Supplementary Material for: Bayesian nonparametric modeling of heterogeneous populations of networks

Francesco Barile*, Simón Lunagómez† and Bernardo Nipoti*

The Supplementary Material is organized as follows. In Section 1 we provide the proof of the results in Section 2.3 of the main paper. Section 2 provides additional details on posterior computations, including strategies to enhance the algorithm's efficiency. Section 3 provides a closed-form expression for the cluster-specific one-step-ahead posterior predictive distribution, conditionally on an estimated partition of the sample. This result is generalized to make a joint prediction on m graphs, assuming that they all belong to a specific cluster of the estimated partition. In addition, the cluster-specific posterior distribution of $\mathcal{C}_k^*$ is also provided in closed-form. Finally, Section 4, Section 5 and Section 6 present further information on the simulation studies and the illustrations in Section 4 Section 5 and Section 6 of the main paper.

1 Proofs of Theorem 2.1 and Corollary 2.1

Proof of Theorem 2.1. $\mathscr{G}_{\mathcal{V}}$ consists of $|\mathscr{G}_{\mathcal{V}}| = 2^M$ possible network configurations, with $M = \binom{N}{2}$. We can then name the elements of $\mathscr{G}_{\mathcal{V}}$ as $\mathscr{G}_{\mathcal{V}} = \{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_{2^M}\}$. We observe that any probability mass function $p_* \in \mathcal{P}_{\mathscr{G}_{\mathcal{V}}}$ is characterized by a set of 2^M weights $p_{0l} = p_*(\mathcal{H}_l)$, with $l = 1, \ldots, 2^M$, as

$$p_*(\cdot) = \sum_{l=1}^{2^M} p_{0l}\delta_{\mathcal{H}_l}(\cdot).$$

By exploiting the stick-breaking representation of the DP (Sethuraman, 1994), we can rewrite $\tilde{f}$ as

$$\tilde{f}(\cdot) = \sum_{j=1}^{\infty} \tilde{p}_j \psi(\cdot; \tilde{\vartheta}_j),$$

where $\tilde{\vartheta}_j = (\tilde{\mathcal{C}}_j, \tilde{\alpha}_j) \overset{\text{iid}}{\sim} P_0$ and the $\tilde{p}_j$'s are positive weights with Griffiths-Engen-McCloskey distribution with parameter c (see, e.g., Ewens, 1990), such that $\sum_{j=1}^{\infty} \tilde{p}_j = 1$ almost surely. We observe that, as done for p_*, also $\tilde{f}$ can be written as a finite sum, that is

$$\tilde{f}(\cdot) = \sum_{l=1}^{2^M} \tilde{q}_l \delta_{\mathcal{H}_l}(\cdot).$$

arXiv: 2410.10354

*Department of Economics, Management and Statistics, University of Milano-Bicocca, Italy, francesco.barile@unimib.it; bernardo.nipoti@unimib.it

†Instituto Tecnológico Autónomo de México, CDMX 01080, Mexico, simon.lunagomez@itam.mx

For any $\omega \in \Omega$, we henceforth use the superscript (ω) to denote a realization of a random variable, e.g. $\tilde{f}^{(\omega)}$.

The remainder of the proof is organized as follows. For any $p_* \in \mathcal{P}_{\mathscr{G}_N}$ and any $\epsilon > 0$:

Part 1. We define a set of conditions and show that, if a realization $\tilde{f}^{(\omega)}$ of $\tilde{f}$ satisfies these conditions, then $\tilde{f}^{(\omega)}$ belongs to $\mathbb{B}_\epsilon(p_*)$, the Kullback–Leibler neighborhood of p_* of radius ϵ. That is, $\mathrm{KL}(p_*; \tilde{f}^{(\omega)}) \leq \epsilon$.

Part 2. We define the events $A_1, A_2 \subseteq \mathcal{P}_{\mathscr{G}_N}$ as

$$A_1 = \{\tilde{f}^{(\omega)} \text{ with } \omega \in \Omega : \texttt{b1}, \texttt{b3} \text{ hold for } (\alpha_\star, \eta_\star) \text{ satisfying } \texttt{a1}, \texttt{a2}\},$$
$$A_2 = \{\tilde{f}^{(\omega)} \text{ with } \omega \in \Omega : \texttt{b2} \text{ holds for } (\alpha_\star, \eta_\star) \text{ satisfying } \texttt{a1}, \texttt{a2}\},$$

and, by exploiting the result in Part 1, we show that

$$\mathbb{B}_\epsilon(p_*) \supseteq A_1 \cap A_2.$$

Part 3. We show that Π assigns positive probability to $A_1 \cap A_2$ and, given the result in Part 2, to $\mathbb{B}_\epsilon(p_*)$.

Part 1. For any $\varepsilon > 0$, we consider $(\alpha_\star, \eta_\star)$ such that

- a1) $0 < \alpha_\star < 1 - \exp\{-\varepsilon/M\}$;
- a2) $0 < \eta_\star < 1 - \exp\{-\varepsilon\}/(1-\alpha_\star)^M$.

