

CONNECTOME REGRESSION

Dake Yang, PhD¹, Bill Shannon, PhD, MBA^{1,2}

¹BioRankings, ²Washington University School of Medicine

INTRODUCTION

What is the problem we are solving? How do connectomes change in relation to patient covariates (e.g., age, disease severity).

How does our solution work? We use **recursive partitioning** and the **Gibbs Distribution** that allows the connectome graphs to be treated as the outcome variable.

Gibbs Distribution

This **probability distribution** applies to **sets of graphs** and can be used to model connectomes from an experiment. It has two parameters describing the ‘central’ connectome and the spread of the observed connectomes around the central connectome (think mean and standard deviation):

$$P(g; g^*, \tau) = c(g^*, \tau) \exp(-\tau d(g, g^*))$$

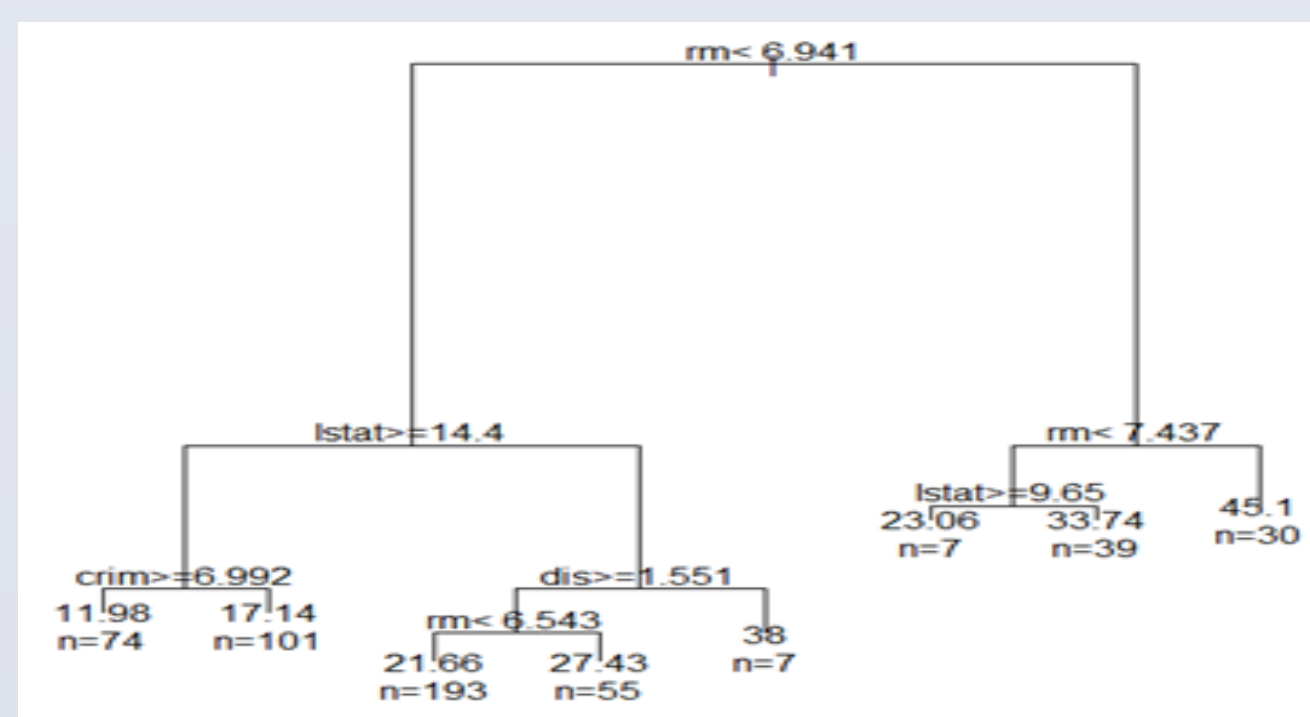
where

$$g^* = \text{central connectome}, \tau = \text{dispersion (spread of data)}$$

$$g = \text{set of observed connectomes}$$

Recursive Partitioning

Recursive Partitioning is a multivariable analysis method that finds a *partition* of a covariate space such that the dependent variable within each partition is in some sense similar. It finds a sequence of splits along the covariates which result in the best separation of the outcome at each step.



