Notes on Multi-level SBM's for Behavioral Modeling Data

Say we have L categories of measurements on N subjects, leading to L adjacency matrices $A_1, \ldots, A_L$ each of dimension $N \times N$ after performing KNN to construct networks for each category of measurements.

We wish to fit an SBM to $A_1, \ldots, A_L$ that accounts for the association among the L different "levels" (i.e., measurement categories) and allows for community detection at each level.

Peng and Carvalho (2016) propose a GLM approach to fitting the SBM by assuming

$$A_{ij} \,|\, \mathbf{z}, \boldsymbol{\theta} \sim \text{Bern}(\theta_{z_i,z_j}),$$

where z_i and z_j are group membership indicators and θ_{z_i,z_j} controls association between communities i and j. Peng and Carvalho model θ_{z_i,z_j} using logistic regression as

$$\text{logit}(\theta_{z_i,z_j}) = \gamma_{z_i,z_j} + \eta_i + \eta_j,$$

where γ_{z_i,z_j} measures the "connectivity" between communities i and j ($i \neq j$), and η_i, η_j are for node specific intercepts measuring the expected degrees of nodes i and j. Peng and Carvalho assume a multivariate normal prior for the $\binom{K}{2}$ γ_{z_i,z_j} parameters.

Perhaps we can extend this GLM representation to model A_{ijl} for $l = 1,\ldots,L$. One direction is to possibly first extend γ_{z_i,z_j} to the multi-level setting by introducing $\gamma^l_{z_i^l,z_j^l}$, which captures the connectivity between communities z_i and z_j in level l. We might organize the $\gamma^l_{z_i^l,z_j^l}$'s in a matrix as

$$\mathbf{G}_{L\times\binom{K}{2}} = \begin{bmatrix} \gamma^1_{z_1,z_1} & \gamma^1_{z_1,z_2} & \cdots & \gamma^l_{z_{K-1},z_K} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma^L_{z_1,z_1} & \gamma^L_{z_1,z_2} & \cdots & \gamma^L_{z_{K-1},z_K} \end{bmatrix}.$$

We might assume a matrix normal distribution for $\mathbf{G}$ to extend the multivariate normal distribution used by Peng and Carvalho. Specifically, we might let

$$\mathbf{G} \sim \text{MatNorm}\left(\mathbf{M}_{L\times\binom{K}{2}}, \mathbf{U}_{L\times L}, \mathbf{V}_{\binom{K}{2}\times\binom{K}{2}}\right),$$

where $\mathbf{U}$ controls dependence among rows (i.e., levels) of $\mathbf{G}$. In this way, we are accounting for association between the different measurement categories.