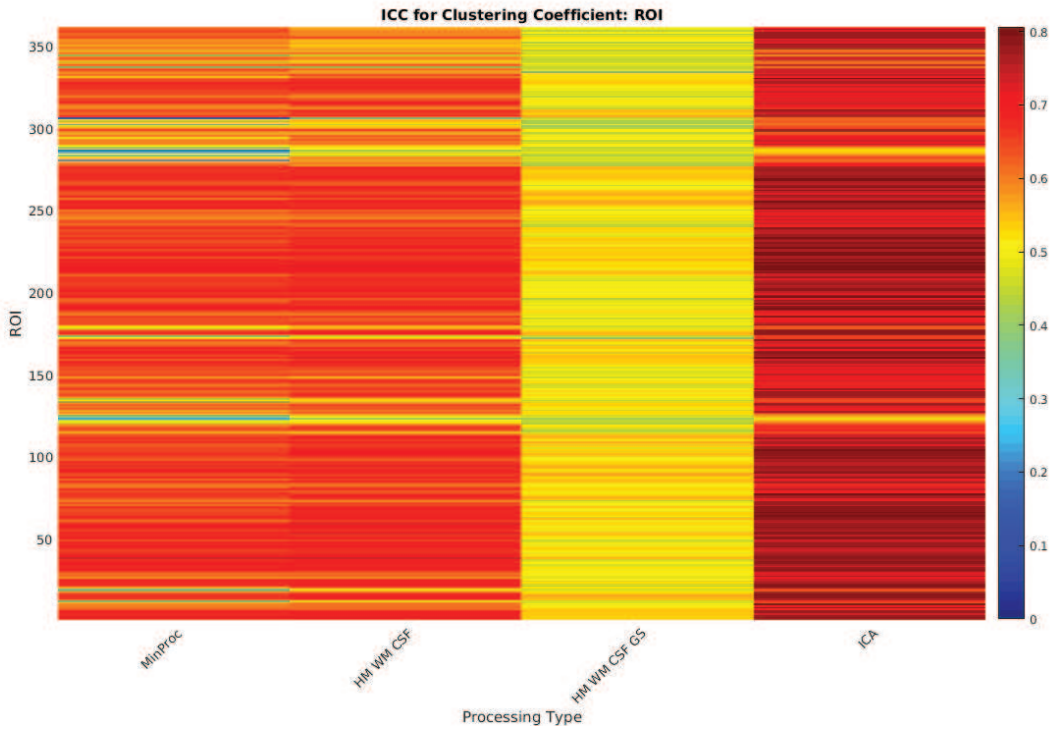

Figure 4.13. ICC values for clustering coefficient graph measures

Figure 4.14 shows population ROI mean Clustering Coefficient results mapped onto brain models. Each processing step removes reported ROI node clustering tendencies.

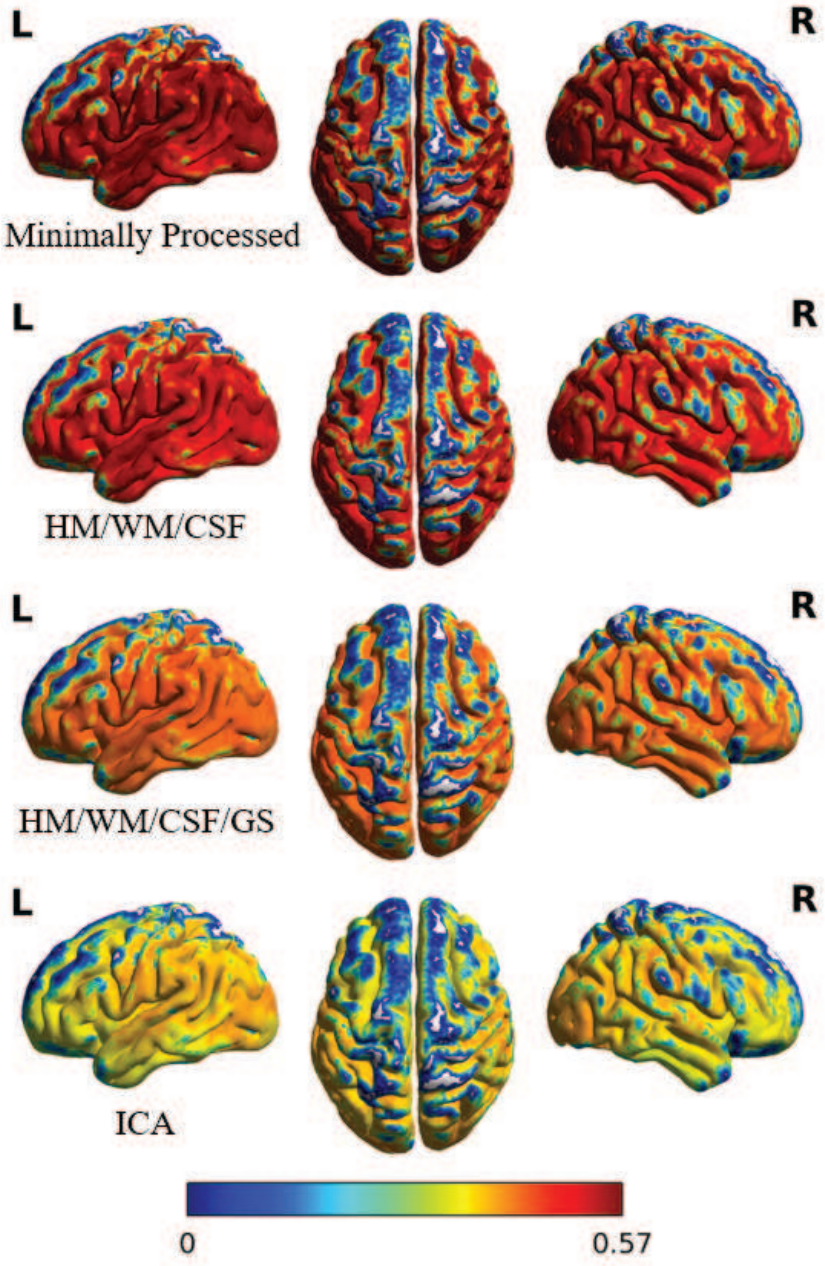

Figure 4.14. Mean clustering coefficient values shown superimposed on brain ROIs.

4.1.3 Effects of Postprocessing Strategy on Topological Analysis

4.1.3.1 ROI Barcode Creation

ROI barcodes represent distances at which ROI nodes merge into clusters as the connectivity threshold is relaxed from strongly to weakly connected. For data visualization purposes, it is often informative to reorder the barcodes from shortest clustering distance to longest, using the endpoints to create a smooth curve. Visualizing the reordered data provides a sense of how quickly nodes are merging into single clusters. Figure 4.15 demonstrates this process.

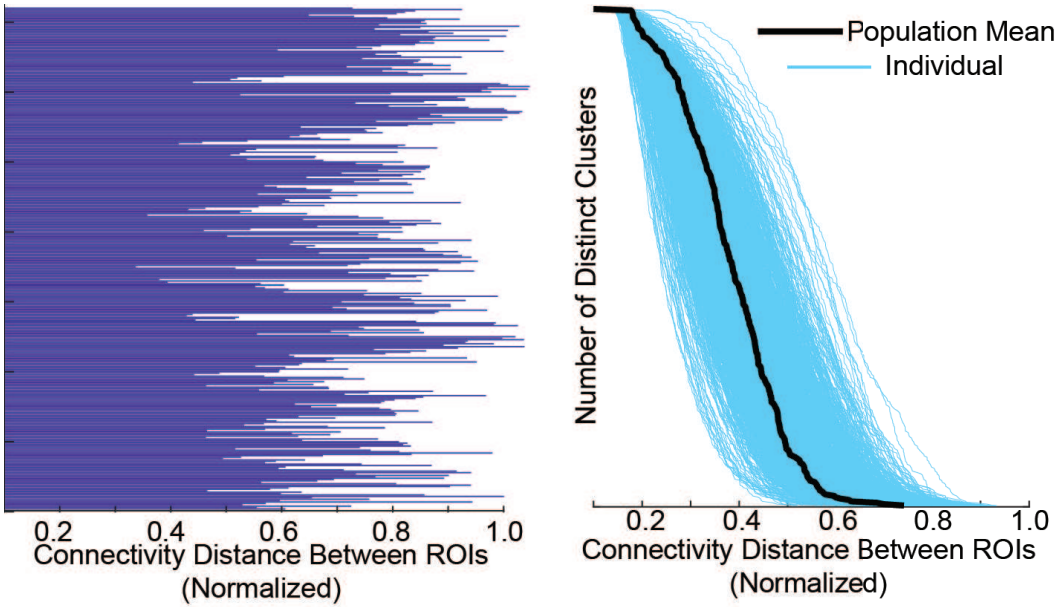

Figure 4.15. Reordering barcodes to create population sigmoid curves.

4.1.3.2 ROI Population Barcodes

Figure 4.16 shows all population barcodes by processing type, with the mean population barcode superimposed in black. This figure provides a sense of the spread of barcodes over the population for each cleaning type.

Figure 4.17 combines the population average barcodes by processing type and their accompanying ICC ROI values. With each progressive cleaning process, average node

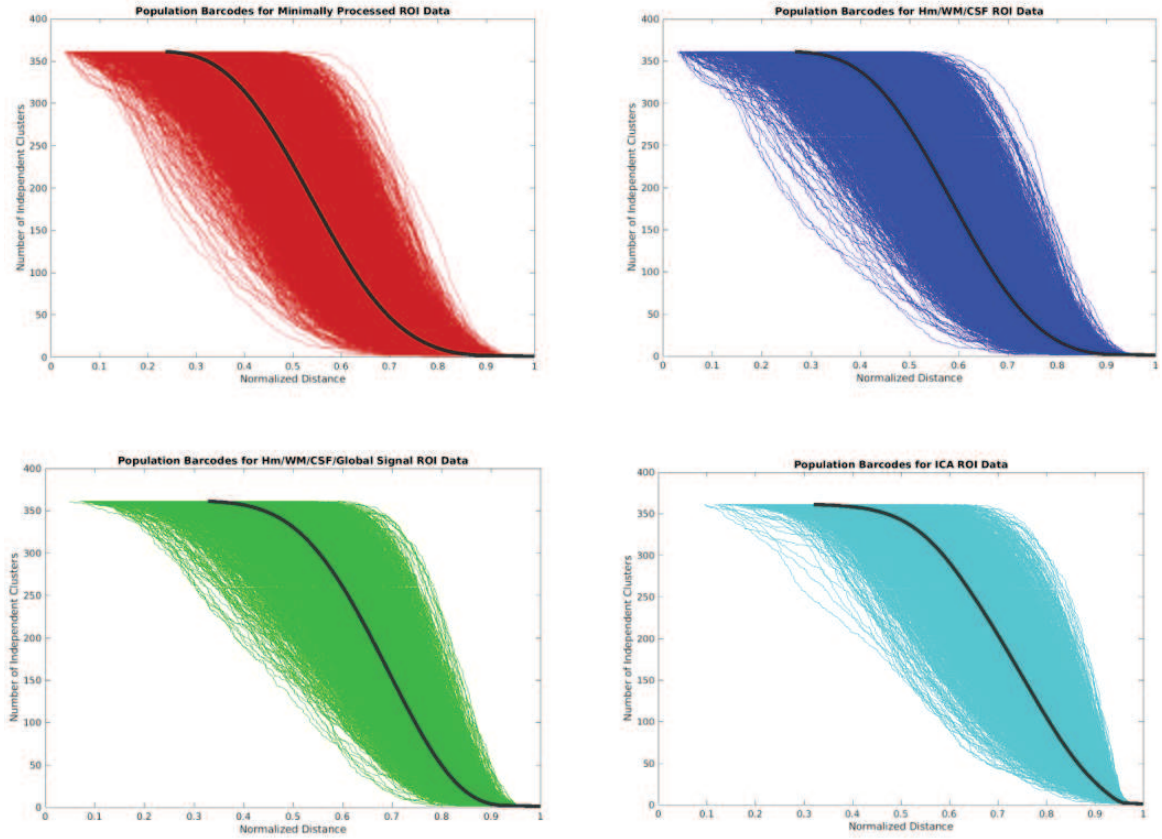

Figure 4.16. Population barcodes by processing type.

clustering is pushed into larger distances. ICC reproducibility presents strong for ICA cleaned data, whereas other processing pipelines are significantly weaker, with global signal processing showing a slight drop in ICC values from minimally processed data.

