Estimating the stochastic blockmodel on a latent space of unknown dimension for an unknown number of communities

Notes

Francesco Sanna Passino

November 25, 2018

1 GRDPG and spectral embeddings

The adjacency matrix $\mathbf{A} \in \{0,1\}^{n\times n}$ of an undirected graph $\mathbb{G}=(V,E)$, with $A_{ij}=\mathbb{1}_E\{(i,j)\}$ is usually embedded into a latent space of dimension d using two different procedures: the spectral adjacency embedding and the spectral Laplacian embedding.

Definition 1 (Spectral adjacency embedding for the undirected graph)**.** *Given an undirected graph with symmetric adjacency matrix $\mathbf{A} \in \{0,1\}^{n\times n}$, and a positive integer $d\geq 1$, consider the spectral decomposition*

$$\mathbf{A} = \hat{\mathbf{\Gamma}}_d\hat{\mathbf{\Lambda}}_d\hat{\mathbf{\Gamma}}_d^\top + \hat{\mathbf{\Gamma}}_d^\perp\hat{\mathbf{\Lambda}}_d^\perp(\hat{\mathbf{\Gamma}}_d^\perp)^\top$$

where $\hat{\mathbf{\Lambda}}_d$ is a $d\times d$ diagonal matrix containing the top d eigenvalues in magnitude, and $\hat{\mathbf{\Gamma}}_d$ is a $n\times d$ matrix containing the corresponding orthonormal eigenvectors. The spectral adjacency embedding $\hat{\mathbf{X}}_d=[\hat{\boldsymbol{x}}_{1d},\dots,\hat{\boldsymbol{x}}_{nd}]^\top$ of $\mathbf{A}$ in $\mathbb{R}^d$ is:

$$\hat{\mathbf{X}}_d = \hat{\mathbf{\Gamma}}_d|\hat{\mathbf{\Lambda}}_d|^{1/2} \in \mathbb{R}^{n\times d},$$

where the operator $|\cdot|$ applied to a matrix returns the absolute value of its entries.

Definition 2 (Spectral Laplacian embedding for the undirected graph)**.** *Given an undirected graph with symmetric adjacency matrix $\mathbf{A} \in \{0,1\}^{n\times n}$, and a positive integer $d\geq 1$, consider the normalised Laplacian matrix*

$$\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2},\ \mathbf{D} = \mathrm{diag}\left(\sum\nolimits_{j=1}^n A_{ij}\right) = \mathrm{diag}\left(\sum\nolimits_{i=1}^n A_{ij}\right),$$

and its spectral decomposition

$$\mathbf{L} = \hat{\mathbf{\Gamma}}_d\tilde{\mathbf{\Lambda}}_d\tilde{\mathbf{\Gamma}}_d^\top + \tilde{\mathbf{\Gamma}}_d^\perp\tilde{\mathbf{\Lambda}}_d^\perp(\tilde{\mathbf{\Gamma}}_d^\perp)^\top$$

where $\tilde{\mathbf{\Lambda}}_d$ is a $d\times d$ diagonal matrix containing the top d eigenvalues in magnitude, and $\tilde{\mathbf{\Gamma}}_d$ is a $n\times d$ matrix containing the corresponding orthonormal eigenvectors. The spectral Laplacian embedding $\tilde{\mathbf{X}}_d=[\tilde{\boldsymbol{x}}_{1d},\dots,\tilde{\boldsymbol{x}}_{nd}]^\top$ of $\mathbf{A}$ in $\mathbb{R}^d$ is:

$$\tilde{\mathbf{X}}_d = \tilde{\mathbf{\Gamma}}_d|\tilde{\mathbf{\Lambda}}_d|^{1/2}.$$

The generalised random dot product graph, a generic latent position model for graphs, is defined below.

