# EN.580.694 ASSIGNMENT # 2

HEATHER PATSOLIC
STATISTICAL CONNECTOMICS

(1) **Sample Space**: The set of all undirected graphs (without loops) on $n$ vertices.

$$\mathcal{G}_n = (\mathcal{V}, \mathcal{E}, \mathcal{Y}),$$

where $\mathcal{V} = \{v_1, \ldots, v_n\}$, $\mathcal{E} = \{e_{11}, \ldots, e_{nn}\}$, and $\mathcal{Y} = \{0,1\}^n$ where $\mathcal{Y}$ is the set of latent labels for the vertices.

(2) **Model**: Stochastic Block Model on $n$ vertices and with $k$ blocks with parameters $\vec{p} \in \Delta_2$ and $\vec{\beta} \in (0,1)^{2\times 2}$.
(CHECK THIS: If I understand this correctly, $\vec{p}$ is an $n \times 1$ vector in which $p_i$ gives the group index of vertex $i$ (that is, $p_i$ denotes which group to which vertex $i$ belongs), either 0 or 1, and $\vec{\beta}$ is a $2 \times 2$ stochastic block matrix where $\beta_{ij}$ gives the probability that a node of type $i$ is adjacent to a node of type $j$. For example, $\beta_{11}$ gives the probability that a node in block 1 is adjacent to another node in block 1.)

(3) **Action Space**: $\mathcal{A} = \{y \in \{0,1\}^n\}$
(This is the space of possible actions, that is the set of all possible $n \times 1$ vectors (of latent variables) $y$.)

(4) **Loss Function**: $\ell : \mathcal{G}_n \times \mathcal{A} \to \mathbb{R}_+$
Often we use $ARI(\hat{y}, y)$, $AdjustedIndex = \frac{Index - \mathbb{E}(Index)}{MaxIndex - \mathbb{E}(Index)}$.
(Generally speaking, the loss function takes an observed graph and compares estimates of the vertex assignments to the actual vertex assignments, returning a 'score' of how good the estimation was with respect to the observed graph. In particular, the ARI (Adjusted Rand Index) measures the overall agreement between the observed graph clusters and the estimates of the vertex clusters by the model. It takes the Rand Index and compares it to the expected value of the Rand Index to see how the two compare.)

(5) **Risk Functional**: $R = \mathbb{E}_p(\ell) : P \times \ell \to \mathbb{R}_+$.
(The risk functional evaluates risk as the expected value of the loss function with respect to a specific SBM, P, with parameters $\vec{p}$ and $\vec{\beta}$. In other words, it is the expected loss incurred while using the particular model (SBM in this case).)

*Date*: February 17, 2015
*Due*: February 17, 2015.