Given $p_* \in \mathcal{P}_{\mathscr{G}_N}$, we let $\tilde{f}^{(\omega)}$ be such that, for any $l = 1, \ldots, 2^M$,

- b1) $\tilde{\mathcal{C}}_l^{(\omega)} = \mathcal{H}_l$;
- b2) $\tilde{p}_l^{(\omega)} \in [p_{0l}(1-\eta_\star), p_{0l}]$;
- b3) $\tilde{\alpha}_l^{(\omega)} \in (0, \alpha_\star]$.

We observe that condition a1 guarantees that the set of solutions $\{\eta_\star : \texttt{a2} \text{ holds}\}$ is not empty. We next show that if $\tilde{f}^{(\omega)}$ satisfies b1, b2, and b3, then $\mathrm{KL}(p_*; \tilde{f}^{(\omega)}) \leq \varepsilon$. We first observe that, for any $l = 1, \ldots, 2^M$,

$$\tilde{q}_l^{(\omega)} \overset{\texttt{b1}}{\geq} \tilde{p}_l^{(\omega)}(1 - \tilde{\alpha}_l^{(\omega)})^M. \tag{1}$$

Then, it follows that

$$\mathrm{KL}(p_*; \tilde{f}^{(\omega)}) = \sum_{l=1}^{2^M} p_{0l} \log\left(\frac{p_{0l}}{\tilde{q}_l^{(\omega)}}\right)$$

$$\begin{aligned}
&\overset{(1)}{\leq} \sum_{l=1}^{2^M} p_{0l} \log\left(\frac{p_{0l}}{\tilde{p}_l^{(\omega)}(1-\tilde{\alpha}_l^{(\omega)})^M}\right)\\
&\overset{\texttt{b2}}{\leq} \sum_{l=1}^{2^M} p_{0l} \log\left(\frac{1}{(1-\eta_\star)(1-\tilde{\alpha}_l^{(\omega)})^M}\right)\\
&\overset{\texttt{b3}}{\leq} \sum_{l=1}^{2^M} p_{0l} \log\left(\frac{1}{(1-\eta_\star)(1-\alpha_\star)^M}\right)\\
&= \log\left(\frac{1}{(1-\eta_\star)(1-\alpha_\star)^M}\right) \overset{\texttt{a2}}{\leq} \varepsilon.
\end{aligned}$$

Part 2. We observe that

$$\begin{aligned}
\mathbb{B}_\varepsilon(p_*) \supseteq &\{\tilde{f}^{(\omega)} \text{ with } \omega\in\Omega : \texttt{b1}, \texttt{b2}, \texttt{b3} \text{ hold for } (\alpha_\star,\eta_\star) \text{ satisfying } \texttt{a1}, \texttt{a2}\}\\
=&\{\tilde{f}^{(\omega)} \text{ with } \omega\in\Omega : \texttt{b1}, \texttt{b3} \text{ hold for } (\alpha_\star,\eta_\star) \text{ satisfying } \texttt{a1}, \texttt{a2}\}\\
&\cap \{\tilde{f}^{(\omega)} \text{ with } \omega\in\Omega : \texttt{b2} \text{ holds for } (\alpha_\star,\eta_\star) \text{ satisfying } \texttt{a1}, \texttt{a2}\} = A_1\cap A_2.
\end{aligned}$$

Part 3. The proof is completed by showing that, for any $\varepsilon > 0$, Π assigns positive probability to $\mathbb{B}_\varepsilon(p_*)$. Given the independence of weights $(\tilde{p}_j)_{j\geq 1}$ and atoms $(\tilde{\vartheta}_j)_{j\geq 1}$ in the definition of $\tilde{f}$, the events A_1 and A_2 are disjoint and thus $\Pi(A_1\cap A_2) = \Pi(A_1)\Pi(A_2)$. To prove that $\Pi(\mathbb{B}_\varepsilon(p_*)) > 0$, it then suffices to check that both $\Pi(A_1)$ and $\Pi(A_2)$ are positive. $\Pi(A_1) > 0$ follows from the fact that P_0 has full support on Θ. Moreover, $\Pi(A_2) > 0$ as, for any $j = 1, 2, \ldots$, the distribution of $\tilde{p}_j$ for the DP has full support on $[0, 1 - \sum_{i=1}^{j-1}\tilde{p}_i]$. □

Proof of Corollary 2.1. A direct application of Example 6.21 in Ghosal and van der Vaart (2017). □

2 Derivation of posterior computations

2.1 Probability of a new value in the generalized Pólya urn scheme

We provide a detailed derivation of Equation 6 from the main paper, which gives the probability π_{l0} of sampling a new pair $\vartheta_l = (\mathcal{C}_l, \alpha_l)$ in the generalized Pólya urn scheme. Specifically,

$$\begin{aligned}
\pi_{l0} &\propto c \int \psi\left(\mathcal{G}_l; \vartheta_l\right) \mathrm{d}P_0(\vartheta_l)\\
&\propto c \int_0^{1/2} \sum_{\mathcal{C}_l\in\mathscr{G}_\mathcal{V}} \alpha_l^{d_\mathrm{H}(\mathcal{G}_l,\mathcal{C}_l)}(1-\alpha_l)^{M-d_\mathrm{H}(\mathcal{G}_l,\mathcal{C}_l)} p_{\mathrm{CER}}\left(\mathcal{C}_l; \mathcal{G}_0, \alpha_l\right) f_{\mathrm{TBeta}}(\alpha_l; 1/2, a, b)\mathrm{d}\alpha_l\\
&\propto c \int_0^{1/2} f_{\mathrm{TBeta}}(\alpha_l; 1/2, a, b)\times \qquad (2)
\end{aligned}$$