Definition 3 (Generalised random dot product graph, GRDPG)**.** *Define a subset* $\mathcal{X} \subseteq \mathbb{R}^d$ *such that* $\boldsymbol{x}_i^\top \mathbf{I}_\pm(p,q)\boldsymbol{x}_j \in [0,1]\ \forall\ \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{X}$ *with*

$$\mathbf{I}_\pm(p,q) = \mathrm{diag}(\underbrace{1,\ldots,1}_{p},\underbrace{-1,\ldots,-1}_{q}),\ p+q=d.$$

Also, let $(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n) \sim \mathcal{F}$ *with* $\mathbf{X} = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)^\top \in \mathbb{R}^{n\times d}$*. Then,* $\mathbf{A}$ *is a hollow matrix such that, conditional on* $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$*, for* $i<j$*,*

$$A_{ij} \overset{iid}{\sim} \mathrm{Bernoulli}\left\{\boldsymbol{x}_i^\top \mathbf{I}_\pm(p,q)\boldsymbol{x}_j\right\},\ A_{ij} = A_{ji},$$

and we write:

$$(\mathbf{A},\mathbf{X}) \sim \mathrm{GRDPG}_{p,q}(\mathcal{F}).$$

2 The stochastic block model as a GRDPG

To represent as a GRDPG model the K-community stochastic block model (SBM) with generating matrix $\mathbf{B} \in [0,1]^{K\times K}$ and community allocation vector $\boldsymbol{\pi} \in \mathcal{S}_{K-1}$, the $K-1$ probability simplex, one must choose $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_K \in \mathbb{R}^d$ such that $\boldsymbol{v}_j^\top \mathbf{I}_\pm(p,q)\boldsymbol{v}_h = B_{jh}\ \forall\ j,h \in \{1,\ldots,K\}$. Rubin-Delanchy et al. (2017) suggests the following algorithm for estimation of the latent positions in a SBM interpreted as a GRDPG:

Algorithm 1: Spectral estimation of the stochastic block model (spectral clustering)

Input: adjacency matrix $\mathbf{A}$ (or the Laplacian matrix $\mathbf{L}$), dimension d, and number of communities $K \geq d$.

1 compute spectral embeddings $\hat{\mathbf{X}}_d = [\hat{\boldsymbol{x}}_{1d},\ldots,\hat{\boldsymbol{x}}_{nd}]^\top$ or $\tilde{\mathbf{X}}_d = [\tilde{\boldsymbol{x}}_{1d},\ldots,\tilde{\boldsymbol{x}}_{nd}]^\top$ into $\mathbb{R}^d$,

2 fit a Gaussian mixture model with K components,

Result: return cluster centres $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_K \in \mathbb{R}^d$ and node memberships $z_1,\ldots,z_n$.

Note that it is not possible to estimate the $\boldsymbol{v}_j$'s directly, but only up to an orthogonal transformation $\boldsymbol{v}_j^\star = \mathbf{Q}_n^{-1}\boldsymbol{v}_j$, where $\mathbf{Q}_n$ is a random matrix in the indefinite orthogonal group $\mathbb{O}(p,q)$. Note that since $\mathbf{Q}_n^{-1} \in \mathbb{O}(p,q)$, $(\mathbf{Q}_n^{-1}\boldsymbol{v}_{z_i})^\top \mathbf{I}_\pm(p,q)(\mathbf{Q}_n^{-1}\boldsymbol{v}_{z_j}) = \boldsymbol{v}_{z_i}^\top \mathbf{I}_\pm(p,q)\boldsymbol{v}_{z_j}$. Assuming the Gaussian mixture modelling for consistent estimation of the latent positions, it is possible to roughly assume that $\hat{\boldsymbol{x}}_i \overset{d}{\sim} \mathbb{N}(\boldsymbol{v}_{z_i}^\star, \boldsymbol{\Sigma}_{z_i}^\star)$ for some community-specific covariance matrix $\boldsymbol{\Sigma}_{z_i}^\star$. Intuitively, this approximately holds because, taking a graph with m nodes, and restricting the attention to the first n nodes, with $n<m$:

$$\mathbf{Q}_m\hat{\boldsymbol{x}}_i \xrightarrow{d} \mathbb{N}\{\boldsymbol{v}_{z_i}, m^{-1/2}\boldsymbol{\Sigma}(\boldsymbol{v}_{z_i})\},\ m\to\infty,\ i=1,\ldots,n,$$

where $\boldsymbol{\Sigma}(\boldsymbol{v}_{z_i})$ can be analytically computed when $\mathbf{Q}_m$ is known.