4.1.4 ROI Behavior Correlation

Figures 4.18 and 4.19 show behavior correlations captured by each processing type. Only correlations with p-values less than 0.05 are shown. Whereas all processing types show some correlation with artifacts such as head motion, ICA finds higher correlation in general across many more relevant behavioral measures. Global signal processing reduces the amount of overall behavioral correlation.

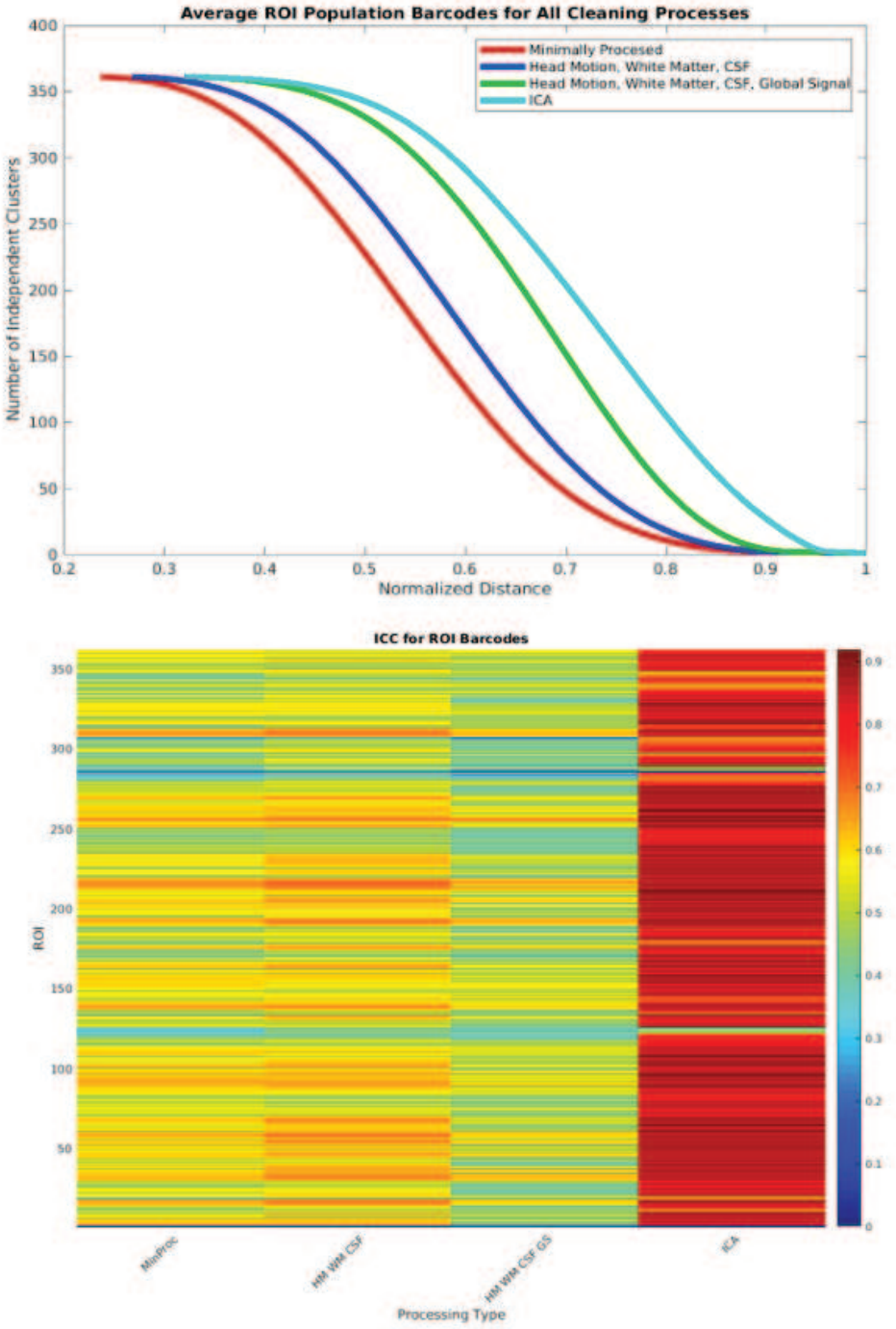

Figure 4.17. Average population barcodes and ROI ICC values by cleaning pipeline.

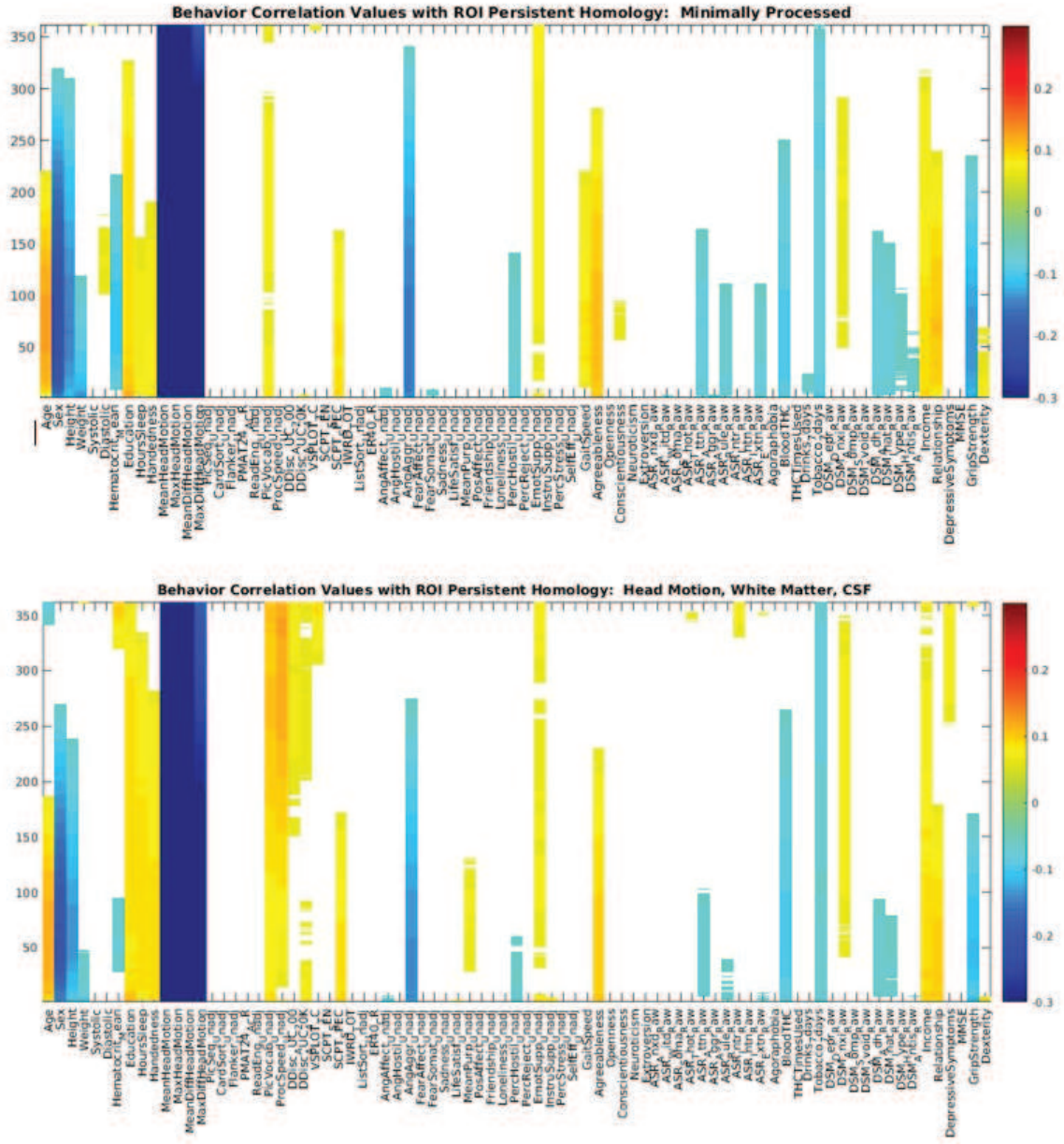

Figure 4.18. Correlation with behavior: minProc, Hm/Wm/Gs.

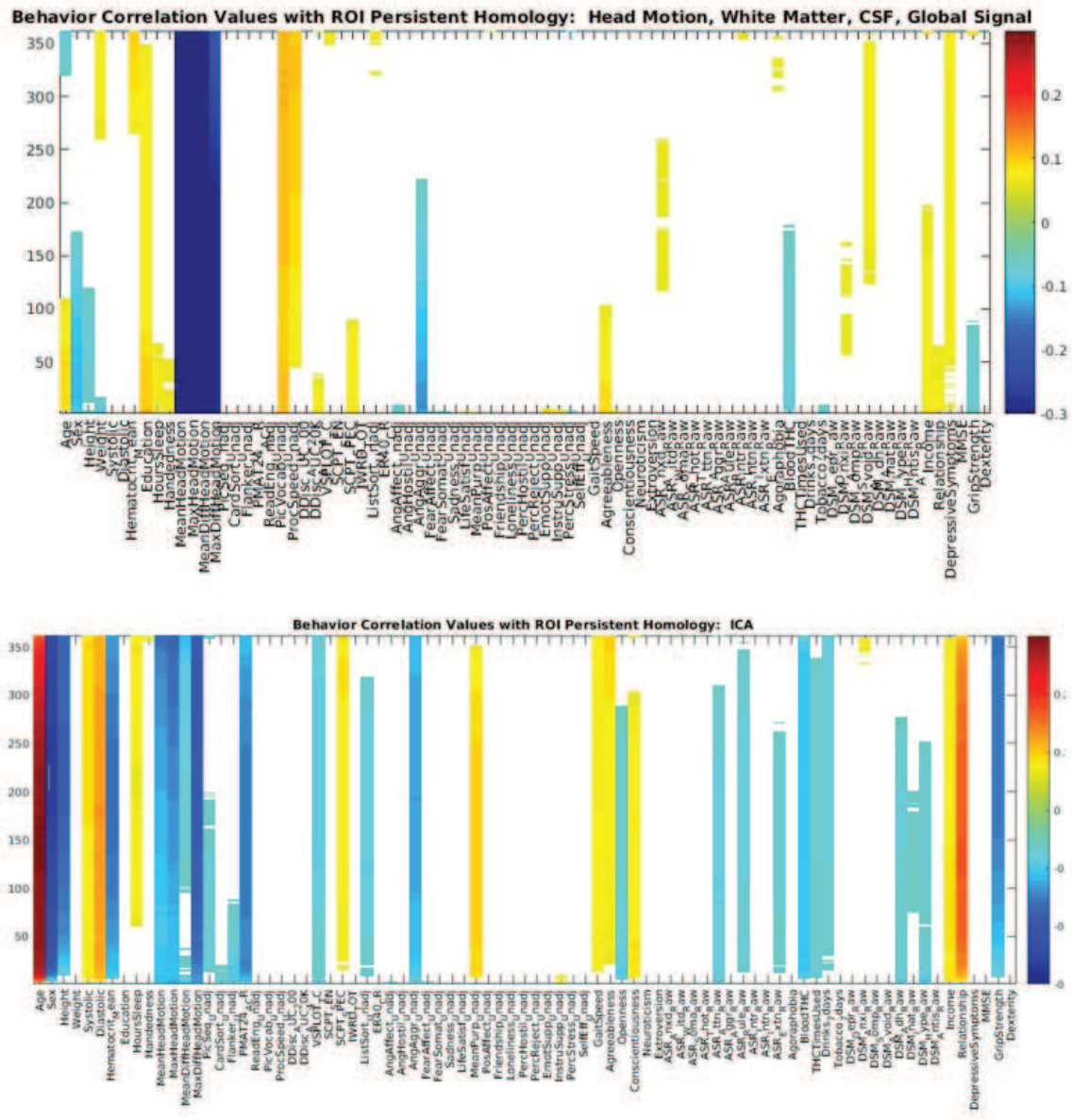

Figure 4.19. Correlation with behavior: Hm/Wm/Csf/Gs, ICA

4.2 Comparing Temporal and Spatial Connectivity Graphs

For all ROI and time domain correlation matrix comparisons, ICA cleaned data was used.