$$\times \sum_{\mathcal{C}_l \in \mathscr{G}_\mathcal{V}} \alpha_l^{d_\mathrm{H}(\mathcal{G}_l,\mathcal{C}_l)}(1-\alpha_l)^{M-d_\mathrm{H}(\mathcal{G}_l,\mathcal{C}_l)}\alpha^{d_\mathrm{H}(\mathcal{C}_l,\mathcal{G}_0)}(1-\alpha_l)^{M-d_\mathrm{H}(\mathcal{C}_l,\mathcal{G}_0)}\,\mathrm{d}\alpha_l$$

$$\propto c\int_0^{1/2} f_{\mathrm{TBeta}}(\alpha_l;1/2,a,b)\times \tag{3}$$

$$\times \sum_{\mathcal{C}_l \in \mathscr{G}_\mathcal{V}} \alpha_l^{d_\mathrm{H}(\mathcal{G}_l,\mathcal{C}_l)+d_\mathrm{H}(\mathcal{C}_l,\mathcal{G}_0)}(1-\alpha)^{2M-[d_\mathrm{H}(\mathcal{G}_l,\mathcal{C}_l)+d_\mathrm{H}(\mathcal{C}_l,\mathcal{G}_0)]}\,\mathrm{d}\alpha_l$$

$$\propto c\int_0^{1/2} f_{\mathrm{TBeta}}(\alpha_l;1/2,a,b)(1-\alpha_l)^{2M}\sum_{\mathcal{C}_l \in \mathscr{G}_\mathcal{V}}\left(\frac{\alpha_l}{1-\alpha_l}\right)^{d_\mathrm{H}(\mathcal{G}_l,\mathcal{C}_l)+d_\mathrm{H}(\mathcal{C}_l,\mathcal{G}_0)}\mathrm{d}\alpha_l. \tag{4}$$

We note that

$$\sum_{\mathcal{C}_l \in \mathscr{G}_\mathcal{V}}\left(\frac{\alpha_l}{1-\alpha_l}\right)^{d_\mathrm{H}(\mathcal{G}_l,\mathcal{C}_l)+d_\mathrm{H}(\mathcal{C}_l,\mathcal{G}_0)} = \sum_{h=0}^{2M} w_{lh}\left(\frac{\alpha_l}{1-\alpha_l}\right)^h, \tag{5}$$

where w_{lh} determines how many graphs $\mathcal{C}_l \in \mathscr{G}_\mathcal{V}$ are such that $d_\mathrm{H}(\mathcal{G}_l,\mathcal{C}_l)+d_\mathrm{H}(\mathcal{C}_l,\mathcal{G}_0) = h$, for $h = 0,1,\ldots,2M$, given that $d_\mathrm{H}(\mathcal{G}_0,\mathcal{G}_l) = d_l$. Conveniently, w_{lh} coincides with

$$w_{lh} = \begin{cases} 0 & \text{if } h < d_l \\ 0 & \text{if } h \geq d_l \text{ and } h-d_l \text{ is odd} \\ 2^{d_l}\binom{M-d_l}{\frac{h-d_l}{2}} & \text{if } h \geq d_l \text{ and } h-d_l \text{ is even.} \end{cases} \tag{6}$$

We observe that when $h-d_l$ is even, then $d_l \leq h \leq 2M-d_l$. Therefore, armed with (6) and setting $r = (h-d_l)/2$, the left hand side of Equation 5 can be written as

$$\begin{aligned}\sum_{\mathcal{C}_l \in \mathscr{G}_\mathcal{V}}\left(\frac{\alpha_l}{1-\alpha_l}\right)^{d_\mathrm{H}(\mathcal{G}_l,\mathcal{C}_l)+d_\mathrm{H}(\mathcal{C}_l,\mathcal{G}_0)} &= \sum_{h=d_l}^{2M-d_l} w_{lh}\left(\frac{\alpha_l}{1-\alpha_l}\right)^h \\ &= \sum_{r=0}^{(M-d_l)} w_{l(2r+d_l)}\left(\frac{\alpha_l}{1-\alpha_l}\right)^{2r+d_l} \\ &= \sum_{r=0}^{(M-d_l)} 2^{d_l}\binom{M-d_l}{r}\left(\frac{\alpha_l}{1-\alpha_l}\right)^{2r+d_l}. \end{aligned} \tag{7}$$