Therefore, the clusters obtained from the embeddings are approximately Gaussian. In this work, it is of interest to consider K and d, the number of communities and latent dimension of the latent positions, to be random and learned from the data. Among other standard techniques for GMMs with unknown number of components, MCMC methods are a common choice, but no attention is devoted to the selection of an appropriate number of dimensions of the latent positions.

3 Learning the number of communities and latent dimension in SBMs from spectral embeddings

For simplicity, the embeddings will be generically denoted as $\mathbf{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]^\top \in \mathbb{R}^{n\times m}$, $\boldsymbol{x}_i \in \mathbb{R}^m$ for some $m > K$, and the non-identifiable latent position $\boldsymbol{v}_j^\star$ are more practically renamed $\boldsymbol{v}_j$. The notation $\boldsymbol{x}_{i:d}$ denotes the first d elements $(x_1, \dots, x_d)$ of the vector $\boldsymbol{x}_i$, and similarly $\boldsymbol{x}_{id:}$ denotes the last $m-d$ elements $(x_{d+1}, \dots, x_m)$ of the vector. We assume that, given the number of communities K and the latent dimension d, the embeddings are generated from K d-dimensional community-specific Gaussians in the first d components, say $\mathbf{X}_{:d}$, and from a generic $(m-d)$-dimensional Gaussian (which is not cluster-specific) for the remaining components $\mathbf{X}_{d:}$. Therefore, introducing latent community assignments $\boldsymbol{z} = (z_1, \dots, z_n)$, the model can be expressed as follows:

$$
\begin{aligned}
\boldsymbol{x}_i|d, K, z_i, \boldsymbol{v}_{z_i}, \boldsymbol{\Sigma}_{z_i}, \boldsymbol{v}_r, \boldsymbol{\Sigma}_r &\overset{d}{\sim} \mathbb{N}_m\left(\begin{bmatrix}\boldsymbol{v}_{z_i}\\ \boldsymbol{v}_r\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{z_i} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_r\end{bmatrix}\right),\ i = 1, \dots, n,\\
(\boldsymbol{v}_k, \boldsymbol{\Sigma}_k)|d, K &\overset{iid}{\sim} \mathrm{NIW}_d(\boldsymbol{v}_{0:d}, \kappa_0, \nu_0 + d - 1, \boldsymbol{\Delta}_{0:d}),\ k = 1, \dots, K\\
(\boldsymbol{v}_r, \boldsymbol{\Sigma}_r)|d, K &\overset{d}{\sim} \mathrm{NIW}_{m-d}(\boldsymbol{v}_{0d:}, \kappa_0, \nu_0 + m - d - 1, \boldsymbol{\Delta}_{0d:}),\\
z_i|\boldsymbol{\theta}, K &\overset{iid}{\sim} \mathrm{Multinoulli}(\boldsymbol{\theta}),\ i = 1, \dots, n,\ \boldsymbol{\theta} \in \mathcal{S}_{K-1},\\
\boldsymbol{\theta}|K &\overset{d}{\sim} \mathrm{Dirichlet}\left(\frac{\alpha}{K}, \dots, \frac{\alpha}{K}\right),\\
k &\overset{d}{\sim} \mathrm{Geometric}(\omega),\\
d &\overset{d}{\sim} \mathrm{Geometric}(\delta).
\end{aligned}
$$

where $\mathcal{S}_{K-1}$ is the $K-1$ probability simplex. Note that the inverse Wishart has been partially re-parametrised using a parameter $\nu_0 > 0$ and adding the corresponding dimension to obtain the required constraint $\nu > d-1$ for the generic parametrisation and interpretation of ν in the inverse Wishart. Also note that m can be chosen to be equal to K, when fixed, for parsimony, or equal to n to have the maximum possible dimension of the embeddings. The notation is kept generic for completeness. Also, note that assuming that K is fixed but potentially infinite in the data generating process, taking $K \to \infty$ gives a variation of the DPGMM. A toy representation of the model is given in Figure 1.