4.2.1 Functional Connectivity Results

Figure 4.20 gives a sense of connectivity in the time domain for a single patient. The bottom graph reorders the Time Domain nodes into eight clusters, each cluster containing time domain nodes that exhibit more similarity in the relative activity pattern across spatial ROIs to each other than to other clusters. Time domain nodes refer to arbitrary time points that are not meaningful to compare across subjects, as opposed to ROI nodes, which represent a physical location on the brain that is analagous across subjects.

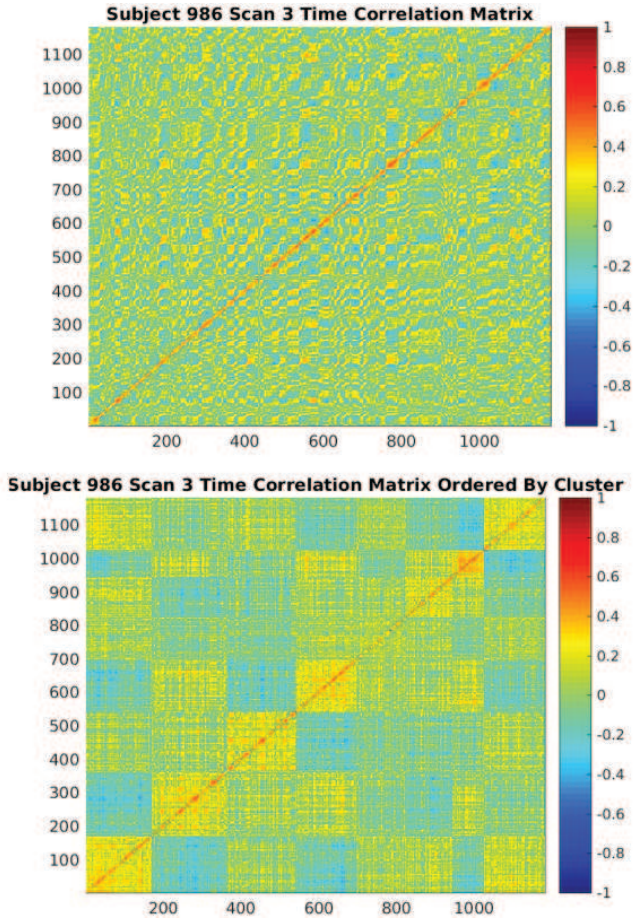

Figure 4.20. ICC values for ROI barcodes.

4.2.2 Graph Theoretical Measures

1. **Modularity.** Modularity results for ROI graphs and time domain graphs are compared in Figure 4.21. Means and variance are similar across both domains, with ICC values slightly higher for the time domain.

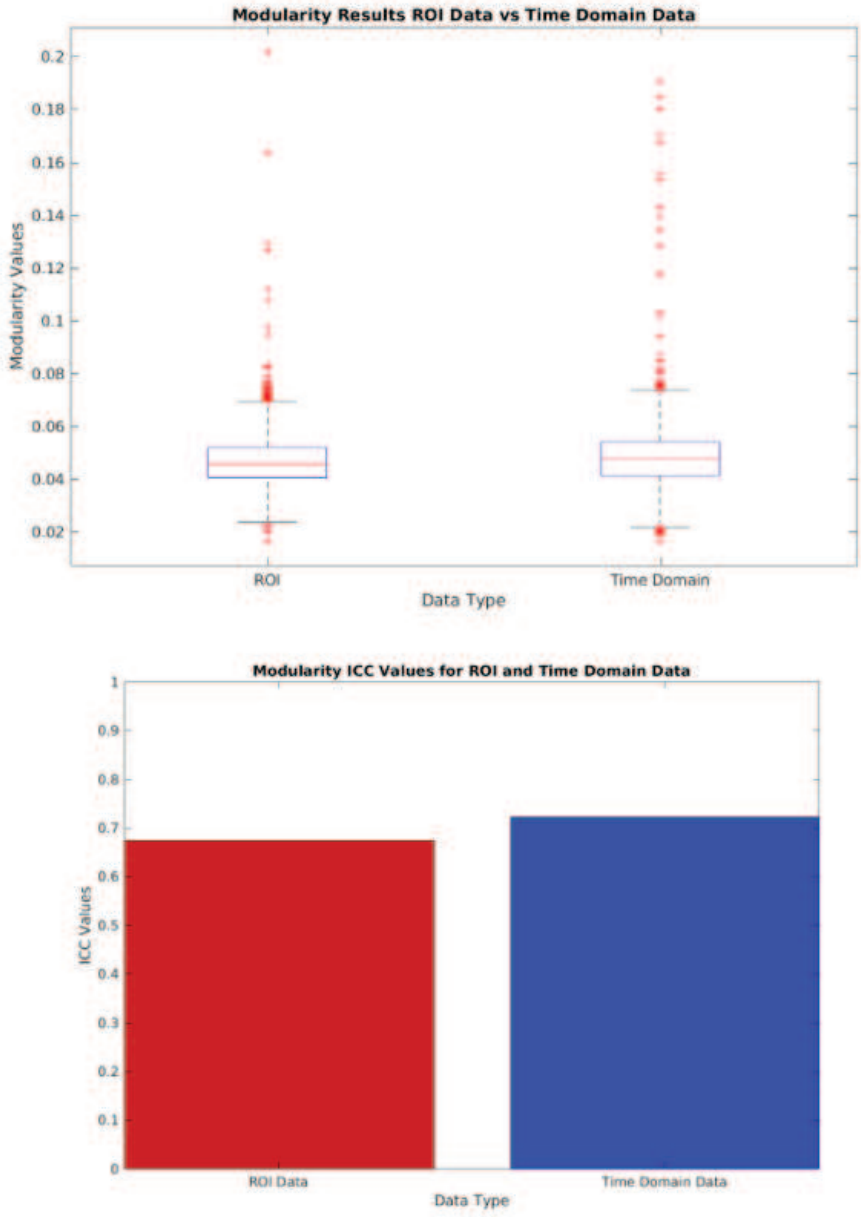

Figure 4.21. Modularity results for ROI and time domain correlation matrices.

Figure 4.22 shows the significant correlations found with behaviors and modularity results for both domains. Although both domains identify signicant correlation to behavior, each captures a different set of behaviors.

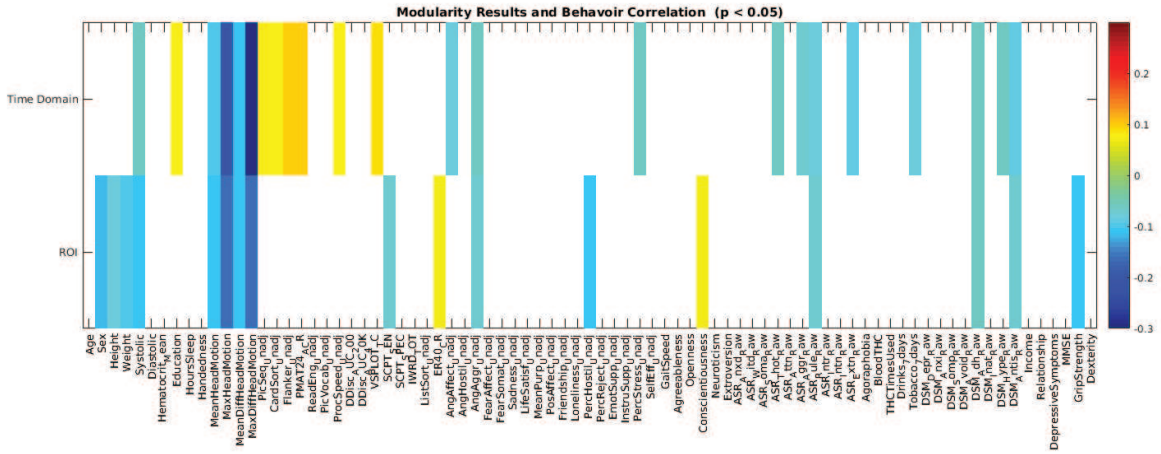

Figure 4.22. Modularity correlation results across ROI and time domains.

2. **Characteristic Path Length.** Characteristic path length results show similar values across both domains as well. However, ICC reproducibility is stronger for the physical ROI domain in this graph measure. Results are shown in Figure 4.23.

