Statistical Connectomics: On Making Decision Theoretic Statements for clustering SBMs

Michelle Chyn

February 17, 2015

1. Sample Space ...Set of potential observations
   - $Adj \, \epsilon \, \{0,1\}^{n \times m}$
   - $y \, \epsilon \, \{0,1\}^{n}$
2. Model $P = \{P_\theta : \theta \, \epsilon \, \Theta\}$collection of probability distributions ex. SBM
   - $SBM_n^k \left(\overrightarrow{\rho}, \overrightarrow{\beta} \right)$
   - $\overrightarrow{\rho} \, \epsilon \, \triangle_2$
   - $\overrightarrow{\beta} \, \epsilon \, (0,1)^{2 \times 2}$
3. Action Space $A = \{a_1, a_2, ..., a_n\}$assigning each vertex to a cluster, label switching when needed
   - $A = \{y \, \epsilon \, (0,1)^n\}$
4. Decision Rule Class $\Phi : \phi ... \rightarrow A$ set of functions that maps sample space to action space
5. Loss Function $l : \, P \times A \rightarrow R_+$mapping from model and action space to R
   - $l : \, G_n \times \, A \, \rightarrow R_+$with permutations or
   - $l : \, ARI \, (\hat{y}, y)$
6. Risk Functional $R : \, L \times \Phi \times P \rightarrow R_+$expected loss under true distribution of data
   - $R : \, A \times A \times P$
   - $E_p \, [l]$