The likelihood associated with the embeddings obtained from a stochastic block model adjacency matrix $\mathbf{A}$ can be expressed as:

$$
L(\mathbf{X}|\mathcal{P}) = \prod_{i=1}^{n} \phi(\boldsymbol{x}_{id:}; \boldsymbol{v}_r, \boldsymbol{\Sigma}_r) \prod_{k=1}^{K} \phi(\boldsymbol{x}_{i:d}; \boldsymbol{v}_k, \boldsymbol{\Sigma}_k)^{\mathbb{1}_k(z_i)},
$$

where $\phi(\cdot)$ denotes the multivariate Gaussian density function and $\mathcal{P}$ is the set of all latent parameters and variables. Hence, the posterior is:

$$
p(\mathcal{P}|\mathbf{X}) \propto L(\mathbf{X}|\mathcal{P})p(\mathcal{P}) = L(\mathbf{X}|\mathcal{P}) \prod_{k=1}^{K} p(\boldsymbol{v}_k, \boldsymbol{\Sigma}_k|d) p(\boldsymbol{v}_r, \boldsymbol{\Sigma}_r|d) \prod_{i=1}^{n} p(z_i|\boldsymbol{\theta}) p(d)
$$

By standard methods for inference in a multivariate Gaussian mixture model with NIW prior, the covariance matrix $\boldsymbol{\Sigma}_k$ can be explicitly integrated out from the posterior to obtain

$$
p(\boldsymbol{v}_k|\mathbf{X}, \boldsymbol{v}_{-h}, \boldsymbol{z}, d) = t_{\nu_{n_k}}\left(\boldsymbol{v}_k \,\middle|\, \boldsymbol{v}_{n_k}, \frac{1}{\kappa_{n_k}\nu_{n_k}}\boldsymbol{\Delta}_{n_k}\right),
$$

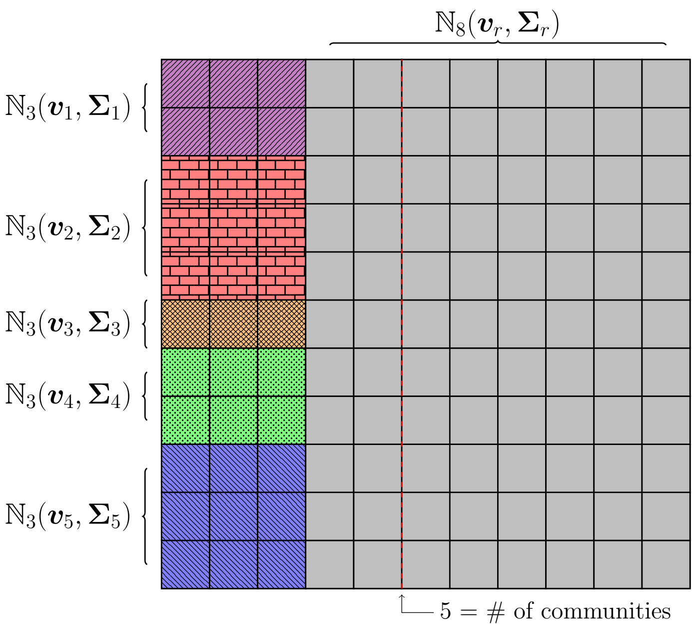

Figure 1: *Toy example for the generating process of the embeddings of a* 11*-nodes graph with* 5 *communities with latent dimension* $d = 3$*.*

where $t.(\cdot)$ is the multivariate Student t distribution. Also:

$$
\begin{aligned}
n_k &= \sum\nolimits_{i=1}^{n} \mathbb{1}_k\{z_i\},\\
\bar{\boldsymbol{x}}_{k:d} &= \sum\nolimits_{i:z_i=h} \boldsymbol{x}_{i:d}/n_k,\ \mathbf{S}_{k:d} = \sum\nolimits_{i:z_i=h} \boldsymbol{x}_{i:d}\boldsymbol{x}_{i:d}^{\top},\\
\boldsymbol{v}_{n_k:d} &= \frac{\kappa_0}{\kappa_0+n_k}\boldsymbol{v}_{0:d} + \frac{n_k}{\kappa_0+n_k}\bar{\boldsymbol{x}}_{k:d},\\
\kappa_{n_k} &= \kappa_0 + n_k,\\
\nu_{n_k} &= \nu_0 + n_k,\\
\boldsymbol{\Delta}_{n_k:d} &= \boldsymbol{\Delta}_{0:d} + \mathbf{S}_{k:d} + \kappa_0\boldsymbol{v}_{0:d}\boldsymbol{v}_{0:d}^{\top} - \kappa_{n_k}\boldsymbol{v}_{n_k:d}\boldsymbol{v}_{n_k:d}^{\top}.
\end{aligned}
$$