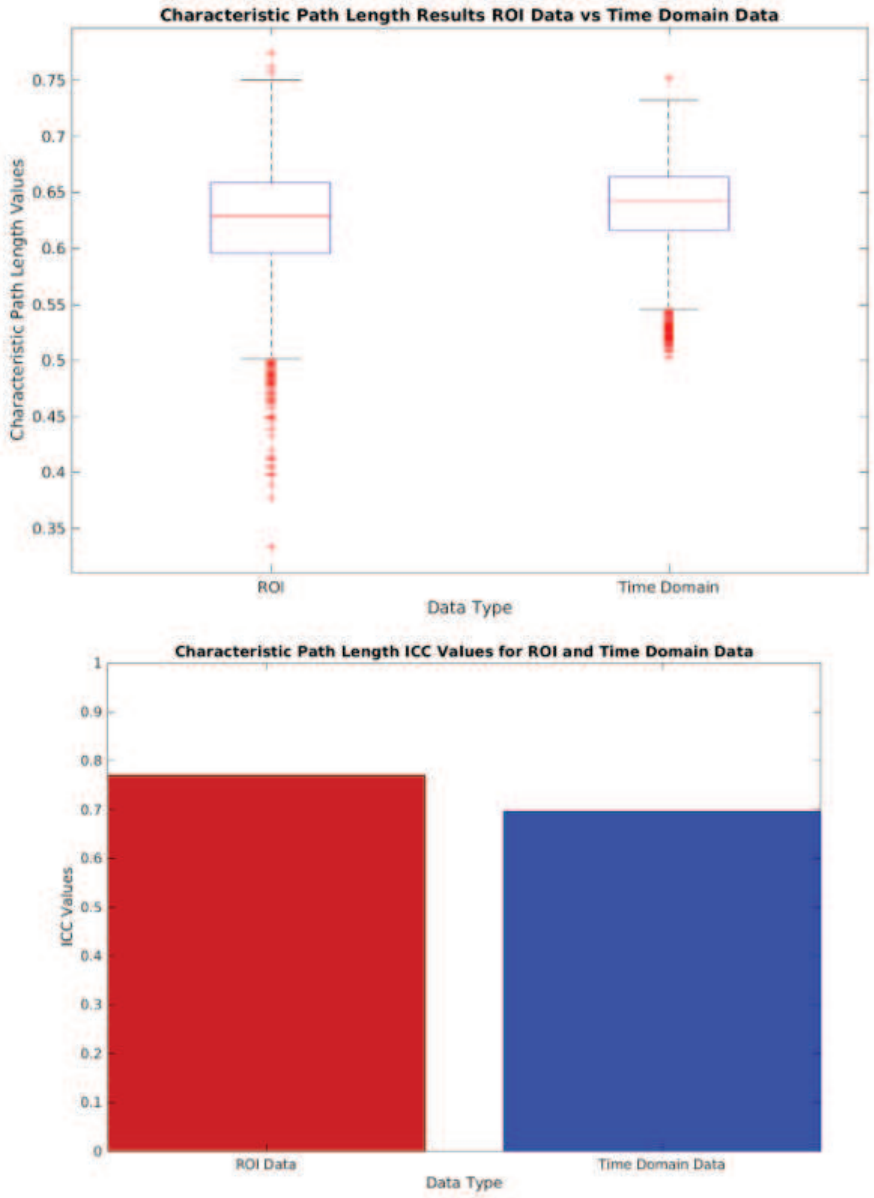

Figure 4.23. Characteristic path length results for ROI and time domain correlation matrices.

Figure 4.24 shows behavioral correlation with characteristic path length results. The time domain characteristic path length results correlate with a larger set of behaviors than do the ROI results.

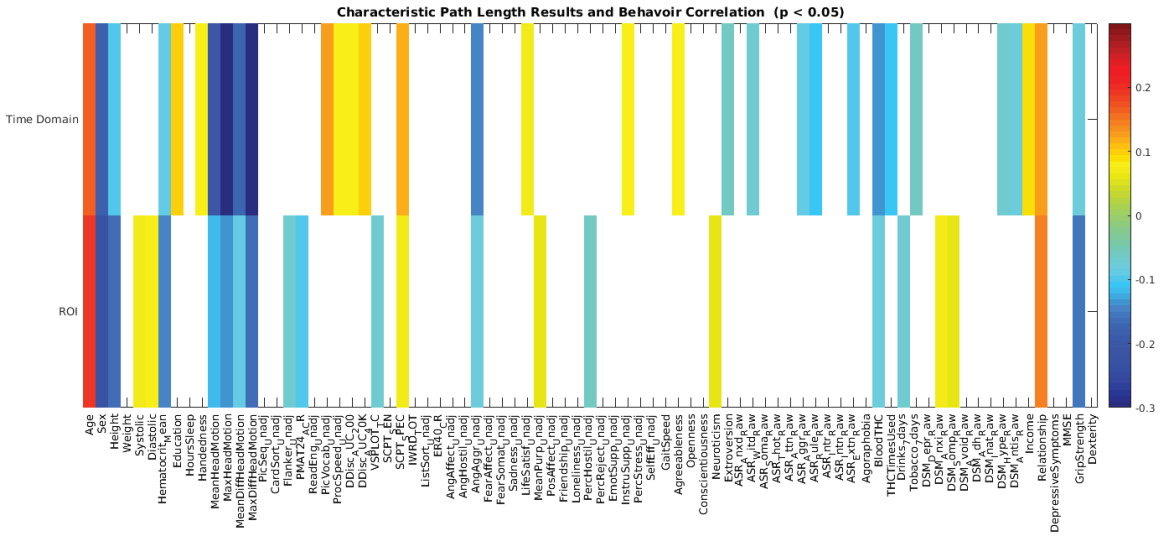

Figure 4.24. Characteristic path length correlation results across ROI and time domains.

3. **Global Efficiency.** Figure 4.25 contains ROI and time domain results for global efficiency. Mean values were roughly equivalent for both with only the variance changing. The time domain global efficiency ICC results were nearly perfect.

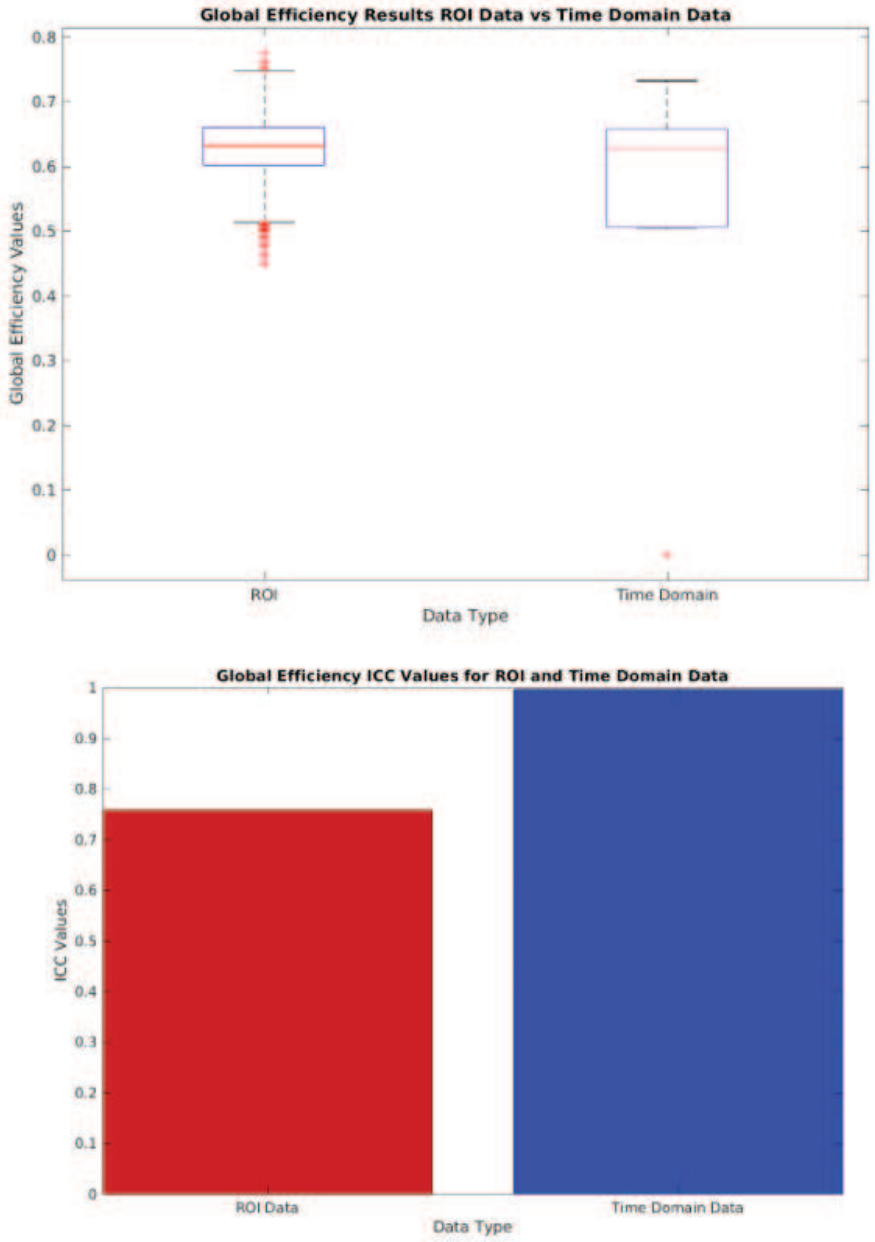

Figure 4.25. Global efficiency results for ROI and time domain correlation matrices.

Figure 4.26 contains Global Efficiency and behavior correlation results for ROI and time domains. Only the physical ROI domain shows numerous significant behavioral correlations.

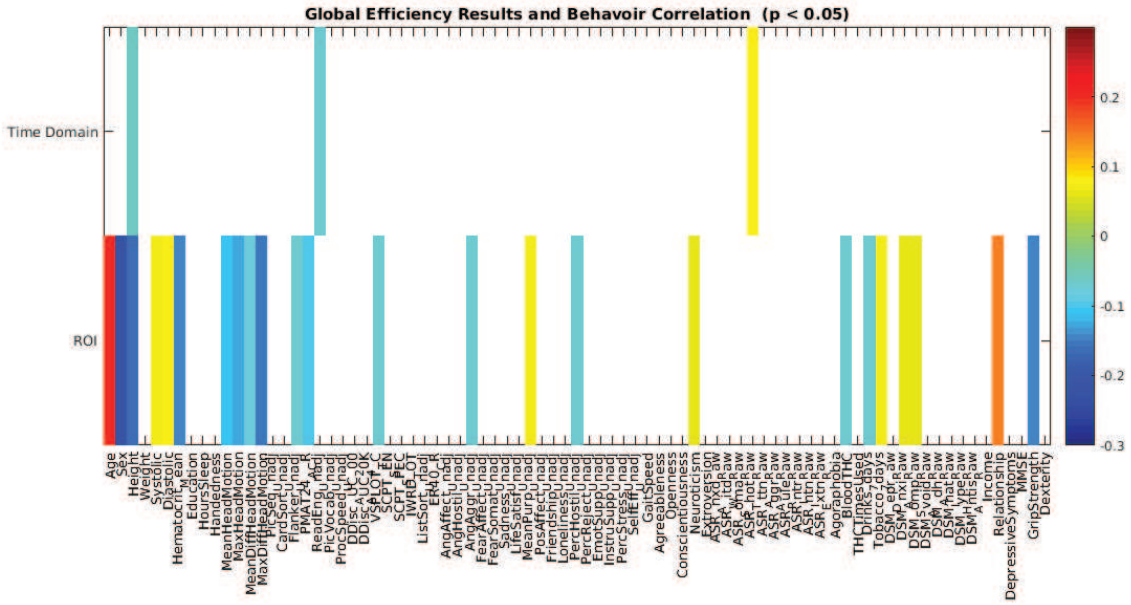

Figure 4.26. Global efficiency correlation results across ROI and time domains.

4. **Betweenness Centrality.** Betweenness centrality also produced sparse vectors for the time domain. ICC values for the time domain did not score high enough to register, and all mean and median values hover around 0. Figure 4.27 shows these results. Figure 4.28 contains behavioral correlation results.

