1 Methods

1.1 Setting

- $\mathbb{G} : \Omega \to \mathcal{G}$, a graph-valued RV with samples $G_i \sim \mathbb{G}$, where $\mathcal{G}$ is the space of possible graphs and $G_i \in \mathcal{G}$.
- For each $G_i \in \mathcal{G}$, we have $G_i = (V, E_i)$; that is, each G_i is defined by a set of vertices V and a set of edges E_i, where $w_i : V \times V \to \mathbb{R}$, and $e_{uv}^{(i)} \in \mathbb{R}$. That is, w_i is the real-valued weight function for each graph.
- $\mathbb{A} : \Omega \to \mathcal{A}$, a adjacency-matrix-valued RV with samples $A_i \sim \mathbb{A}$, where $\mathcal{A}$ is the space of possible adjacency-matrices and $A_i \in \mathcal{A}$.
- $A_i \in \mathcal{A}$, and $\mathcal{A} \subseteq \mathbb{R}^{V \times V}$.
- Each graph G_i can be represented as an adjacency-matrix A_i.
- Let $Y : \Omega \to \mathcal{Y}$ be a discrete RV with samples y_i, where $y_i \sim Y$ denotes the class of graph/adjacency matrix i, and $y_i \in \mathcal{Y}$.
- We have some joint pdf given by $F_{\mathbb{G},Y}$, where we have a collection of n samples such that $\{(G_i, y_i)\}_{i=1}^n \sim F_{\mathbb{G},Y}$.

1.2 Statistical Goal

Given a new graph G_n, correctly estimate its class y_n assuming $(G_n, y_n) \sim F_{\mathbb{G},Y}$.

1.3 Model

1.3.1 Assumptions

- each graph has the same set of uniquely labeled vertices, V.
- edges are independent; that is, $F_{A,Y} = \prod_{(u,v) \in \mathcal{E}} F_{A_{uv},Y}$. Then by definition of the joint probability, it follows that $F_{A,Y} = F_{A|Y} F_Y$, and $F_{A_{uv}|y} = F_{A_{uv}|Y=y}$, and $\pi_y = F_{Y=y}$. Then $\sum_{y \in Y} \pi_y = 1$ denotes the prior parameter for the classes our graphs are sampled from.
- There exists a set of edges $\mathcal{S} = \{(u,v) \in \mathcal{E} | F_{uv|y=0} \neq F_{uv|y=1}\}$; that is, that there is some difference in the conditional probabilities for the edges $e_{uv} \in \mathcal{S}$ between the two classes.
- Assume the graphs are undirected and loop-less, where $A_{uv} \in [0,1]$ (TODO: generalize to $[a,b]$). Then the likelihood is given by the standard Beta RV with the scalar probability parameter:

$$\begin{aligned} F_{uv|y}(\mathbb{A}_{uv}) &= Beta(\mathbb{A}_{uv}; \alpha_{uv|y}, \beta_{uv|y}) \\ &= \frac{1}{Beta(\alpha_{uv}, \beta_{uv})} \mathbb{A}_{uv}^{\alpha_{uv|y}-1} (1 - \mathbb{A}_{uv})^{\beta_{uv|y}-1} \end{aligned} \tag{1}$$

where we have the normalizing constant:

$$Beta(\alpha_{uv|y}, \beta_{uv|y}) = \int_0^x y^{\alpha_{uv|y}-1} (1-y)^{\beta_{uv|y}} dy$$

Then we are left with the following model:

$$\begin{aligned} F_{\mathbb{G},Y} &= \{F_{\mathbb{A},Y}(A_i, y_i; \theta) \; \forall \; A_i \in \mathcal{A}, \; \forall \; y_i \in \mathcal{Y} : \theta \in \Theta\} \\ F_{\mathbb{A},Y}(A_i, y_i; \theta) &= \prod_{(u,v) \in \mathcal{S}} Beta(A_{uv}^{(i)}; \alpha_{uv|y}, \beta_{uv|y}) \pi_y \prod_{(u,v) \notin \mathcal{S}} Beta(A_{uv}^{(i)}; \alpha_{uv|y}, \beta_{uv|y}) \end{aligned} \tag{2}$$

1.4 Test Statistic

We use a simple hypothesis defined on each edge as follows:

- $H_0 : F_{uv|y=0} = F_{uv|y=1}$ with parameters indicating the null hypothesis given by Θ_0
- $H_a : F_{uv|y=0} \neq F_{uv|y=1}$ with parameters indicating the alternative hypothesis given by Θ_a

Then we can construct test statistics $T_{uv}^{(n)} : \mathcal{T}_n \to \mathbb{R}_+$ using the generalized likelihood test framework:

$$\Lambda = \frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})} = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)}$$

where $\Theta = \Theta_0 \cup \Theta_a$, where $\Theta_0 \cap \Theta_a = \emptyset$. The intuition here is that when the maximum likelihood estimate for the data is found outside of Θ_0 and consequently in Θ_a, then $L(\hat{\Theta}_0) < L(\hat{\Theta}) = L(\hat{\Theta}_a)$ and we should consider rejecting the null hypothesis H_0 in factor of the alternative hypothesis H_a. It is obvious that since $0 < L(\hat{\Theta}_0) \leq L(\hat{\Theta})$, then $0 \leq \Lambda \leq 1$.