In turn, exploiting (7), Equation 2 can be rewritten as

$$\begin{aligned}\pi_{l0} &\propto c\int_0^{1/2} f_{\mathrm{TBeta}}(\alpha_l;1/2,a,b)(1-\alpha_l)^{2M}\sum_{\mathcal{C}_l \in \mathscr{G}_\mathcal{V}}\left(\frac{\alpha_l}{1-\alpha_l}\right)^{d_\mathrm{H}(\mathcal{G}_l,\mathcal{C}_l)+d_\mathrm{H}(\mathcal{C}_l,\mathcal{G}_0)}\mathrm{d}\alpha_l \\ &\propto c\int_0^{1/2} f_{\mathrm{TBeta}}(\alpha_l;1/2,a,b)(1-\alpha_l)^{2M}\sum_{r=0}^{(M-d_l)} 2^{d_l}\binom{M-d_l}{r}\left(\frac{\alpha_l}{1-\alpha_l}\right)^{2r+d_l}\mathrm{d}\alpha_l \\ &\propto c\sum_{r=0}^{(M-d_l)} 2^{d_l}\binom{M-d_l}{r}\int_0^{1/2}\frac{\alpha_l^{a-1}(1-\alpha_l)^{b-1}}{\mathcal{B}(1/2;a,b)}(1-\alpha_l)^{2M}\left(\frac{\alpha_l}{1-\alpha_l}\right)^{2r+d_l}\mathrm{d}\alpha_l \end{aligned}$$

$$\propto c \sum_{r=0}^{(M-d_l)} 2^{d_l} \binom{M-d_l}{r} \int_0^{1/2} \frac{\alpha_l^{a+2r+d_l-1}(1-\alpha_l)^{b+2(M-r)-d_l-1}}{\mathcal{B}(1/2;a,b)} \,\mathrm{d}\alpha_l$$
$$\propto c \sum_{r=0}^{(M-d_l)} w_{lr} \frac{\mathcal{B}(1/2;a_{lr},b_{lr})}{\mathcal{B}(1/2;a,b)}, \tag{8}$$

where $w_{lr} = 2^{d_l}\binom{M-d_l}{r}$, $a_{lr} = a + 2r + d_l$ and $b_{lr} = b + 2M - 2r - d_l$.

2.2 Reshuffling step

We provide an explicit derivation of the characterization of the full conditional distribution for each ϑ_k^*, with $k = 1, \ldots, K$, given in Equations 11, 12 and 13 of the main manuscript. We start by recalling that we let $\mathcal{D}_k = \{l \in \{1, \ldots, n\} : \vartheta_l = \vartheta_k^*\}$ denote the index set of observations belonging to the k-th cluster, with $|\mathcal{D}_k| = n_k$, and define $\mathcal{D}_k^\dagger = \mathcal{D}_k \cup \{0\}$. For any index set $\mathcal{D} \subseteq \{0, 1, \ldots, n\}$, we denote $\mathcal{G}^{(\mathcal{D})} = \{\mathcal{G}_l : l \in \mathcal{D}\}$, and we let $n_{ij}^{(k)} = \sum_{l \in \mathcal{D}_k^\dagger} A_{\mathcal{G}_l[ij]}$ denote the number of graphs in $\mathcal{G}^{(\mathcal{D}_k^\dagger)}$ that present an edge connecting the nodes $\{i, j\}$. Finally, throughout this section, we use $p(x)$ to denote the distribution of x and $p(x \mid y)$ to denote the conditional distribution of x given y.
The full conditional distribution of ϑ_k^* can be factorized as:

$$p(\vartheta_k^* \mid \mathcal{G}^{(\mathcal{D}_k)}) = p(\alpha_k^* \mid \mathcal{G}^{(\mathcal{D}_k)}) p(\mathcal{C}_k^* \mid \alpha_k^*, \mathcal{G}^{(\mathcal{D}_k)}). \tag{9}$$

To study the distributions appearing in the right-hand side of (9), it is instructive to start from the joint distribution $p(\alpha_k^*, \mathcal{C}_k^*, \mathcal{G}^{(\mathcal{D}_k)})$. Namely,

$$\begin{aligned}
p(\alpha_k^*, \mathcal{C}_k^*, \mathcal{G}^{(\mathcal{D}_k)}) &= p(\alpha_k^*, \mathcal{C}_k^*) p(\mathcal{G}^{(\mathcal{D}_k)} \mid \alpha_k^*, \mathcal{C}_k^*) \\
&= p(\alpha_k^*) p(\mathcal{C}_k^* \mid \alpha_k^*) \prod_{l \in \mathcal{D}_k} p(\mathcal{G}_l \mid \alpha_k^*, \mathcal{C}_k^*) \\
&= p(\alpha_k^*) p_{\mathrm{CER}}(\mathcal{C}_k^*; \mathcal{G}_0, \alpha_k^*) \prod_{l \in \mathcal{D}_k} p_{\mathrm{CER}}(\mathcal{G}_l; \mathcal{C}_k^*, \alpha_k^*) \\
&= p(\alpha_k^*) {\alpha_k^*}^{d_{\mathrm{H}}(\mathcal{G}_0, \mathcal{C}_k^*)} (1-\alpha_k^*)^{M - d_{\mathrm{H}}(\mathcal{G}_0, \mathcal{C}_k^*)} \prod_{l \in \mathcal{D}_k} {\alpha_k^*}^{d_{\mathrm{H}}(\mathcal{G}_l, \mathcal{C}_k^*)} (1-\alpha_k^*)^{M - d_{\mathrm{H}}(\mathcal{G}_l, \mathcal{C}_k^*)} \\
&= p(\alpha_k^*) \prod_{l \in \mathcal{D}_k^\dagger} {\alpha_k^*}^{d_{\mathrm{H}}(\mathcal{G}_l, \mathcal{C}_k^*)} (1-\alpha_k^*)^{M - d_{\mathrm{H}}(\mathcal{G}_l, \mathcal{C}_k^*)} \\
&= p(\alpha_k^*) (1-\alpha_k^*)^{(n_k+1)M} \left(\frac{\alpha_k^*}{1-\alpha_k^*} \right)^{\sum_{l \in \mathcal{D}_k^\dagger} d_{\mathrm{H}}(\mathcal{G}_l, \mathcal{C}_k^*)}.
\end{aligned} \tag{10}$$

