

## Graph-Theoretical Methods for Statistical Inference on MR Connectome Data

### Abstract No:

180

### Authors:

Joshua Vogelstein<sup>1</sup>, John Bogovic<sup>1</sup>, Aaron Carass<sup>1</sup>, William Gray<sup>2</sup>, Jerry Prince<sup>1</sup>, Bennett Landman<sup>3</sup>, Dzung Pham<sup>1</sup>, Luigi Ferrucci<sup>4</sup>, Susan Resnick<sup>4</sup>, Carey Priebe<sup>1</sup>, R. Jacob Vogelstein<sup>2</sup>

### Institutions:

<sup>1</sup>Johns Hopkins University, Baltimore, MD, <sup>2</sup>Johns Hopkins University Applied Physics Laboratory, Laurel, MD, <sup>3</sup>Vanderbilt University, Nashville, TN, <sup>4</sup>National Institute on Aging, Baltimore, United States

### Introduction:

We are developing analytical tools to perform statistical inference on the connectome. Previous work has shown that simple measures of brain connectivity (e.g. total volume of white matter) are correlated with general cognitive features such as intelligence. Because the connectome can be represented as a large interconnected graph (in which nodes are neurons and synapses are edges), we hypothesize that the development of algorithms based on principles of graph theory will allow for greater prediction of performance on measures of specific cognitive functions like episodic memory. To this end, we have developed classification algorithms that are asymptotically optimal on brain-graph data under various simplifying assumptions. We applied these classifiers to a simulated dataset derived from the neuroimaging substudy of the Baltimore Longitudinal Study of Aging (BLSA), which includes multimodal MR images sufficient to create low-resolution connectomes for individual subjects.

### Methods:

Fiber tracking algorithms are applied to diffusion MRI data to identify and trace white matter fiber bundles from source to destination. By overlaying these results onto neuroanatomically parcellated gray matter regions, a graph is created with 70 nodes corresponding to cortical regions and edge values corresponding to the mean fractional anisotropy of tracts connecting any two regions. To evaluate algorithm performance, we simulated two classes of brain graphs derived from the actual brain graphs of BLSA participants to obtain a larger (artificial) population. Edge values were assumed to be Poisson random variables with parameters equal to observed values from one individual. In one simulated class, edges originating at five nodes had modified Poisson parameters ( $\pm 20\%$ ) relative to the other class. All other parameters were equal in the two classes. Performance of our classifiers was assessed by results on a held-out test set from this simulated dataset.

### Results:

Our simulations demonstrate good classification performance given limited training data using a number of different classifiers evaluated on the same dataset. The most general algorithm, the k-nearest neighbor (kNN) classifier, requires relatively more data to achieve good accuracy (50% with 80 samples, 60% with 180 samples), but is also universally consistent -- it attains the best possible performance given infinite data regardless of model assumptions. In contrast, the naive Bayes (NB) classifier requires relatively less data to obtain the same accuracy (60% with fewer than 10 samples, 85% with 180 samples), but is not universally consistent. We can further improve upon the naive Bayes classifier by using dimensionality reduction techniques. Standard metrics on graphs, such as small-worldness and measures of degree, are poor at discriminating between these two simulated classes of brain-graphs, implying that a good discrimination function must be sensitive to higher-order graph structure. The attached figure shows our basic results.

### Conclusions:

We demonstrate the ability to discriminate between two simulated classes of brain graphs derived from statistics of real data, with realistic assumptions on the amount and type of data available. Our novel algorithms outperform standard metrics on graphs, and some have the attractive property of asymptotically approaching optimal performance regardless of the statistics of the graphs themselves. In addition to the mathematical results, we show the utility of the TOADS and CRUISE algorithms, the CATNAP layout and the JIST toolkit for automated extraction of brain-graphs from multi-modal MR images. Current work focuses on using this pipeline to process the entire BLSA dataset and perform classification on real data.

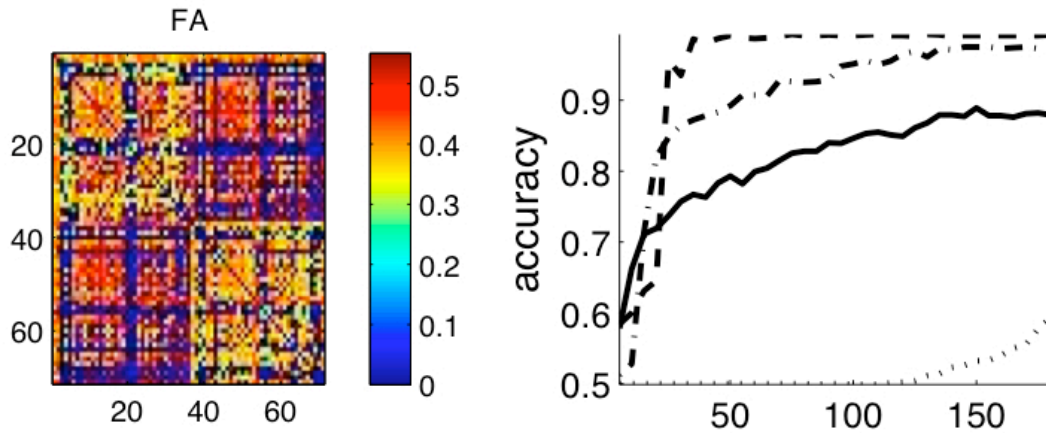


Figure 1: On the left is the connectome of a single subject, obtained using the TOADS and CRUISE algorithms with the CATNAP layout and the Jist toolkit for automated extraction of brain-graphs from multi-modal MR images. The colorbar indicates the magnitude of the mean fractional anisotropy for each projection. The right panel shows the performance of the various classifiers on the simulated data derived from this connectome. The k-Nearest Neighbors classifier (dotted line) converges relatively slowly compared with the Nave Bayes classifier (solid line), and the reduced dimensionality versions of the Nave Bayes algorithm (dashed and dotted-dashed lines).

#### Categories

- Diffusion MRI (Imaging Techniques and Contrast Mechanism)
- Multivariate Modeling, PCA and ICA (Modeling and Analysis)
- DTI Studies, Application (Neuroanatomy)
- Cognitive Aging (Cognition and Attention)
- Classification and Predictive Modeling (Modeling and Analysis)