1.4.1 Likelihood Estimators

Beta parameters To estimate the class-conditional likelihood parameters $\alpha_{uv|y}$ and $\beta_{uv|y}$, we can use the following identities:

$$\mu_{uv|y} = \mathbb{E}[A_{uv|y}] = \frac{\alpha_{uv|y}}{\alpha_{uv|y} + \beta_{uv|y}}$$
$$\sigma^2_{uv|y} = \mathbb{E}[(A_{uv|y} - \mu_{uv|y})^2] = \frac{\alpha_{uv|y}\beta_{uv|y}}{(\alpha_{uv|y} + \beta_{uv|y})^2(\alpha_{uv|y} + \beta_{uv|y} + 1)}$$

solving this system of equations, we find that:

$$\alpha_{uv|y} = \left(\frac{1 - \mu_{uv|y}}{\sigma^2_{uv|y}} - \frac{1}{\mu_{uv|y}}\right)\mu^2_{uv|y} \tag{3}$$
$$\beta_{uv|y} = \alpha_{uv|y}\left(\frac{1}{\mu_{uv|y}} - 1\right) \tag{4}$$

Prior Parameters (TODO: Add section once we develop some sort of classification technique).

1.5 Algorithm

1.5.1 High Level

- For each edge uv:
  1. For each observation i:
     (a) compute $\hat{\theta}_{0,1} = \alpha_{uv|0,1}, \beta_{uv|0,1}$ from class 0 and class 1 data together without the observation i
     (b) compute $\hat{\theta}_0 = \alpha_{uv|0}, \beta_{uv|0}$ from class 0 data and $\hat{\theta}_1 = \alpha_{uv|1}, \beta_{uv|1}$ from class 1 data, without the observation i
     (c) compute $\Lambda_{uv}^{(i)}$ as the ratio of the likelihood of the held out observation given the parameters $\hat{\theta}_{0,1}$ with respect to the max of the set $\hat{\theta}_{0,1}$, $\hat{\theta}_0$, $\hat{\theta}_1$.
  2. TODO: Ask Jovo how to get a p-value for edge uv from the set of $\{\Lambda_{uv}^{(i)}\}_{i=1}^{n}$... is it just the average of $\Lambda_{uv}^{(i)}$ over all i?
- choose the set of edges that satisfy $\Lambda_1 \leq \Lambda_2 \leq \ldots \leq k$ for some threshold k.

1.5.2 Pseudocode

Algorithm 1: weighted-Signal-Subgraphs(A, Y, k)

Input:
$A = \{A_i\}_{i=1}^n, A_i \in \{[0,1]\}^{|V| \times |V|}$ where n is the number of subjects, $|V|$ is the number of vertices in our graph, and our graph is undirected and loop-less. Then let A_i denote the adjacency-matrix associated with the i^{th} observation, and $A_{uv}^{(i)}$ denote the weight of the (u, v) edge for observation i.
$Y = \{y_i\}_{i=1}^n, y_i \in \{0, 1\}$ where y_i denotes the class that observation i is part of.
$k \in \{[0, 1]\}$: the p-value below which we will not consider edges when greedily searching.
Result:
$S = \{(u, v) \in \mathcal{E} : T_{uv} \leq k\}$, the set of edges that satisfy our given stopping criterion k, forming our signal subgraph.

$S = \{\}$
for $u \in 2 : |V|$ **do**
  for $v \in u + 1 : |V|$ **do**
    for $i = 1 : n$ **do**
      `// ignore held-out observation for now`
      Let $A^\dagger = A \setminus A_i$, and $Y^\dagger = Y \setminus y_i$.
      `// set of class 0 graphs without the held-out observation`
      Let $A^0 = \left\{A_i^\dagger : y_i = 0\right\}$.
      `// set of class 1 graphs without the held-out observation`
      Let $A^1 = \left\{A_i^\dagger : y_i = 1\right\}$.
      `// Compute the MLE from equation (3) given the` μ `and` σ `of the sets of values at a given edge.`
      Compute $\hat{\theta}_{uv}^{0,1}$ from $A_{uv}^\dagger$.
      Compute $\hat{\theta}_{uv}^j$ from the set A_{uv}^j for $j = \{0, 1\}$. `// Compute the General Likelihood Ratio, producing a value` $\Lambda_{uv}^{(i)}$ `between 0 and 1.`
      Compute $\Lambda_{uv}^{(i)} = \frac{L(\hat{\theta}_{uv}^{0,1})}{\max_{q \in \{0,1,\{0,1\}\}}\{L(\hat{\theta}_{uv}^q)\}}$
    end
    $\lambda_{uv} = \frac{1}{n}\sum_{i=1}^n \Lambda_{uv}^{(i)}$, the mean General-Likelihood ratio for a given iteration. (TODO: Jovo is this valid?) `// If the p-value of our likelihood-ratio achieves a particular threshold, add it to our subgraph`
    if $\Lambda_{uv} \leq k$ **then**
      $S = S \cup (u, v)$
    end
  end
end
return S