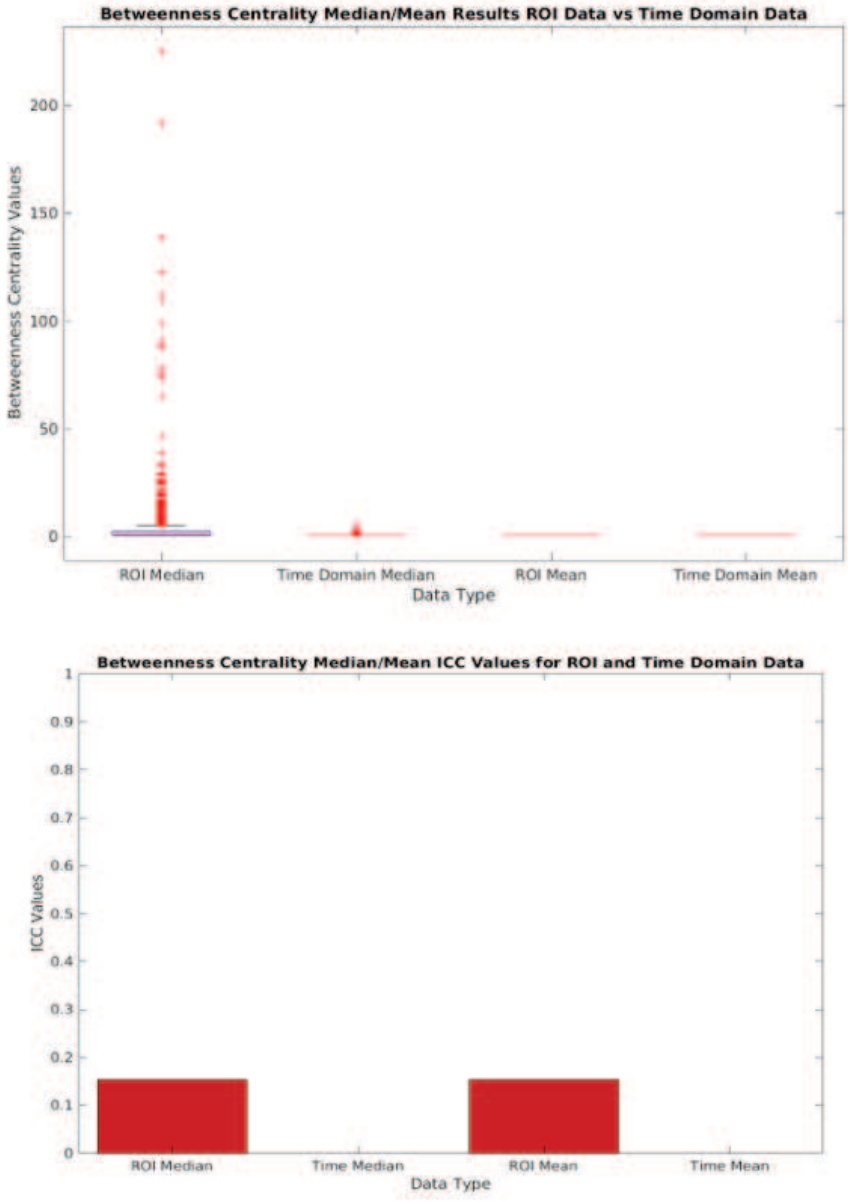

Figure 4.27. Betweenness centrality results for ROI and time domain correlation matrices.

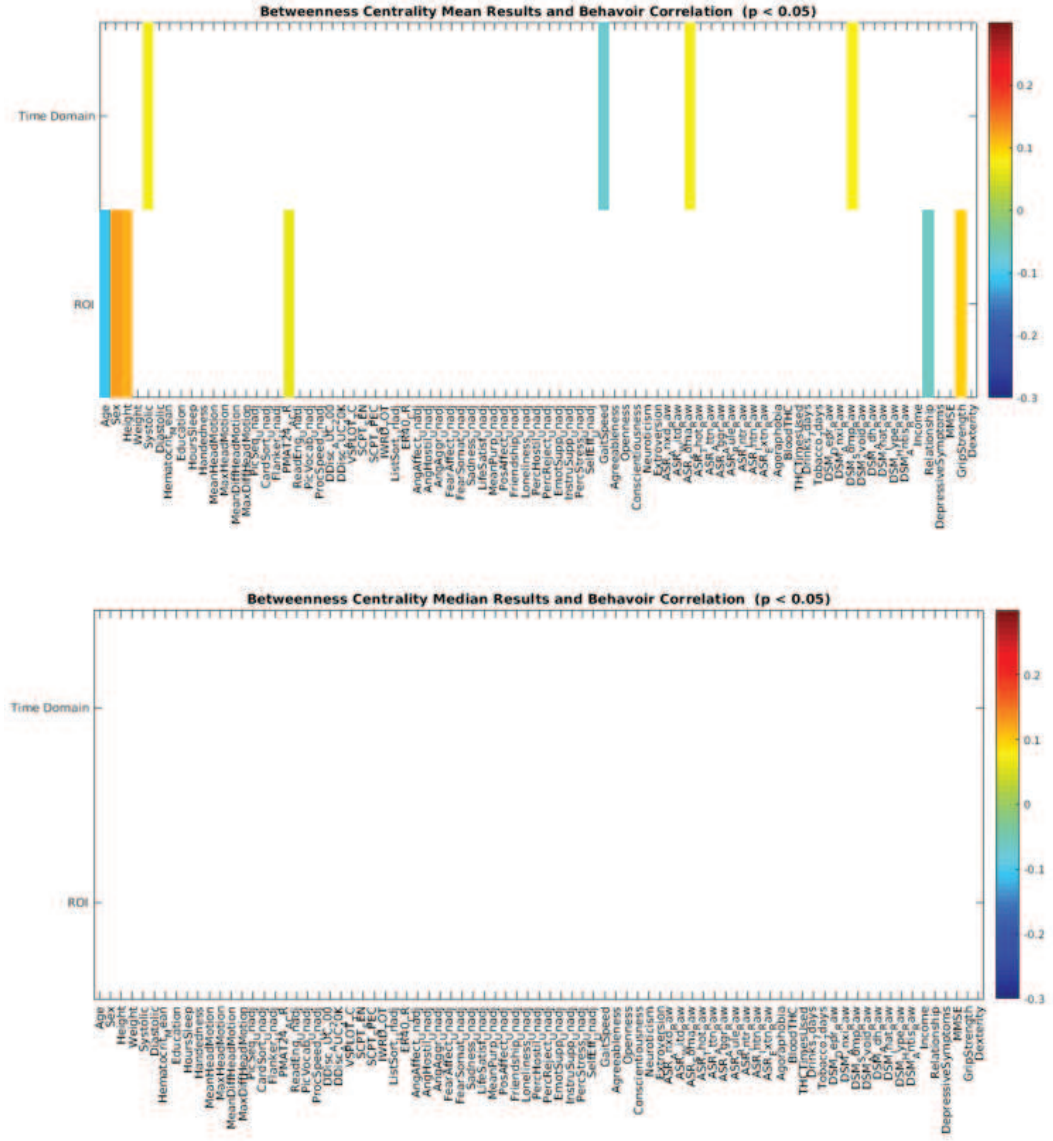

Figure 4.28. Betweenness centrality mean and median correlation across ROI and time domains.

5. **Eigenvector Centrality.** Eigenvector centrality results for ROI and time domains are contained in Figure 4.29. The ROI domain notably provides stronger results, both in eigenvector centrality and ICC scores. Figure 4.30 shows behavior correlation with eigenvector centrality, with the ROI domain showing more correlation.

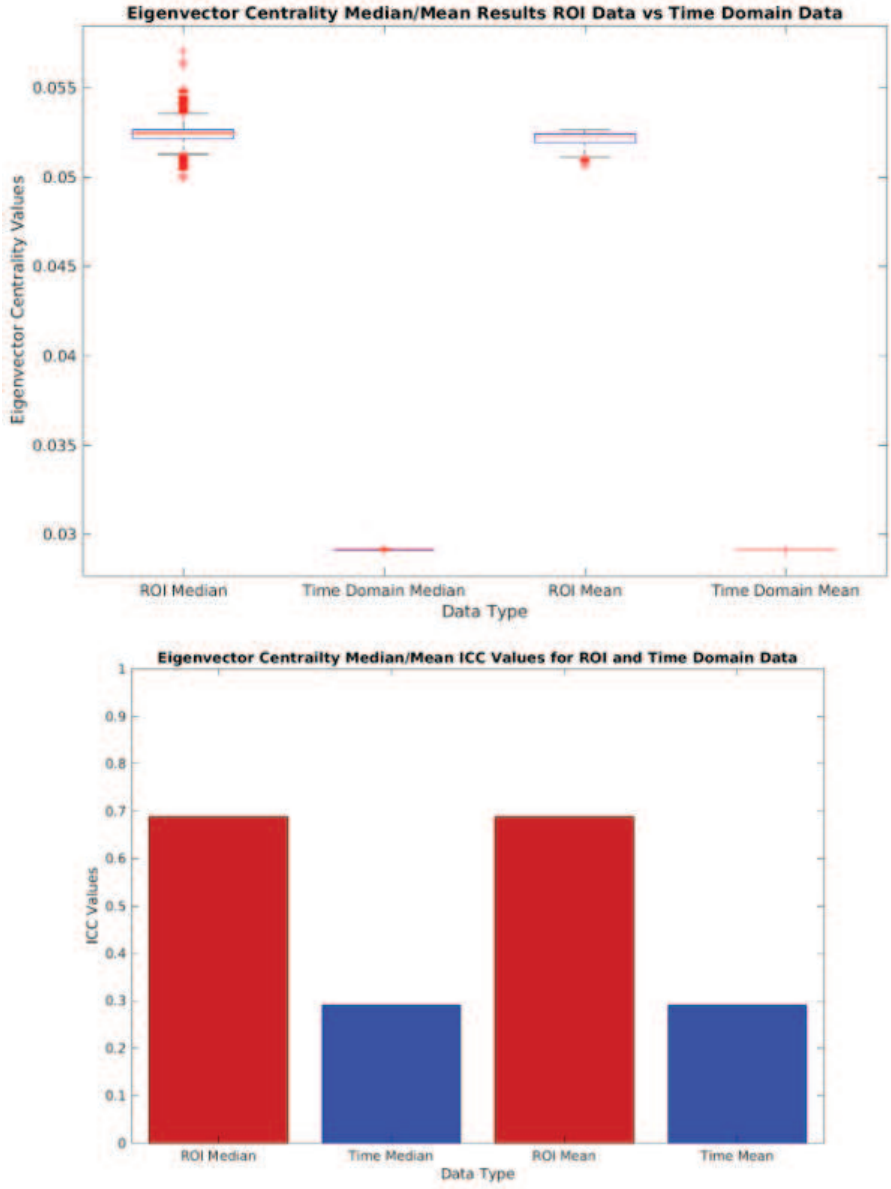

Figure 4.29. Eigenvector centrality results for ROI and time domain correlation matrices.

