EN.580.694: Statistical Connectomics
Final Project Proposal

Michael Seung · April 2, 2015

Evaluating the Mixed Membership Stochastic Block Model

Opportunity
Some networks such as neural networks may be modeled with a modified stochastic model called the mixed membership model. This model allows a vertex in a graph to be a member of more than one block, which may be more accurate for neural analysis because some members of one block may be interconnected to a member of another block

Challenge
We must modify the stochastic block model to allow a vertex (neuron) to be potentially assigned to multiple blocks. To quantify degree of membership to each block, there must be some sort of weighting ranging from 0 to 1 where 0 indicates no association to the block and 1 indicates no association to any blocks other than the one it belongs in.

Action
Different graphs with varying degrees of clustering will be generated. On one end of the spectrum, blocks will be well defined and at the other end of the spectrum, blocks will be ill-defined due to high interconnections between some members of different blocks. The mixed-membership stochastic model will run these graphs and try to generate block membership for each vertex.

Resolution
We will evaluate the clustering results given by the mixed membership stochastic block model by comparing them to the regular stochastic block model and the null distribution.

Future Work
Apply the mixed membership model to real data (i.e. sequenced connectomes in a species) to see if it does a better job than the regular stochastic block model.

Statistical Decision Theoretic

Sample space The set of graphs representing all vertices and edges. The sample space is given as Adjacency Matrices $A = \{0, 1\}^{N \times N}$

Model The Stochastic Block Model will be modified from the equation

$$P = \mathrm{SBM}^{k}_{n}\,(\rho, \beta), \text{ where } \rho \in \Delta_k \text{ and } \beta \in (0, 1)^{k \times k}$$

Action Space The action space will provide all possible actions with the given data or cluster assignments, given as

$$A = \{y \in \{0, 1\}^n\}$$

Loss function The function quantifying a loss during parameter estimation, given by

$$l: G_n \times A \rightarrow R_+$$

Risk function Expected value of the loss function with respect to the sample space, specifically the Stochastic Block Model

$$R = P \times l$$