0.1 Schema of our interest

To explain the aim of this report, recall the results of the seminal paper [?].

$$L \to \mathscr{L}_{\text{sym}} \xrightarrow{\text{Sp.Cl.}} Z \tag{1}$$

The idea is that the eigenvectors of an observed graph Laplacian L converge to a rotated version of the eigenvectors of the *population* Laplacian $\mathscr{L}_{\text{sym}}$. In a sense this tells us that all Laplacians of observations coming from a fixed model are essentially $\mathscr{L}_{\text{sym}}$. Then, it is proved that applying the spectral clustering algorithm to the population Laplacian of an SBM model with k communities provides precisely the true community assignments Z. This happens because the matrix Q, which is the matrix with eigenvectors of $\mathscr{L}_{\text{sym}}$ as columns, contains only k distinct rows which correspond to the community structure (see Lemma 3.1. of [?]).

Our point of view is to see the recovery of Z as the result of the maximization of the expected ELBO:

$$Z = \underset{\tau}{\operatorname{argmax}}\, \mathbb{E}\left[\text{ELBO}\left(\tau, \gamma^*\right)\right], \tag{2}$$

where γ^* is the true parameter of the underlying SBM. If the equality above is true, then one can use classical theory to show that the maximizer $\hat{\tau}$ of the ELBO (without the expectation) converges to Z.

0.2 Spectral structure of an SBM

0.2.1 Expected Laplacian

An important object is the expected Laplacian of an SBM, since one could expect for the observed Laplacian to converge to it in some sense as the size of the SBM grows.

Expected unnormalized Laplacian Consider the expected unnormalized Laplacian, where the expectation is taken conditionally on the model parameters and the cluster assignments. Denote it by $\mathscr{L} := \mathbb{E}[L|Z]$. Since $A_{ij} \sim \text{Ber}\left(\gamma_{\Omega_i\Omega_j}\right)$, it follows that $\mathbb{E}\left[A_{ij}\right] = \gamma_{\Omega_i\Omega_j}$. Here Ω_i denotes the community that node i belongs to. Reorder the rows and columns such that the first n_1 rows and columns correspond to the first community, the subsequent n_2 rows and columns correspond to the second community, and so on. Then, a straightforward calculation reveals that

$$\mathscr{L} = \begin{pmatrix} \mathscr{L}_1 & -\gamma_{12}\mathbf{1}_{n_1\times n_2} & \dots & -\gamma_{1k}\mathbf{1}_{n_1\times n_k} \\ & \mathscr{L}_2 & & \vdots \\ & & \ddots & \\ & & & \mathscr{L}_k \end{pmatrix},$$

where the entries in the diagonal are the expected Laplacians of each community. They can be written as

$$\begin{aligned} \mathbf{n} &:= (n_1, \dots, n_k), \\ \bar{d}_i &:= \Gamma_{i\cdot} \cdot \mathbf{n} - \gamma_{ii} \\ \mathscr{L}_i &= \bar{d}_i I_{n_i} - \gamma_{ii} J_{n_i}. \end{aligned}$$

Eigenvectors of $\tilde{L}$. One can easily verify that the following vectors are eigenvectors of $\mathscr{L}$:

$$\begin{aligned} v_1 = \mathbf{1}_n &\implies \lambda_1 = 0 \\ v_2 = \begin{pmatrix} -\frac{n_2}{n_1}\mathbf{1}_{n_1} \\ \mathbf{1}_{n_2} \end{pmatrix} &\implies \lambda_2 = n\gamma_{12}. \end{aligned}$$