We next focus on the first distribution in the factorization on the right-hand side of (9).

$$\begin{aligned}
p(\alpha_k^*, \mathcal{G}^{(\mathcal{D}_k)}) &= \sum_{\mathcal{C}_k^* \in \mathscr{G}_\mathcal{V}} p(\alpha_k^*, \mathcal{C}_k^*, \mathcal{G}^{(\mathcal{D}_k)}) \\
&= p(\alpha_k^*) (1-\alpha_k^*)^{(n_k+1)M} \sum_{\mathcal{C}_k^* \in \mathscr{G}_\mathcal{V}} \left(\frac{\alpha_k^*}{1-\alpha_k^*} \right)^{\sum_{l \in \mathcal{D}_k^\dagger} d_{\mathrm{H}}(\mathcal{G}_l, \mathcal{C}_k^*)}.
\end{aligned} \tag{11}$$

We let $U_k = \sum_{l \in \mathcal{D}_k^\dagger} d_{\mathrm{H}}\left(\mathcal{G}_l, \mathcal{C}_k^*\right)$ and note that

$$U_k = \sum_{l \in \mathcal{D}_k^\dagger} d_{\mathrm{H}}\left(\mathcal{G}_l, \mathcal{C}_k^*\right) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[(n_k + 1 - n_{ij}^{(k)}) A_{\mathcal{C}_k^*[ij]} + n_{ij}^{(k)} (1 - A_{\mathcal{C}_k^*[ij]}) \right]$$
$$= \sum_{i<j} \left[(n_k + 1 - 2n_{ij}^{(k)}) A_{\mathcal{C}_k^*[ij]} + n_{ij}^{(k)} \right], \tag{12}$$

has support $u_k \in \{d_k^*, d_k^* + 1, \ldots, D_k^* - 1, D_k^*\}$, where

$$d_k^* = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \min\{n_{ij}^{(k)}, n_k + 1 - n_{ij}^{(k)}\}, \quad D_k^* = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \max\{n_{ij}^{(k)}, n_k + 1 - n_{ij}^{(k)}\}.$$

We let $m_{kh} = \# \left\{ \{i,j\} \in \mathcal{V}^2 : n_{ij}^{(k)} = h \right\}$ denote the number of pairs of distinct nodes that are connected by an edge in exactly h graphs in $\mathcal{G}^{(\mathcal{D}_k^\dagger)}$ and define $M_{kh} = m_{kh} + m_{k(n_k+1-h)}$. In addition, we let $\gamma_{kh}(s_h) = (n_k + 1 - 2h)\, s_h + h M_{kh}$. At this stage, it is worth noticing that (11) involves a polynomial in the variable $x_k = \alpha_k^*/(1-\alpha_k^*)$ of the form

$$\mathcal{P}(x_k) = \sum_{\mathcal{C}_k^* \in \mathscr{G}_\mathcal{V}} x_k^{\sum_{l \in \mathcal{D}_k^\dagger} d_{\mathrm{H}}(\mathcal{G}_l, \mathcal{C}_k^*)} = \sum_{\mathcal{C}_k^* \in \mathscr{G}_\mathcal{V}} x_k^{\sum_{i<j} \left[(n_k + 1 - 2n_{ij}^{(k)}) A_{\mathcal{C}_k^*[ij]} + n_{ij}^{(k)} \right]}$$
$$= \sum_{A_{\mathcal{C}_k^*[12]} \in \{0,1\}} x_k^{(n_k+1-2n_{12}^{(k)}) A_{\mathcal{C}_k^*[12]} + n_{12}^{(k)}} \times \cdots \times$$
$$\sum_{A_{\mathcal{C}_k^*[(N-1)N]} \in \{0,1\}} x_k^{(n_k+1-2n_{(N-1)N}^{(k)}) A_{\mathcal{C}_k^*[(N-1)N]} + n_{(N-1)N}^{(k)}}$$
$$= \left(x_k^{n_{12}^{(k)}} + x_k^{n_k+1-n_{12}^{(k)}} \right) \times \cdots \times \left(x_k^{n_{(N-1)N}^{(k)}} + x_k^{n_k+1-n_{(N-1)N}^{(k)}} \right)$$
$$= \prod_{i=1}^{N-1} \prod_{j=i+1}^{N} \left(x_k^{n_{ij}^{(k)}} + x_k^{n_k+1-n_{ij}^{(k)}} \right) = \prod_{h=0}^{n_k+1} \left(x_k^h + x_k^{n_k+1-h} \right)^{m_{kh}}$$
$$= \begin{cases} \prod_{h=0}^{n_k/2} \sum_{s_h=0}^{M_{kh}} \binom{M_{kh}}{s_h} x_k^{\gamma_{kh}(s_h)} & \text{if } n_k \text{ is even} \\ \left(2x_k^{\lfloor n_k/2 \rfloor + 1} \right)^{m_{k(\lfloor n_k/2 \rfloor + 1)}} \prod_{h=0}^{\lfloor n_k/2 \rfloor} \sum_{s_h=0}^{M_{kh}} \binom{M_{kh}}{s_h} x_k^{\gamma_{kh}(s_h)} & \text{if } n_k \text{ is odd} \end{cases}$$
$$= \sum_{u=d_k^*}^{D_k^*} w_{k(u-d_k^*)}^* x_k^u = \sum_{r=0}^{D_k^* - d_k^*} w_{kr}^* x_k^{d_k^* + r}, \tag{13}$$