Henceforth, $\boldsymbol{v}_k$ can potentially be resampled in a simple Gibbs sampling step, conditional on the actual value of d and on the community allocations $\boldsymbol{z}$. Similar calculations allow to integrate out $\boldsymbol{\Sigma}_r$ and obtain the marginal posterior for $\boldsymbol{v}_r$ conditional on d. Note that it is possible to treat the components separately since $\boldsymbol{v}_{k:d}$ and $\boldsymbol{v}_{kd:}$ are independent, since these are both Gaussian with covariance $\mathbf{0}$. In this work, the location vectors $\boldsymbol{v}_k$ are collapsed out too, but the distribution is instructive to present other distributional results below.

In a multivariate Gaussian model with normal inverse Wishart prior, it is also possible to analytically express the marginal likelihood of the observed data. Here, conditioning on a community-specific Gaussian, on the assignments $\boldsymbol{z}$ and on the dimension of the latent space d:

$$
p(\mathcal{X}_{k:d}|d,\boldsymbol{z}) = \pi^{-n_k d/2}\frac{\kappa_0^{d/2}|\boldsymbol{\Delta}_{0:d}|^{(\nu_0+d-1)/2}}{\kappa_{n_k}^{d/2}|\boldsymbol{\Delta}_{n_k:d}|^{(\nu_{n_k}+d-1)/2}}\prod_{i=1}^{d}\frac{\Gamma\{(\nu_{n_k}+d-i)/2\}}{\Gamma\{(\nu_0+d-i)/2\}} \tag{3.1}
$$

where $\mathcal{X}_{k:d}$ is the subset of rows of $\mathbf{X}_{:d}$ such that $z_i = k$. Similar calculations give the full marginal likelihood for the remaining portion of the embeddings $\mathbf{X}_{d:}$. Note that the components are multivariate Gaussian with null covariance for each group, and therefore can be treated independently.

$$p(\mathbf{X}_{d:}|d, \boldsymbol{z}) = \pi^{-n(m-d)/2} \frac{\kappa_0^{(m-d)/2} |\boldsymbol{\Delta}_{0d:}|^{(\nu_0+m-d-1)/2}}{\kappa_n^{(m-d)/2} |\boldsymbol{\Delta}_{nd:}|^{(\nu_n+m-d-1)/2}} \prod_{i=1}^{m-d} \frac{\Gamma\{(\nu_n + m - d - i)/2\}}{\Gamma\{(\nu_0 + m - d - i)/2\}}.$$

Also, note that the probabilities $\boldsymbol{\theta}$ associated to the community assignment can be easily integrated out, resulting in the following marginal likelihood of community assignments, conditional on K:

$$p(\boldsymbol{z}|K) = \frac{\Gamma(\alpha) \prod_{k=1}^{K} \Gamma(n_k + \alpha/K)}{\Gamma(\alpha/K)^K \Gamma(n + \alpha)}. \tag{3.2}$$

Essentially, when the locations $\boldsymbol{v}_k$ and covariances $\boldsymbol{\Sigma}_k$ are collapsed out, four types of moves are available:

a. Propose a change in the community allocations $\boldsymbol{z}$,
b. Propose a change in the latent dimension d,
c. Propose to split (or merge) two communities (hence change K),
d. Propose to create (or remove) an empty community (hence change K).

3.1 Change in the community allocations

It immediately follows that a fully collapsed Gibbs update for each community assignment is available:

$$p(z_i = k|\boldsymbol{z}_{-i}, \mathbf{X}, d, K) \propto p(z_i = k|\boldsymbol{z}_{-i}, K) p(\boldsymbol{x}_i|\{\boldsymbol{x}_j\}_{j\neq i:z_j=k}, z_i = k, d, K) \tag{3.3}$$

Clearly, from (3.2):

$$p(z_i = k|\boldsymbol{z}_{-i}, K) = \frac{p(z_i = k, \boldsymbol{z}_{-i}|K)}{p(\boldsymbol{z}_{-i}|K)} = \frac{n_k^{-i} + \alpha/K}{n - 1 + \alpha}.$$

