Creating Classifiers for the Identification of Subject Covariates in Brain Graph Data

Addison Wright, Chris Micek, Monica Rodriguez

January 14, 2016

1 Introduction

An abundance of brain graph data exist that have yet to be analyzed, specifically the southwest China data set. Unfortunately, the data is corrupted by errors in the form of misidentified edges or synapses, which will prove to be a challenge in our analysis. Our team proposes to train an array of classifiers to identify certain binary criteria, evaluating the strength of each one along the way. After the classifiers have been created, our team would also like to gain insight into the topological data features deemed the most important for the identification of graphs. In addition to the classification above, our team will also explore the potential benefits of performing elementary optmization analyses on graphs, hoping to use them as features of the data to see if they provide any further insight.

2 Project Outline

2.1 A Detailed List

The tasks below compose our plan of attack for this project. They are subject to change and addendum if our team deems it necessary.

1. Get the Southwest China data set

2. Understand and parse the data

3. Choose a programming language - MATLAB

4. Choose an atlas to use as a proof of concept. This will require the consultation of Will, Greg and Alex.

5. Merge files. This will require seperate code that likely will not be written in MATLAB. The goal of this task is to create files that will be easier to upload into MATLAB.

6. Identify particular binary classes, for the purposes of classification

7. Explore different classifiers, comparing their efficacy

8. Explore different graph statistics

9. Explore classifiers again using the aforementioned graph statistics as added information

10. Concurrent: Explore the important graph features and their relationship with Nueroscience; explore methods outlined in the paper, "Graph Classification using Signal-Subgraphs: Applications in Statistical Connectomics," written by Joshua T. Vogelstein, William R. Gray, R. Jacob Vogelstein, and Carey E. Priebe

11. Create Poster/ one-sheeter for presentation to the class

2.2 Schedule

Thurs. Jan. 14 - Have preliminary classifier code that works
Mon. Jan. 18 - Refine and debug classifier code, record and discuss results
Thurs. Jan 21 - Have completed poster

2.3 Allocation of Tasks

Chris - Try to combine .graphml files for easier loading, help code classifier
Addison - Read up on signal subgraphs and Josh Vogelstein's work, help code classifier
Monica - Help code classifier, research underlying neuronal mechanisms for results