Graph Classification using Signal-Subgraphs: Applications in Statistical Connectomics

Joshua T. Vogelstein, William R. Gray, R. Jacob Vogelstein, and Carey E. Priebe

Presented by Addison Wright

1/21/2016

Background

- A connectome is a brain-graph!
- MR connectomes utilize multi-modal Magnetic Resonance (MR) imaging to determine both the vertex and edge set for each individual

Opportunity

- A widely held view is that many psychiatric issues are fundamentally “connectopathies”
- It is desirable to know which vertices and/or edges are malfunctioning, such that therapy can be targeted to those locations. This is the motivating application for the work.

Challenge

- While statistical inference techniques for vector-valued data are widespread, statistical tools for the analysis of graph-valued data are relatively rare

Action

Action

- Consider the following “graph classification” question: given a collection of graphs and associated classes, how can one predict the class of a newly observed graph?
- Authors propose a statistical model for graph/class pairs which leads to a set of estimators to identify the class-conditional signal, or “signal-subgraph”

Methods

- Finding the signal-subgraph amounts to providing an understanding of the differences between the two graph classes
- This graph-model based approach is different from previous approaches.
- Simply representing the adjacency matrix with a vector and applying standard machine learning techniques ignores graph structure
- Computing a set of graph invariants and then classifying using only these invariants ignores vertex labels

Assumptions

- Adopt the common practice of identifying graphs with their adjacency matrices
- Assume that each graph has the same set of uniquely labeled vertices
- Assume edges are independent
- Assume the existence of a class conditional difference
- The likelihood of an edge between vertex u and v is given by a Bernoulli random variable with a scalar probability parameter

Finding a Signal Subgraph

- One might consider a search over all possible signal-subgraphs by plugging each one in to the classifier and selecting the best performing option. This strategy is intractable because the number of signal-subgraphs scales super-exponentially with the number of vertices. Aka: computationally taxing

Finding a Signal Subgraph

- Consider a hypothesis test for each edge.
  - The simple null hypothesis is that the class-conditional edge distributions are the same
  - The composite alternative hypothesis is that they differ
  - Given such hypothesis tests, one can construct test statistics and reject the null in favor of the alternative whenever the value of the test statistic is greater than some critical-value
    - Example test statistics include Fisher's and chi-squared

Finding a Signal Subgraph

- If you know the size of the signal subgraph, then an estimate of the signal-subgraph is the collection of s edges with minimal test statistics
- Once subgraph is found, can be used for various applications including classification and identification of class-specific connections -> can treat connectopathies

Pros:

- Signal Subgraph classifier performs better than traditional classifiers.

classifier	error	p-val
prior	0.50	< 0.01
naïve Bayes	0.41	< 0.01
lasso	0.27	< 0.02
graph-kNN	0.35	< 0.02
invariant-kNN	0.43	< 0.01
signal-subgraph	**0.16**	n\a

- Runs an order of magnitude quicker than the next most accurate classifier (lasso)
- Allows examination of Signal Subgraph which can beneficial implications for treatment

Cons:

- Independence between edges is almost certainly nonsense for real connectome data.
- Analysis suggests that even if the model were true, only about half the edges in the estimated signal-subgraph are in the actual signal-subgraph