Also, $p(\boldsymbol{x}_i|\{\boldsymbol{x}_j\}_{j\neq i:z_j=k}, z_i = k, d, K)$ can be obtained as the ratio of marginal likelihoods

$$p(\boldsymbol{x}_i|\{\boldsymbol{x}_j\}_{j\neq i:z_j=k}, z_i = k, d, K) = \frac{p(\boldsymbol{x}_i, \{\boldsymbol{x}_j\}_{j\neq i:z_j=k}|z_i = k, d, K)}{p(\{\boldsymbol{x}_j\}_{j\neq i:z_j=k}|d, K)}. \tag{3.4}$$

Using (3.1), the exact expression for the ratio in (3.4) is:

$$p(\boldsymbol{x}_i|\{\boldsymbol{x}_j\}_{j\neq i:z_j=k}, z_i = k, d, K) = \frac{(\kappa_{n_k^{-i}} + 1)^{-d/2} |\boldsymbol{\Delta}_{\{n_k^{-i}+i\}}|^{-(\nu_{n_k^{-i}}+d)/2} \Gamma\{(\nu_{n_k^{-i}} + d)/2\}}{\pi^{d/2} \kappa_{n_k^{-i}}^{-d/2} |\boldsymbol{\Delta}_{n_k^{-i}}|^{-(\nu_{n_k^{-i}}+d-1)/2} \Gamma\{\nu_{n_k^{-i}}/2\}},$$

which is equivalent to the following multivariate Student t distribution:

$$p(\boldsymbol{x}_i|\{\boldsymbol{x}_j\}_{j\neq i:z_j=k}, z_i = k, d, K) = t_{\nu_{n_k^{-i}}} \left(\boldsymbol{x}_{i:d} \middle| \boldsymbol{v}_{n_k^{-i}:d}, \frac{\kappa_{n_k^{-i}} + 1}{\kappa_{n_k^{-i}} \nu_{n_k^{-i}}} \boldsymbol{\Delta}_{n_k^{-i}:d} \right).$$

3.2 Change in the latent dimension

Given a sampled value of d, birth-death moves of the type $d^\star \in \{d-1, d+1\}$ can be proposed with different probabilities $q(d^\star|d)$. Since the fully collapsed sampler is used, there is no need to also propose values for $\boldsymbol{v}_k$ and $\boldsymbol{\Sigma}_k$ together with d, and therefore the acceptance ratio reduces to the likelihood ratio of marginal likelihoods for the two different values of d, conditional on the allocations $\boldsymbol{z}$ and K, times their prior ratio for the proposed values, weighted by the inverse of the proposal ratio:

$$\alpha(d^\star|d) = \min\left\{1, \frac{p(\mathbf{X}|d^\star, K, \boldsymbol{z})p(d^\star)}{p(\mathbf{X}|d, K, \boldsymbol{z})p(d)} \frac{q(d|d^\star)}{q(d^\star|d)}\right\}.$$

Note that here $p(\mathbf{X}|d, K, \boldsymbol{z})$ reduces to a product of marginal likelihoods for the data points assigned to each of the K components, limited to the first d dimensions, times the marginal likelihood for the entire embedding for the last $m-d$ dimensions:

$$p(\mathbf{X}|d, K, \boldsymbol{z}) = p(\mathbf{X}_{d:}|d) \prod_{k=1}^{K} p(\mathcal{X}_{k:d}|d).$$

where $\mathcal{X}_{k:d}$ is the subset of the rows of $\mathbf{X}_{:d}$ such that $z_i = k$. Using (3.1), for $d^\star = d+1$:

$$\frac{p(\mathbf{X}|d+1, K, \boldsymbol{z})}{p(\mathbf{X}|d, K, \boldsymbol{z})} = \frac{\Gamma\{(\nu_0+m-d-1)/2\}\kappa_n^{1/2}\kappa_0^{(K-1)/2}}{\Gamma\{(\nu_n+m-d-1)/2\}\Gamma\{\nu_0+d)/2\}^K} \frac{|\boldsymbol{\Delta}_{0(d+1):}|^{(\nu_0+m-d-2)/2}}{|\boldsymbol{\Delta}_{0d:}|^{(\nu_0+m-d-1)/2}}$$
$$\frac{|\boldsymbol{\Delta}_{nd:}|^{(\nu_n+m-d-1)/2}}{|\boldsymbol{\Delta}_{n(d+1):}|^{(\nu_n+m-d-2)/2}} \frac{|\boldsymbol{\Delta}_{0:(d+1)}|^{K(\nu_0+d)/2}}{|\boldsymbol{\Delta}_{0:d}|^{K(\nu_0+d-1)/2}} \prod_{k=1}^{K} \frac{\Gamma\{(\nu_{n_k}+d)/2\}}{\kappa_{n_k}^{1/2}} \frac{|\boldsymbol{\Delta}_{n_k:d}|^{(\nu_{n_k}+d-1)/2}}{|\boldsymbol{\Delta}_{n_k:(d+1)}|^{(\nu_{n_k}+d)/2}}.$$

3.3 Split (or merge) two communities

A new community configuration must be proposed for this type of move, and a sampler inspired to Sequentially-Allocated Merge-Split sampling (Dahl, 2003) is used here. Two indices i and j are sampled at random from the n nodes. If $z_i = z_j = k$ is proposed, z_i is allocated to k, and z_j to $K+1$. The nodes in $\mathcal{C}_{k\setminus\{i,j\}} = \{h \in \{1,\dots,n\} \setminus \{i,j\} : z_h = k\}$ are randomly permuted and allocated to clusters k or $K+1$ sequentially with probability proportional to the respective sequential posterior predictive probability, restricted to the two clusters. Otherwise, if $z_i \neq z_j$, which means that i and j belong to different clusters, then the two clusters are merged together. The acceptance probability becomes:

$$\alpha(K^\star, \boldsymbol{z}^\star|K, \boldsymbol{z}) = \min\left\{1, \frac{p(\mathbf{X}_{:d}|d, K^\star, \boldsymbol{z}^\star)p(\boldsymbol{z}^\star|K^\star)p(K^\star)}{p(\mathbf{X}_{:d}|d, K, \boldsymbol{z})p(\boldsymbol{z}|K)p(K)} \frac{q(K, \boldsymbol{z}|K^\star, \boldsymbol{z}^\star)}{q(K^\star, \boldsymbol{z}^\star|K, \boldsymbol{z})}\right\}.$$

Importantly, $q(K^\star, \boldsymbol{z}^\star|K, \boldsymbol{z})$ is evaluated jointly up to a proportionality constant that cancels out in the ratio. In particular:

Split move (or inverse merge)	$q(K+1, \boldsymbol{z}^\star|K, \boldsymbol{z}) \propto \prod_{u\in\sigma(\mathcal{C}_{k\setminus\{i,j\}})} \pi_{z_u^\star},$
Inverse split move (or merge)	$q(K, \boldsymbol{z}|K+1, \boldsymbol{z}^\star) \propto 1.$

where, in a very compact notation, $\sigma(\mathcal{C}_{k\setminus\{i,j\}})$ is a random permutation of the indices in $\mathcal{C}_{k\setminus\{i,j\}}$, and $\pi_{z_u^\star}$ is the normalised probability obtained from the sequential posterior predictive, essentially obtained using a sequential version of (3.3), restricted to two groups. Also note that when the acceptance ratio involves a merge move, then it is necessary to calculate an "imaginary" split using, crucially, a *random* permutation of the indices.

3.4 Create (or remove) an empty community

The birth-death moves for empty communities do not change the likelihood for the observed data, hence the acceptance probability reduces to the ratio of marginal likelihoods for group assignments under two different settings:

$$\alpha(K^\star|K) = \min\left\{1, \frac{p(\boldsymbol{z}|K^\star)p(K^\star)}{p(\boldsymbol{z}|K)p(K)}\frac{q(K|K^\star)}{q(K^\star|K)}\right\}.$$

References

Dahl, D.B. (2003). *An improved merge-split sampler for conjugate Dirichlet process mixture models.* Tech. rep. 1086. Department of Statistics, University of Wisconsin, Madison.

Rubin-Delanchy, P. et al. (2017). “A statistical interpretation of spectral embedding: the generalised random dot product graph”. In: *ArXiv e-prints.* arXiv: 1709.05506 [stat.ML].