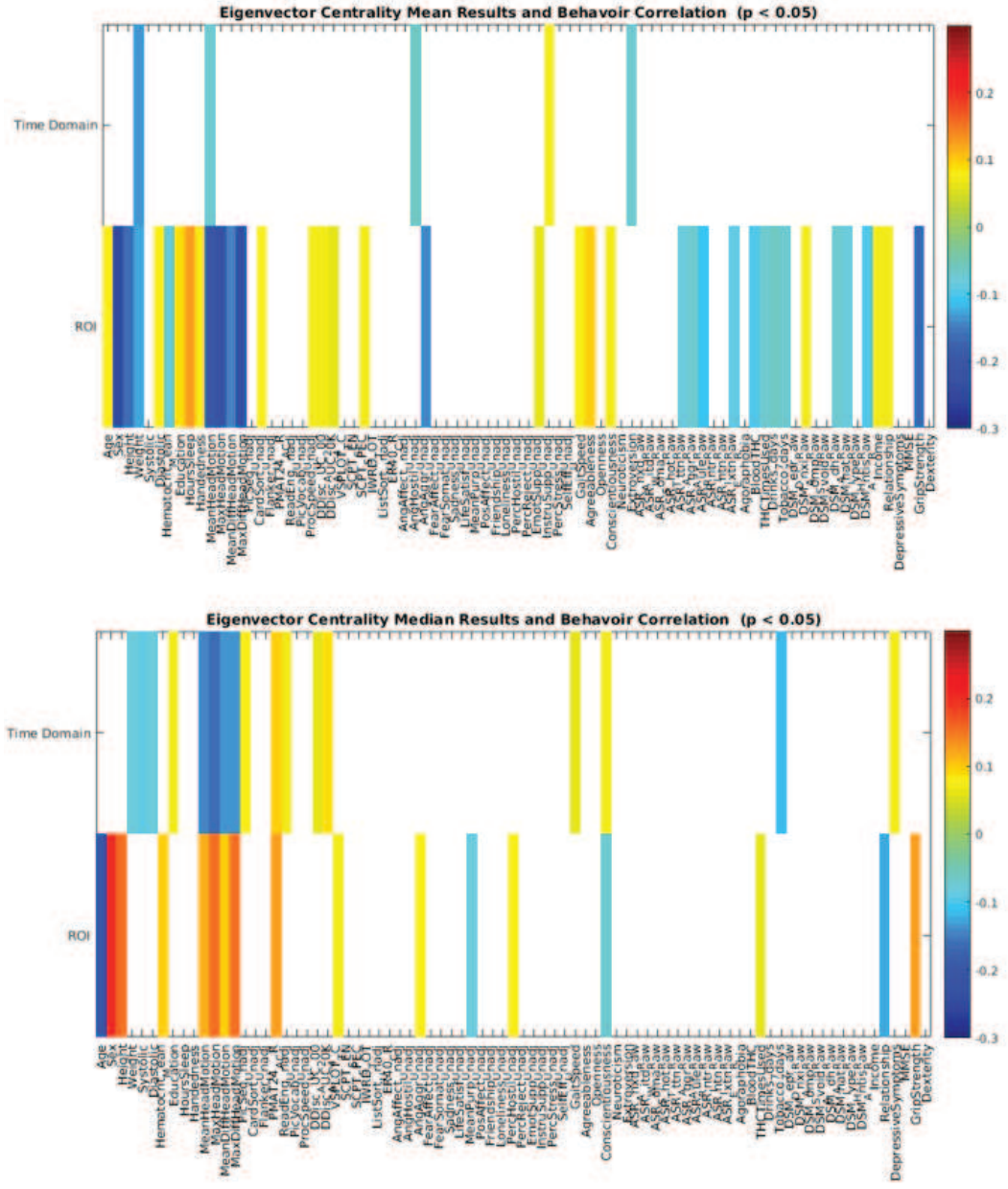

Figure 4.30. Eigenvector centrality mean and median correlation across ROI and time domains.

6. **Clustering Coefficient.** ROI and time domain clustering coefficient results are shown in Figure 4.31. Although mean and median results remain steady across both domains (with ROI ICC values registering slightly higher), the time domain contains stronger clustering coefficient correlation with behavior, shown in Figure 4.32. ROI and time domain each capture a different set of behaviors with this metric.

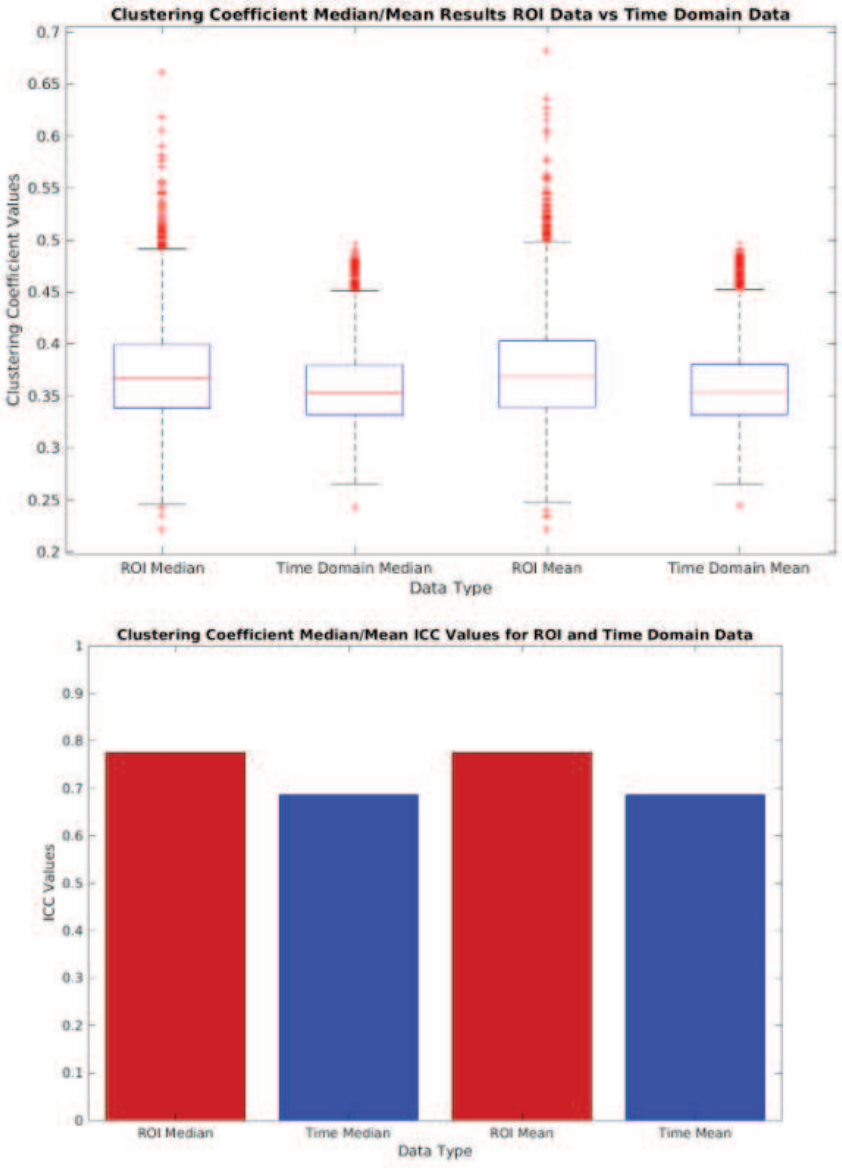

Figure 4.31. Clustering coefficient results for ROI and time domain correlation matrices.

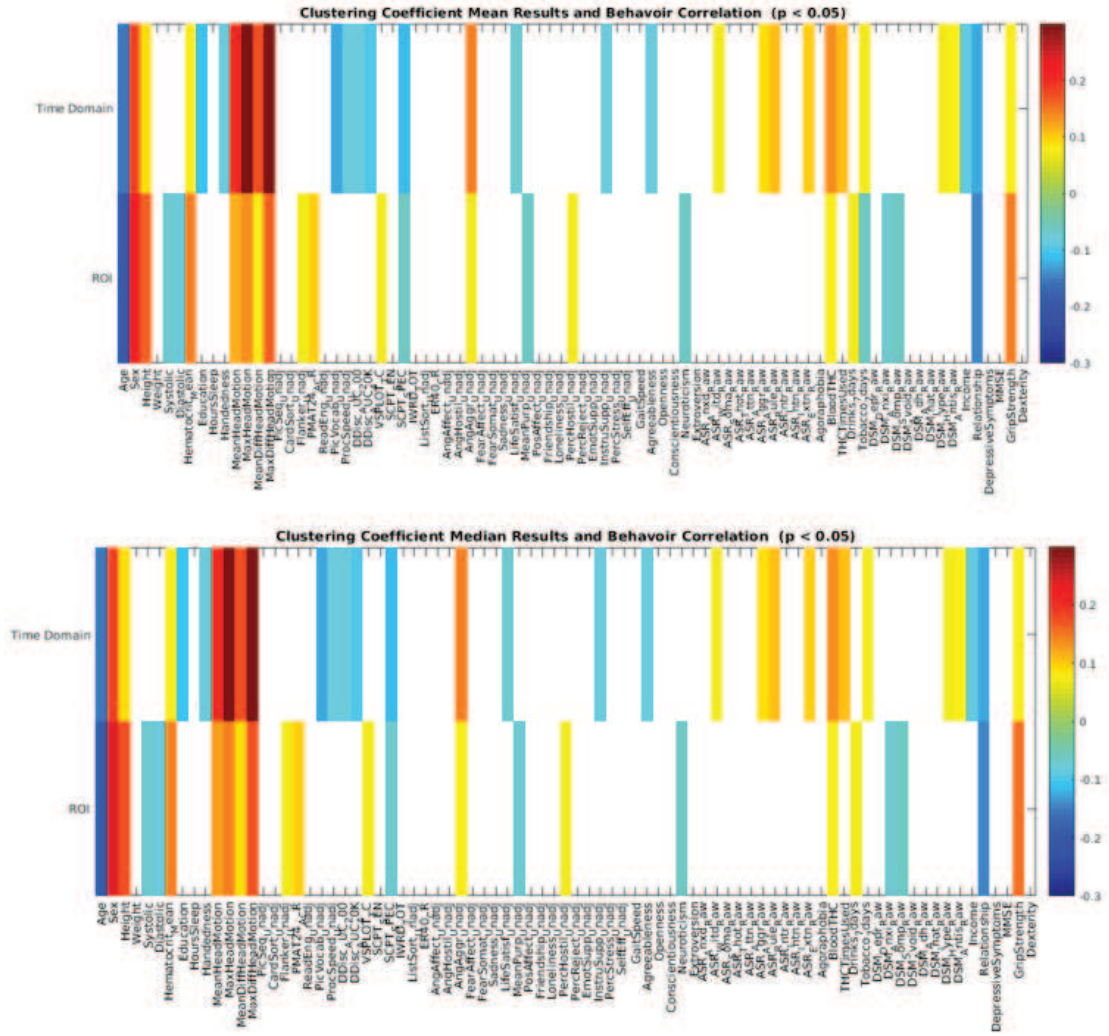

Figure 4.32. Clustering coefficient mean and median correlation across ROI and time domains.

4.2.3 Topological Analysis

4.2.3.1 ROI and Time Domain Barcodes

Figure 4.33 shows the population barcodes for scan 3 ICA data, first created for ROI followed by barcodes created for the time domain. Nodes in the time domain require initially larger distances to begin merging into clusters, but the curve drops sharply once significant merging begins.

