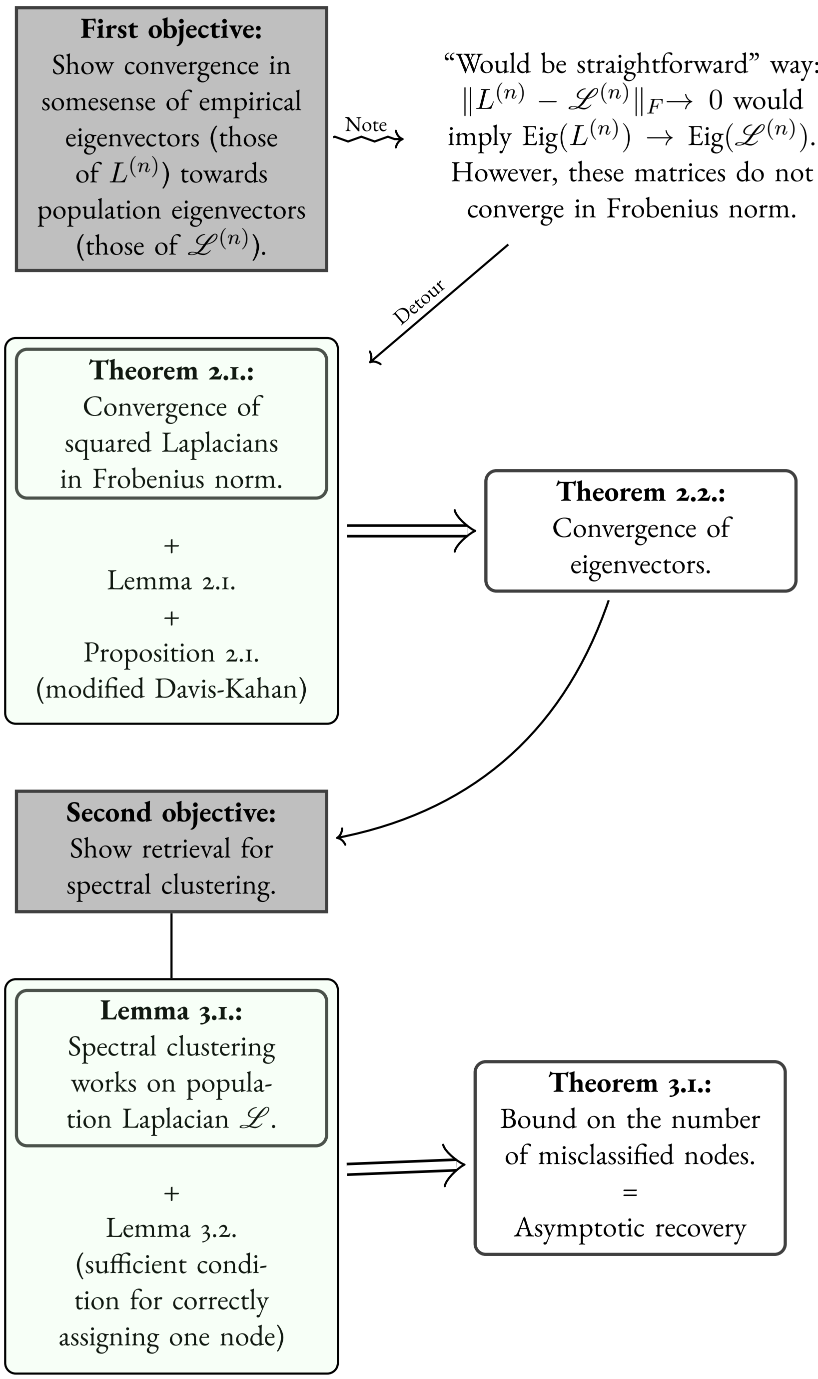
0.1 Resuming the article of Rohe

0.1.1 General comments

A natural question one might ask when studying the relationship between SBM models and spectral clustering algorithms is when the spectral clustering algorithm can recover the community partitions of graphs generated under some SBM. The seminal work [?] was among the first to study this question.

Diagram of results. Here is a simple diagram explaining their results.


First objective:
Show convergence in somesense of empirical eigenvectors (those of $L^{(n)}$) towards population eigenvectors (those of $\mathscr{L}^{(n)}$).
Note
"Would be straightforward" way: $\|L^{(n)} - \mathscr{L}^{(n)}\|_F \to 0$ would imply $\mathrm{Eig}(L^{(n)}) \to \mathrm{Eig}(\mathscr{L}^{(n)})$. However, these matrices do not converge in Frobenius norm.
Detour
Theorem 2.1.:
Convergence of squared Laplacians in Frobenius norm.
+
Lemma 2.1.
+
Proposition 2.1.
(modified Davis-Kahan)
Theorem 2.2.:
Convergence of eigenvectors.
Second objective:
Show retrieval for spectral clustering.
Lemma 3.1.:
Spectral clustering works on population Laplacian $\mathscr{L}$.
+
Lemma 3.2.
(sufficient condition for correctly assigning one node)
Theorem 3.1.:
Bound on the number of misclassified nodes.
=
Asymptotic recovery

Notation and definitions. I will concentrate here all notations and definitions needed.

Definition 1 (Latent space model)**.** For i.i.d. random vectors $z_1, \ldots, z_n \in \mathbb{R}^k$ and random adjacency matrix $A \in \{0,1\}^{n\times n}$, let $\mathbb{P}(A_{ij}|z_i, z_j)$ be the probability mass function of A_{ij} conditioned on z_i, z_j. If a probability distribution on A has the conditional dependence relationships

$$\mathbb{P}(A|z_1, \ldots, z_n) = \prod_{i<j} \mathbb{P}(A_{ij}|z_i, z_j),$$

and $\mathbb{P}(A_{ii} = 0) = 1$ for all i, then it is called an *undirected latent space model*.

Choice of Laplacian

They use the matrix $L = D^{-1/2}AD^{-1/2}$ as Laplacian for the algorithm. This is justified, since this matrix has the same eigenvectors as the more common normalized Laplacian $\tilde{L} = I - L$, and the eigenvectors are the only thing that matters in their case. However, due care should be taken if the estimators found in this report end up depending on eigenvalues, as these are not the same and should be transformed from one Laplacian to the other.

DEFINE THE SPECTRAL CLUSTERING ALGORITHM !

0.1.2 Convergence of eigenvectors

All the results of this first part are valid for latent space models, which include the SBM.

0.1.3 Retrieval for spectral clustering

The results of this part focus on the special case of the latent space model being an SBM, and show the asymptotic consistency when using the spectral clustering algorithm to estimate the latent variables $Z^\star$.