where $\lfloor z \rfloor$ denotes the integer part of z. The polynomial $\mathcal{P}(x_k)$ represents a generating-function with coefficients of the form

$$w_{kr}^* = \begin{cases} \sum_{\mathcal{S}_{kr}} \prod_{h=0}^{n_k/2} \binom{M_{kh}}{s_h} & \text{if } n_k \text{ is even} \\ \sum_{\mathcal{R}_{kr}} 2^{m_{k(\lfloor n_k/2 \rfloor + 1)}} \prod_{h=0}^{\lfloor n_k/2 \rfloor} \binom{M_{kh}}{s_h} & \text{if } n_k \text{ is odd,} \end{cases} \tag{14}$$

where the sums in (14) are taken over the sets

$$\mathcal{S}_{kr} = \Big\{(s_0, \dots, s_{n_k/2}) : s_h \in \{0, \dots, M_{kh}\}\, \forall h,\ \sum_{h=0}^{n_k/2} \gamma_{kh}(s_h) - d_k^* = r\Big\},$$

$$\mathcal{R}_{kr} = \Big\{(s_0, \dots, s_{\lfloor n_k/2 \rfloor}) : s_h \in \{0, \dots, M_{kh}\}\, \forall h,\ \sum_{h=0}^{\lfloor n_k/2 \rfloor} \gamma_{kh}(s_h) + (\lfloor n_k/2 \rfloor + 1) m_{k(\lfloor n_k/2 \rfloor + 1)} - d_k^* = r\Big\}.$$

Thanks to (13) and (14), and recalling that each α_k^* is distributed a priori as a Truncated-Beta on $(0, 1/2)$ with parameters $a, b > 0$, (11) can be written as

$$\begin{aligned} p(\alpha_k^*, \mathcal{G}^{(\mathcal{D}_k)}) &= p(\alpha_k^*)(1-\alpha_k^*)^{(n_k+1)M} \sum_{\mathcal{C}_k^* \in \mathscr{G}_\mathcal{V}} \left(\frac{\alpha_k^*}{1-\alpha_k^*}\right)^{\sum_{l \in \mathcal{D}_k^\dagger} d_{\mathrm{H}}(\mathcal{G}_l, \mathcal{C}_k^*)} \\ &= p(\alpha_k^*)(1-\alpha_k^*)^{(n_k+1)M} \sum_{r=0}^{D_k^* - d_k^*} w_{kr}^* \left(\frac{\alpha_k^*}{1-\alpha_k^*}\right)^{d_k^* + r} \\ &= \frac{1}{\mathcal{B}(1/2; a, b)} \sum_{r=0}^{D_k^* - d_k^*} w_{kr}^* {\alpha_k^*}^{a + d_k^* + r - 1} (1-\alpha_k^*)^{b + (n_k+1)M - d_k^* - r - 1}. \end{aligned} \tag{15}$$

Marginalizing (15) with respect to α_k^*, we obtain the marginal distribution of $\mathcal{G}^{(\mathcal{D}_k)}$ as

$$\begin{aligned} p(\mathcal{G}^{(\mathcal{D}_k)}) &= \int_0^{1/2} p(\alpha_k^*, \mathcal{G}^{(\mathcal{D}_k)}) \mathrm{d}\alpha_k^* \\ &= \sum_{r=0}^{D_k^* - d_k^*} \frac{w_{kr}^*}{\mathcal{B}(1/2; a, b)} \int_0^{1/2} {\alpha_k^*}^{a + d_k^* + r - 1} (1-\alpha_k^*)^{b + (n_k+1)M - d_k^* - r - 1} \mathrm{d}\alpha_k^* \\ &= \sum_{r=0}^{D_k^* - d_k^*} w_{kr}^* \frac{\mathcal{B}(1/2; a_{kr}^*, b_{kr}^*)}{\mathcal{B}(1/2; a, b)}, \end{aligned} \tag{16}$$

where $a_{kr}^* = a + d_k^* + r$ and $b_{kr}^* = b + (n_k+1)M - d_k^* - r$.
Combining (15) and (16), the first distribution on the right hand side of (9) can be written as

$$p(\alpha_k^* \mid \mathcal{G}^{(\mathcal{D}_k)}) = \frac{p(\alpha_k^*, \mathcal{G}^{(\mathcal{D}_k)})}{p(\mathcal{G}^{(\mathcal{D}_k)})} = \sum_{r=0}^{D_k^* - d_k^*} \varphi_{kr}^* f_{\mathrm{TBeta}}(\alpha_k^*; 1/2, a_{kr}^*, b_{kr}^*), \tag{17}$$

where the mixture weights φ_{kr}^* are given by:

$$\varphi_{kr}^* = \frac{w_{kr}^* \mathcal{B}(1/2; a_{kr}^*, b_{kr}^*)}{\sum_{r=0}^{D_k^* - d_k^*} w_{kr}^* \mathcal{B}(1/2; a_{kr}^*, b_{kr}^*)}. \tag{18}$$