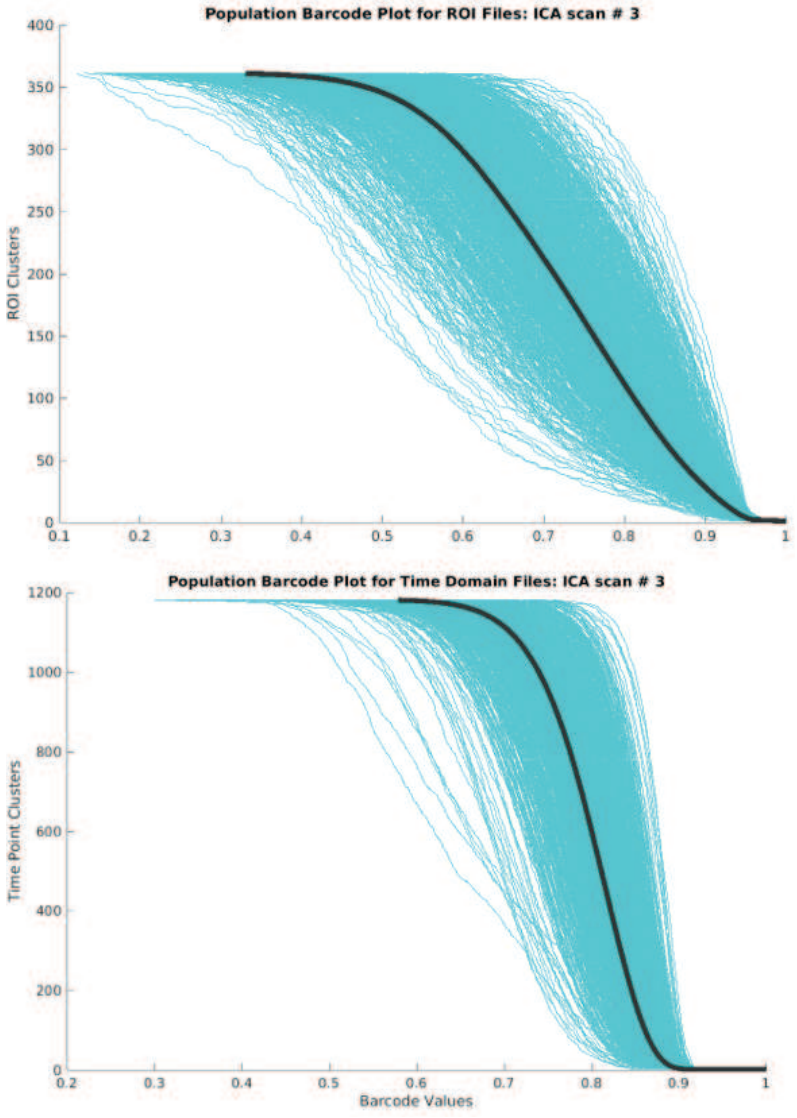

Figure 4.33. Comparing ROI and time domain ICA barcodes.

4.2.3.2 ICC for Time Domain.

Figure 4.34 displays ICC reproducibility for time domain barcodes. Barcode values were first sorted from shortest merging distances to longest, because there is no spatial counterpart to physically placed ROI nodes. Regardless, sorted barcodes across the four scans for a subject resembled each other more than they did results from other patient scans. Because the ICC results were unexpectedly high, the available different processing pipeline data was also used for comparison.

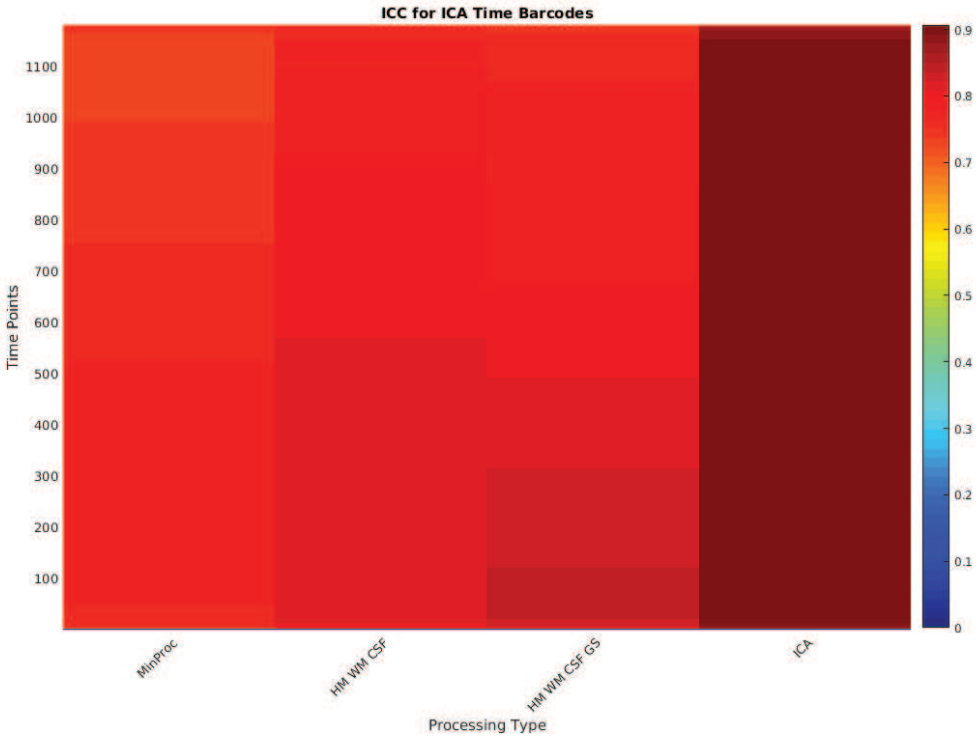

Figure 4.34. ICC values for time domain barcodes.

4.2.3.3 Time Domain Behavior Correlation.

While ROI Behavior and barcode correlation picked up a few more correlated behaviors than did the time domain, what is remarkable is how similar the results present, both in

capturing the same behaviors as well as in the scores for the behavior. Figure 4.35 shows behavior correlation for both ROI and time domains.

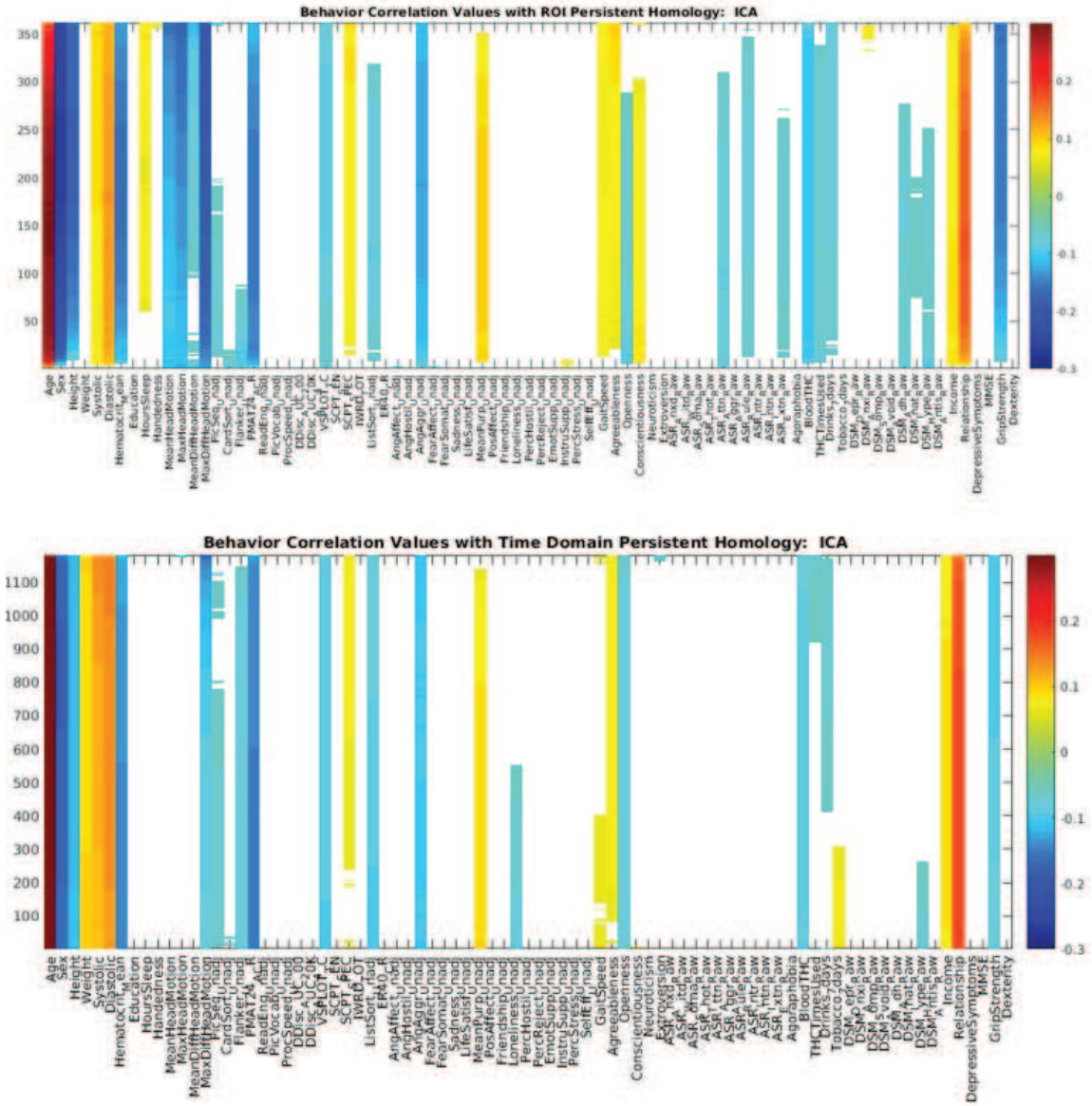

Figure 4.35. Correlation with behavior: ROI vs time domain.

CHAPTER 5

CONCLUSIONS

5.1 Postprocessing Pipeline Findings

5.1.1 Summary of Core Postprocessing Findings

The postprocessing strategy used will affect both graph and topological analytical results, and therefore deserves consideration. In general, more aggressive cleaning strategies will remove artificial connectivity between nodes created by shared noise. Removing global signal (the mean time series of the whole brain), however, goes a step too far and removes or dilutes important signal and reliability. The results consistently showed that ICA cleaned data had the highest ICC scores, significantly removed misleading results attributed more to noise than to signal, and presented the strongest correlation results with behavior.

5.1.2 Graph Measure Results by Pipeline

For all graph theoretic results, the aggressively cleaned ICA data suggested a less connected graph than did the other cleaning processes. Graph measures obtained from minimally cleaned data suggested higher connectivity between nodes, shorter path lengths, and higher clustering between nodes, presumably from noise causing the nodes to produce artificially high levels of connectivity. ICA, however, reduced the noise captured in the signals and produced graph theoretic results more likely to be those relating to a genuine brain signal rather than noise. In the case of eigenvector Centrality, mean values went down for ICA data but variance increased, allowing detection of more specific influential nodes in the network. For global efficiency, ICA data reported more efficient networks than did the more noisy data collected from less aggressive processing.

5.1.3 ICC Results by Pipeline

The most prominent result from graph theoretic measures was the consistent drop in ICC reproducibility scores for removing global signal, and the simultaneous rise in ICC reproducibility scores for ICA processed data. With four scans per subject, ICC reproducibility evaluation could detect results that supported more similarity between scans of the subject as opposed to similarity with scans in the population at large. Of particular note is the result from eigenvector centrality. Although it might initially appear questionable that ICA data could produce Eigenvector Centrality results not captured in other data, the strong ICC values found only in ICA data substantiate the finding and suggest it is a reliable result. Global signal data showed a significant drop in ICC performance for the clustering coefficient compared to other data as well.

5.1.4 Topological Results by Pipeline

The topological barcodes produced by cleaning type showed a gradual progression from more quickly merging clusters to slower merging clusters as the cleaning method became more aggressive. This is consistent with the premise that noise will artificially create connectivity where there is none, consequently merging nodes together faster than clean signal will. The ICC results for the barcodes was remarkable. Not only did ICA data produce persistently high ICC results for the ROI nodes, but global signal data once again showed a drop in ICC outcomes, lower even than minimally processed data. Topological analysis performed on ICA data will produce accurate results with a higher degree of confidence than the other cleaning processes will.

5.1.5 Behavior Results by Pipeline

All cleaning pipelines picked up similar sets of behavior correlations with barcodes. There was a slight drop in behavior correlation with Global Signal data as compared to the other processes, and a slight rise in correlation for ICA data. Of interest is the strong correlation the first three pipelines produced with artifact, especially head motion. ICA data, however, picked up correlation with head motion, but to a lesser degree. ICA data is the only set that correlated strongly with age, suggesting that slower merging clusters are an indication of an older person. ICA data barcodes also correlated with high level behaviors such as whether or not the subject was in a relationship, income levels, how

open to experience the individual is, and I.Q.