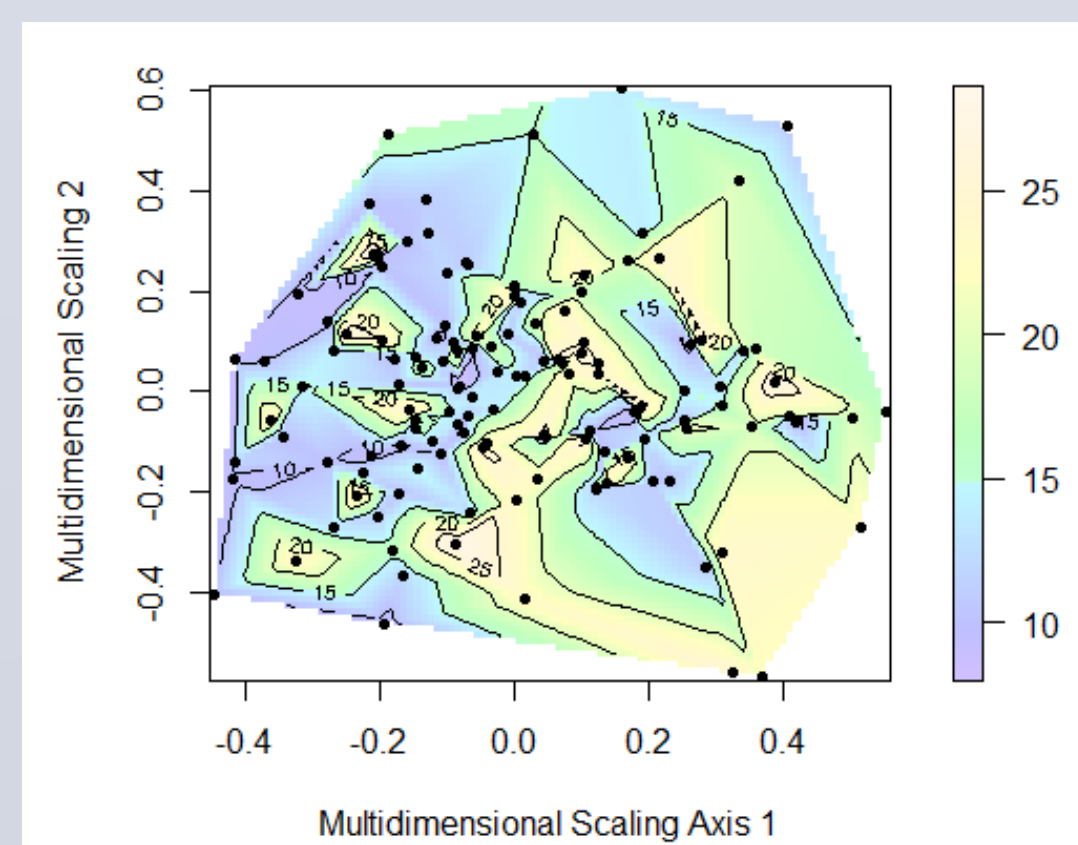
An example shown in the adjacent figure shows how neighborhood variables (e.g., average number of rooms in a house, % lower status population, crime rates) impact median cost of houses (in \$1,000’s of dollars) within the neighborhood.

In this example the decision tree shows how small houses with high % lower status population and high crime rates have a median house price of \$11,980. In contrast larger homes have a median cost of \$45,100.

Connectome Regression

We use recursive partitioning with subject graph-valued connectomes as the dependent variable, and subject-level covariates as the splitting independent variables. The **splitting rule** is based on the Gibbs Log Likelihood Ratio Test (LRT) statistic. The **pruning** is based on minimizing the within-group distance.