Focusing on the second distribution on the right hand side of (9), we start from (10) and use (12) to get

$$p(\alpha_k^*, \mathcal{C}_k^*, \mathcal{G}^{(\mathcal{D}_k)}) = p(\alpha_k^*)(1-\alpha_k^*)^{(n_k+1)M} \left(\frac{\alpha_k^*}{1-\alpha_k^*}\right)^{\sum_{i<j}\left[(n_k+1-2n_{ij}^{(k)})A_{\mathcal{C}_k^*[ij]}+n_{ij}^{(k)}\right]}. \quad (19)$$

As mentioned in Section 2.1 of the main paper, modeling a graph $\mathcal{G}$ is equivalent to modeling the M-dimensional vector $V_{\mathcal{G}} = \text{vech}(A_{\mathcal{G}})$ defined as the half-vectorization of $A_{\mathcal{G}}$, whose components coincide with the elements of the lower triangular half of $A_{\mathcal{G}}$. We let $V_{\mathcal{C}_k^*(-[ij])}$ denote the $(M-1)$-dimensional vector after removing the element encoding the binary relation $A_{\mathcal{C}_k^*[ij]}$ between the nodes $\{i,j\}$ from the half-vectorization $V_{\mathcal{C}_k^*}$ of $A_{\mathcal{C}_k^*}$. We note that (19) highlights the conditional independence property of the elements of $V_{\mathcal{C}_k^*}$, allowing for independent edge-specific distributions, conditionally on α_k^* and $\mathcal{G}^{(\mathcal{D}_k)}$. Thus, for $g \in \{0,1\}$, the full conditional of $A_{\mathcal{C}_k^*[ij]}$ is such that:

$$\begin{aligned}
p_{kij}^* &= \mathbb{P}(A_{\mathcal{C}_k^*[ij]} = g \mid \alpha_k^*, V_{\mathcal{C}_k^*(-[ij])}, \mathcal{G}^{(\mathcal{D}_k)}) \\
&= \mathbb{P}(A_{\mathcal{C}_k^*[ij]} = g \mid \alpha_k^*, \mathcal{G}^{(\mathcal{D}_k)}) \\
&= \frac{\left(\frac{\alpha_k^*}{1-\alpha_k^*}\right)^{(n_k+1-2n_{ij}^{(k)})g+n_{ij}^{(k)}}}{\left(\frac{\alpha_k^*}{1-\alpha_k^*}\right)^{(n_k+1-2n_{ij}^{(k)})g+n_{ij}^{(k)}} + \left(\frac{\alpha_k^*}{1-\alpha_k^*}\right)^{(n_k+1-2n_{ij}^{(k)})(1-g)+n_{ij}^{(k)}}} \\
&= \left[1 + \left(\frac{\alpha_k^*}{1-\alpha_k^*}\right)^{-2\left(n_{ij}^{(k)}-(n_k+1)/2\right)(1-2g)}\right]^{-1}. \quad (20)
\end{aligned}$$

The conditional distribution $p(\mathcal{C}_k^* \mid \alpha_k^*, \mathcal{G}^{(\mathcal{D}_k)})$ in (9) is fully determined.

2.3 Distribution P_l in the generalized Pólya urn scheme

When $n_k = 1$, that is $\mathcal{D}_k = \{l\}$ for some $l = 1, \ldots, n$, the distribution in (11) and (12) simplifies to the distribution P_l in (8) and (9) of the main paper, for ϑ_l, conditionally on $\mathcal{G}_l$ and given that ϑ_l takes a new value. To see this, it can be easily verified from (12) that, for $n_k = 1$, U_k has support $u_k \in \{d_k^* = d_l, \ldots, D_k^* = 2M - d_l\}$. Moreover, $n_{ij}^{(k)} = A_{\mathcal{G}_0[ij]} + A_{\mathcal{G}_l[ij]} \in \{0,1,2\}$, $m_{k1} = d_l$, $M_{k0} = m_{k0} + m_{k2} = M - d_l$, $\gamma_{k0}(s_0) = 2s_0$. Thus, for $r = u - d_l$, the coefficient in (14)

$$w_{k(u-d_l)}^* = 2^{d_l}\binom{M-d_l}{(u-d_l)/2} = 2^{d_l}\binom{M-d_l}{s_0},$$

as there is a constrained scalar decision variable $s_0 \in \{0, \ldots, M - d_l\}$ in the condition $2s_0 = u - d_l$ defining the set $\mathcal{R}_{k(u-d_l)}$ for fixed u and d_l. Moreover, from (16), $a_{k(u-d_l)}^* = a + u = a + 2s_0 + d_l$ and $b_{k(u-d_l)}^* = b + 2(M - d_l) - u = b + 2M - 2s_0 - d_l$. That is, the coefficients a_{kr}^*, b_{kr}^* in (14) and w_{kr}^* in (15), boil down respectively to a_{ls_0}, b_{ls_0} and w_{ls_0} in (6) of the main paper. Finally, the distribution specified by (9) and (10) directly follows from (20) by setting $n_k = 1$.