5.1.6 Other Pipelines To Consider

Although ICA data performed exceptionally well for this study, it is not the only approach to data cleaning that produces reliable results. Another method often used is that of scrubbing: the technique of removing an entire volume (time point) that is particularly corrupted by high levels of head motion and concatenating the study without this volume [30]. The scrubbing technique has shown to produce similarly high ICC reproducibility scores to those of ICA data, even over finer grained spatial regions [32], when combined with regression of head motion, white matter, and CSF signals (not global signal).

5.1.7 Limitations

One limitation is the lack of ground truth observations to compare results to. The tools available allow us to infer reliability only based on reproducibility and behavioral correlations, but the results remain uncertain. Reproducibility and behavioral correlation results could be further evaluated with simulated brain connectivity data with added noise where ground truth is known. Many other graph theoretical and topological methods that could be added to verify results obtained from different cleaning pipelines. fMRI data that is more sparse may perform differently than the Human Connectome data.

5.2 Time vs Spatial Domain Findings

5.2.1 Summary of Core Time vs Spatial Findings

Graph theoretical measures gave mixed results when comparing Time Domain and Spatial (ROI) domain data. The time domain performed reasonably well in most cases in terms of ICC reproducibility, but in some cases the Spatial domain outperformed the time domain. Most often the set of correlated behaviors was different with graph results, suggesting different information is to be found in the flipped graph structure. In the topological analysis, however, both the time domain and spatial domain correlated with nearly identical behavior sets, with ICC reproducibility scores registering remarkably high. One advantage the time domain barcodes has is that there appears to be less noise from head motion reflected in the results.

5.2.2 Graph Measure Differences Time vs Space

The specific graph analyses used for this study produced mixed results in terms of stronger performance for either spatial ROI data or time domain data. For modularity, global efficiency, and characteristic path length, both the ROI and time domains showed similar population mean values, but the time domain showed stronger support for these findings with higher ICC reproducibility scores (especially strong for global efficiency, slightly lower for characteristic path length). Although the time domain produced stronger and more correlation with behavior for modularity and characteristic path length, surprisingly the time domain global efficiency results produced almost no correlation with behavior. Betweenness centrality was not a metric that produced interesting results for either ROI or time domains: by that measurement, the respective graphs have few strong hubs. The hub measurement of eigenvector centrality did have some notable results for the ROI domain, but the time domain contained no reliable eigenvector centrality information. Finally, for the clustering coefficient, mean and median values were consistent across both domains, with ICC reproducibility only slightly higher for ROI connectivity matrices.

5.2.3 Topology Differences Time vs Space

Side-by-side plots of population barcodes show that spatial ROI barcodes begin to coalesce faster and within shorter distances. However, once Time Domain barcodes do start to merge, the merging increases at a faster rate, resulting in a single cluster before the ROI barcodes do. The ICC reproducibility values for the time domain were conspicuously high, and higher than those for ROI barcodes. Time domain barcode information appears to be reliably reproducible for subjects and can readily distinguish one subject from another, analogous to a brain fingerprint.

Perhaps one of the most compelling findings is that were not many differences between the time domain and spatial domain in barcode and behavior correlation. The time domain barcodes retain information found in ROI barcodes and both produce similar sets of behavior correlation. As might be expected, the time domain barcodes avoided significant correlation with head motion, present in the ROI results. Because the time domain looks across values of all nodes at a given time rather than the value of a single node across time, the interfering noise introduced by head motion is lost, which might be considered

an advantage for the time domain.

For both domains, the topological data appears to discriminate between two brain phenotypes, as illustrated in Figure 5.1.

Faster Convergence	Slower Convergence
More Common in Males	More Common in Females
Younger Participants	Older Participants
Higher Fluid Intelligence	Higher Meaning/Life Purpose
Increased Physical Strength	Improved Attention/Inhibitory Control
Improved Visuospatial Ability	Higher Income
Improved Working Memory	More likely to have Long-term Relationship
Higher Openness to Experience	Higher Agreeableness
More Rulebreaking Behaviors	Higher Conscientiousness
More likely to Use Alcohol/THC	Faster Walking Speed

Figure 5.1. Behaviors Associated with Faster and Slower Barcode Convergence

Faster barcode convergence is associated with higher fluid intelligence, working memory, and openness to experience, while slower convergence is associated with improved attention and inhibitory control, higher agreeableness and conscientiousness, improved income, personal relationships, and life purpose. One possible explanation is that relatively slower vs faster convergence may be related to brain inhibition, with faster convergence in participants with overall weaker inhibitory connections in the brain.

5.2.4 Future Directions for Analysis using Time-Based Correlation Matrices

Because the time domain retained important structural information with topological analysis, there are more options to explore here. For this study, only dimension 0 barcodes were used, measuring the rate of node clustering. Dimension 1 barcodes could also be calculated to sense more complex structures in the connectivity matrix and also correlated with behavior. Other graph theoretical results could be performed to see if different metrics will capture a stronger correlation with behavior. Because the spatial domain maps to physical locations in the brain, there is an opportunity for additional investigation of the regional effects of topology measures across the brain. For the time domain, there is future opportunity to explore optimized clustering of temporal nodes, what spatial

patterns across the brain these may represent, and how consistently these patterns are seen across subjects.

REFERENCES

[1] E. Allen et al., *Tracking whole-brain connectivity dynamics in the resting state*, Cerebral Cortex, 24(3) (2014), pp. 663–676.

[2] J. Anderson et al., *Network anticorrelations, global regression, and phase-shifted soft tissue correction*, Human Brain Mapping, 32(6) (2011), pp. 919–934.

[3] R. Birn et al., *Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI*, Neuroimage, 31(4) (2006), pp. 1536–1548.

[4] R. Boubela et al., *Beyond noise: Using temporal ICA to extract meaningful information from high-frequency fMRI signal fluctuations during rest*, Frontiers in Human Neuroscience, 7 (2013), p. 168.

[5] R. Buckner et al., *The organization of the human cerebellum estimated by intrinsic functional connectivity*, J Neurophysiology, 106(5) (2011), pp. 2322–2345.

[6] E. Bullmore and O. Sporns, *Complex brain networks: graph theoretical analysis of structural and functional systems*, Nature Reviews Neuroscience, 10(3) (2009), pp. 186–198.

[7] C. Caballero-Gaudes and R. Reynolds, *Methods for cleaning the BOLD fMRI signal*, Neuroimage, (2017), pp. 128–149.

[8] G. Carlsson, *Topology and data*, Bullentin of the American Mathematical Society, 46 (2009), pp. 255–308.

[9] C. Change, J. Cunningham, and G. Glover, *Influence of heart rate on the BOLD signal: the cardiac response function*, Neuroimage, 44 (2009), pp. 857–869.

[10] C. Change and G. Glover, *Relationship between respiration, end-tidal co2, and BOLD signals in resting-state fMRI*, Neuroimage, 47 (2009), pp. 1381–1393.

[11] Y. Chao-Gan and Z. Yu-Feng, *DPARSF: A MATLAB toolbox for "pipeline" data analysis of resting-state fMRI*, Front Syst Neurosci, 4 (2010), p. 13.

[12] D. V. Cicchetti, *Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology*, Psychological Assessment, 6(4) (1994), pp. 284–290.

[13] H. Edelsbrunner and J. Harer, *Persistent homology - a survey*, Contemporary Mathematics, 453 (2008), pp. 257–281.

[14] B. T. Fasy, J. Kim, F. Lecci, and C. Maria, *Introduction to the r package tda*, Hal, (2015), pp. 1–17.

[15] M. Ferguson, J. Anderson, and R. Spreng, *Fluid and flexible minds: Intelligence reflects synchrony in the brain's intrinsic network architecture*, Network Neuroscience, 1(2) (2017), pp. 192–207.

[16] B. Fischl et al., *Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain*, Neuron, 33(3) (2002), pp. 341–355.

[17] M. Fox, D. Zhang, A. Snyder, and M. Raichle, *The global signal and observed anticorrelated resting state brain networks*, Journal of Neurophysiology, 101 (2009), pp. 3270–3283.

[18] K. Friston, J. Hobson, C. Frith, J. Poline, J. Heather, and R. Frackowiak, *Registration and normalization of images*, Human Brain Mapping, 2 (1995), pp. 165–189.

[19] A. Golestani et al., *Mapping the end-tidal co2 response function in the resting-state BOLD fMRI signal: spatial specificity, test-retest reliability and effect of fMRI sampling rate*, Neuroimage, 104 (2015), pp. 266–277.

[20] E. Gordon et al., *Generation and evaluation of a cortical area parcellation from resting-state correlations*, Cereb Cortex, 26(1) (2016), pp. 288–303.

[21] L. Griffanti et al., *ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging*, Neuroimage, 95 (2014), pp. 232–247.

[22] H. Jo et al., *Effective preprocessing procedures virtually eliminate distance-dependent motion artifacts in resting state fMRI*, J Appl Math, (2013).

[23] S. W. K.O. McGraw, *Forming inferences about some intraclass correlation coefficients*, Psychological Methods, 1 (1996), pp. 30–46.

[24] R. Muller and P. Buttner, *A critical discussion of intraclass correlation coefficients*, Statisitics in Medicine, 13(23-24) (1994), pp. 2465–2476.

[25] K. Murphy et al., *The impact of global signal regression on resting state correlations: are anti-correlated networks introduced?*, Neuroimage, 44(3) (2009), pp. 893–905.

[26] K. Murphy and M. Fox, *Towards a consensus regarding global signal regression for resting state functional connectivity MRI*, Neuroimage, 154 (2017), pp. 169–173.

[27] M. Newman, *Modularity and community structure in networks*, PNAS, 103(23) (2006), pp. 8577–8582.

[28] S. Palande et al., *Revisiting abnormalities in brain network architecture underlying autism using topology-inspired statistical inference*, International Workshop on Connectomics in Neuroimaging, (2017), pp. 98–107.

[29] J. Power et al., *Methods to detect, characterize, and remove motion artifact in resting state fMRI*, Neuroimage, 84 (2014), pp. 320–341.

[30] ——, *Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion*, Neuroimage, 59(3) (2019), pp. 2142–2154.

[31] J. Power, B. Schlaggar, and S. Petersen, *Recent progress and outstanding issues in motion correction in resting state fMRI*, Neuroimage, 105 (2015), pp. 536–531.

[32] L. Shah et al., *Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state*, Brain and Behavior, (2016).

[33] K. Shmueli et al., *Low-frequency fluctuations in the cardiac rate as a source of variance in the resting-state fMRI BOLD signal*, Neuroimage, 38(2) (2007), pp. 306–320.

[34] B. Wang, *Visual detection of structural changes in time-varying graphs using persistent homology*, Nature Reviews Neuroscience, (2017). Unpublished paper, personal communication.

[35] A. Weissenbacher et al., *Correlations and anticorrelations in resting-state functional connectivity MRI: a quantitative comparison of preprocessing strategies*, Neuroimage, 47(4) (2019), pp. 1408–1416.

[36] E. Wong, S. Palande, B. Wang, B. Zielinski, J. Anderson, and P. T. Fletcher, *Kernel partial least squares regression for relating functional brain network topology to clinical measures of behavior*, International Symposium on Biomedical Imaging (ISBI), (2016).