Graph Explorer

Gautam Prabhu, Elliot Liu, Devin Ramsden

January 25, 2018

1 Demo Response (ScienceInTheCloud)

The program worked for two out of three people on our team, although for one person, the program does not finish running; The program and its result look very technical, and people with less background knowledge in connectomics can have a hard time interpreting the results; The program could benefit from some links to pages offering basic knowledge on the subject.

On the bright side, running the program was easier than running other files in the class; The interface was clean and easy to navigate; The illustration in the beginning was helpful for understanding the program.

2 Mean Connectomes

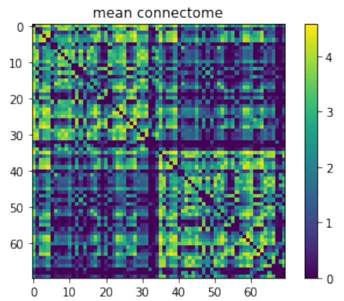

Figure 1: Mean Connectome for SWU4 Dataset (n=167).

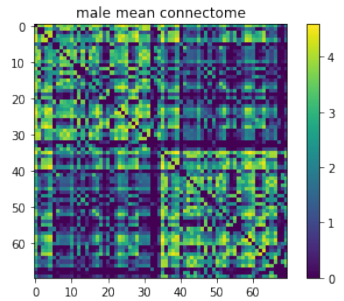

Figure 2: Mean Connectome for Males (n=81)

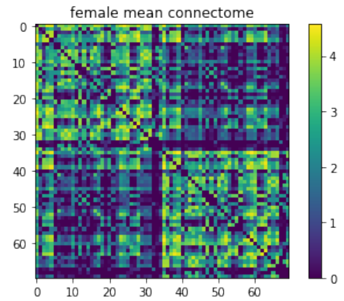

Figure 3: Mean Connectome for Females (n=86)

3 Covariates

3.1 Node Connectivity

The node connectivity is the minimun amount of nodes or edges that must be taken away to disconnect the graph. The classifier uses the node connectivty algorithm to see patterns in male and female brains to be able to decipher which one is which.

3.2 Clustering

The clustering algorithm creates models that are based upon the spacing of the nodes. A coefficient is then created and this coefficient is used to decide the sex of a given brain.

3.3 Triangles

The triangle algorithm finds out the number of triangles that include a node as a single vertex.The difference in triangles between male and female brain graphs are used in this classifier to make educated guesses on newly obtained graphs.

3.4 Degree Centrality

Degree centrality measures the fraction of nodes that each node is adjacent to. It helps us differentiate between male and female brains, since some regions in male brains might be more "connected" than those same regions in a female brain, and some might be less "connected".

3.5 Bipartite

A bipartite graph is one whose vertices can be divided into two disjoint, independent sets. This function had little impact on our score. Likely it is because the regions of human brain cannot be divided into two groups, where a region is not connected to any other region in that group. It does not help us differentiate between male and female brains since it should return false for all brains.

4 Classifier Results

With all five features above, our random forest returns a score of 0.58. Our classifier used 1000 estimators.

4.1 Alternative Classification Methods

We also tested Support Vector Classifiers with a variety of kernels (Gaussian, linear, sigmoid); the best classification accuracy achieved here was 0.66 with the sigmoid kernel (with scaled features)

We additionally tested logarithmic regression trained with stochastic gradient descent, for a 0.56 accuracy (with scaled features).

Next, we tested gradient-boosted decision trees with 1000 estimators and a learning rate of 1.0, which got an overall accuracy of 0.59. Gradient-boosted decision trees differ from random forests in that the individual trees are weak classifiers (high bias) which are not fully grown. In contrast, random forest uses fully-grown decision trees, which are low bias but high variance.

Our final test used a Multi-Layer Perceptron (MLP), a neural network with two hidden layers, of size 100 and 10 with a tanh() activation function, had an accuracy of 0.623 (with scaled features).