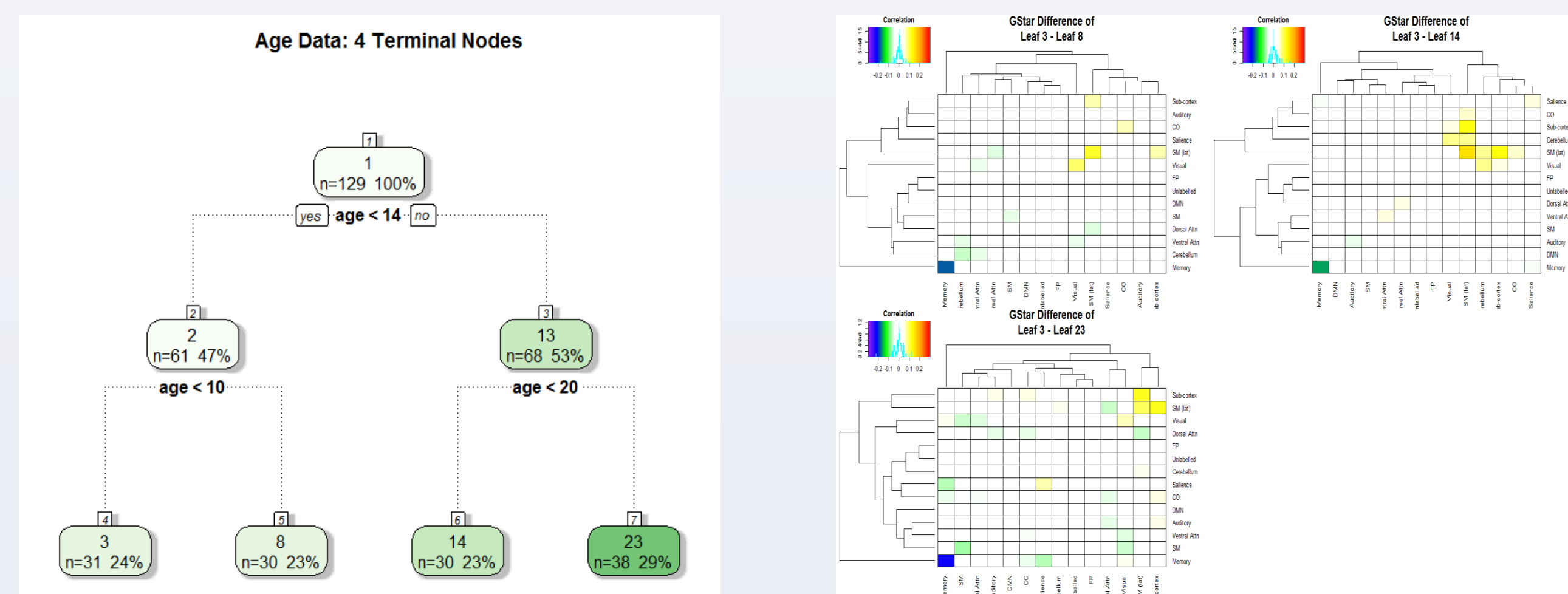
Age Data



fMRI data on 129 subjects ranging in age from 7 – 31 years old was provided by the Petersen/Schlaggar Lab from WUSM. These connectomes were parcellated in 105 ROIs (**nodes of the graphs**) with activity correlations between ROIs (**edges of the graph**).

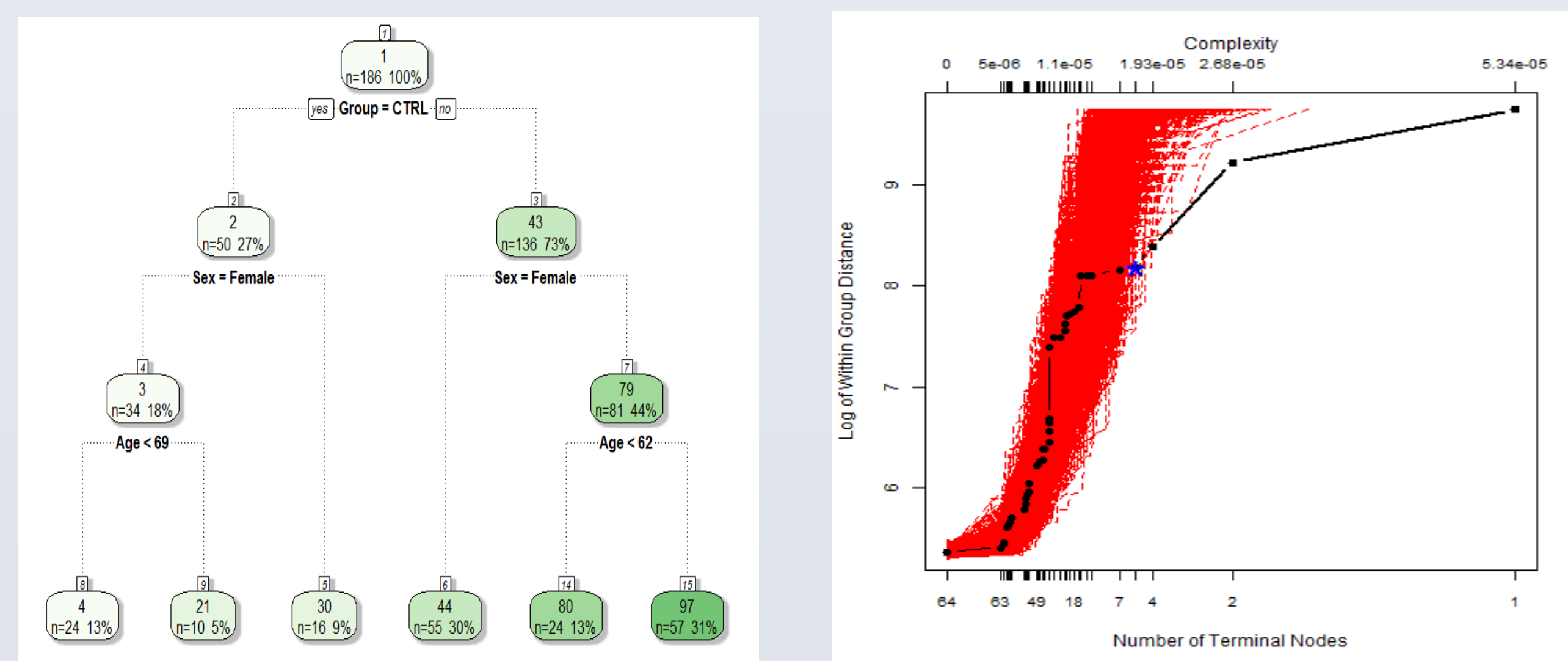
The MDS plot showing the projection from 105 dimensions to 2 uncovers little pattern in connectome structure as a function of age. This indicates the complexity of the problem and limitations of the MDS exploratory method for analyzing the data.

Recursive Partitioning of Age Data



Connectome regression split the 129 age samples into 4 groups by age: < 10; 10-14; 14 – 20, and > 20 years old. The heatmap difference plots compare the Gibbs central connectome graph of the youngest group to the three older groups. This indicates how the networks change with age.

Recursive Partitioning of Case/Control Data



186 connectomes from a case/control study were partitioned on 3 covariates: disease, gender, and age. In practice recursive partitioning produces a large tree which is pruned back to an optimal level based on a cost-complexity parameter which is a subjective measure. To make these analyses automatic, we developed a permutation test (right plot) for selecting the right sized tree.

Discussion

The Gibbs Distribution offers a new way for doing **translational statistical analysis of connectome data**. Here we presented connectome regression to measure how the connectome changes as covariates change. Previously we have used the Gibbs Distribution for performing formal hypothesis testing, regressing covariates onto connectomes, and mixture modeling.

Acknowledgements and Contact Information

We would like to thank the Petersen/Schlaggar at WUSM for invaluable collaborative support and access to scientists and data. This work has been supported by NIH R03 and SBIR Phase II funding, and the McDonnell Foundation.

REF: Statistics In Medicine, Volume 35, Issue 4, February 2016.

CONTACT: dake@biorankings.com, bill@biorankings.com