2.4 Additional details on posterior computations

The Gibbs sampling in Algorithm 1 allows us to sample from the posterior distribution of $\vartheta^{(1:n)}$ conditionally on $\mathcal{G}^{(1:n)}$. We discuss some details that can help improving its efficiency.

The first one is that in the first step of the sampler, generating a new ϑ_l, for all l, has a computational complexity of the order of $\mathcal{O}(nM)$, which can become an issue for larger network dimensions and/or population sizes. However, the unnormalized probability of sampling a new value for ϑ_l, given in (6) and represented by the r.h.s. of (8), and the distribution P_l of new values for ϑ_l, conditionally on $\mathcal{G}_l$, defined in (8)-(9), are iteration-invariant, if the hyperparameter $\mathcal{G}_0$ is kept fixed. These quantities can thus be computed once before running the Gibbs sampler.

Secondly, in the reshuffling step, the update of $\vartheta_k^* = (\mathcal{C}_k^*, \alpha_k^*)$ has a computational complexity of the order of $\mathcal{O}(KM)$, with K being the number of clusters at a given iteration. Although, this step of the sampler may present a lower computational complexity, as $K \leq n$, calculations are more involved because the solution of a set of linear Diophantine equations is required for each k to compute the coefficients w_{kr}^* defined in (15) and appearing in the mixture weights given in (14). The number of equations to solve for each k depends on N and n_k through $D_k^* - d_k^*$. The complexity of each equation directly depends on n_k. Some considerations are worth noting. Let $e_{kh} = n_k + 1 - 2h$ for $h = 0, \ldots, \lfloor n_k/2 \rfloor$, and let g_k denote the common greatest divisor of the vector $e_k = (e_{k0}, e_{k1}, \ldots, e_{k\lfloor n_k/2 \rfloor})$, namely $g_k = cgd(e_k)$. The linear Diophantine equation defining the set $\mathcal{S}_{kr}$ ($\mathcal{R}_{kr}$) has no solution when $r + d_k^* - \sum_{h=0}^{n_k/2} hM_{kh}$ $\left(r + d_k^* - (\lfloor n_k/2 \rfloor + 1)m_{k(\lfloor n_k/2 \rfloor + 1)} - \sum_{h=0}^{\lfloor n_k/2 \rfloor} hM_{kh}\right)$ is not a multiple of g_k. In this case, $w_{kr}^* = 0$. Moreover, the coefficient w_{kr}^* is symmetric with respect to the index r, that is $w_{kr}^* = w_{k\bar{r}}^*$, with $\bar{r} = D_k^* - d_k^* - r$. With these arguments, the overall computational time needed to define the set $\mathcal{S}_{kr}$ ($\mathcal{R}_{kr}$) can be more than halved. From a practical perspective, this can be solved with the algorithm based on a generating function of Hardy and Littlewood (1966), used by Voinov and Nikulin (1997) and implemented in the `nilde` R package (Arnqvist et al., 2022), by imposing $\sum_{h=0}^{\lfloor n_k/2 \rfloor} s_h \leq M_k$, where $M_k = \sum_{h=0}^{\lfloor n_k/2 \rfloor} M_{kh}$ and retaining only the feasible solutions, that is those satisfying $s_h \in \{0, \ldots, M_{kh}\}$ $\forall h$. Yet, defining the set $\mathcal{S}_{kr}$ ($\mathcal{R}_{kr}$) at each iteration for any $k = 1, \ldots, K$ and $r = 0, \ldots, D_k^* - d_k^*$ substantially increases the computational time required in the reshuffling step. An alternative and cheaper strategy consists in replacing the conditional distribution of α_k^* given $\mathcal{G}^{(\mathcal{D}_k)}$ in (11) by its full conditional. That is $\alpha_k^* \mid \mathcal{C}_k^*, \mathcal{G}^{(\mathcal{D}_k)} \sim \text{TBeta}(1/2; a_k^*, b_k^*)$ where $a_k^* = a + \sum_{l \in \mathcal{D}_k^\dagger} d_{\text{H}}(\mathcal{G}_l, \mathcal{C}_k^*)$ and $b_k^* = b + (n_k + 1)M - \sum_{l \in \mathcal{D}_k^\dagger} d_{\text{H}}(\mathcal{G}_l, \mathcal{C}_k^*)$. Sampling α_k^* from its full conditional finds justification in that, along the chain, it targets the distribution in (11). Yet in this step, noting that the Bernoulli random variables given in (12) are identically distributed for all pairs of nodes $\{i, j\}$ sharing the same value of $n_{ij}^{(k)} = h$, only $m_{kh} \leq M$ Bernoulli parameters p_{kij}^* defined in (13) must be computed.

3 Posterior prediction

3.1 Cluster-specific posterior predictive distribution

The posterior predictive distribution implied by the statistical model defined in (2.1) is

$$p(\mathcal{G}^\star \mid \mathcal{G}^{(1:n)}) = \int_\Theta \psi\left(\mathcal{G}^\star; \vartheta\right) \mathrm{d}p(\vartheta \mid \mathcal{G}^{(1:n)}) \tag{21}$$

where $p(\vartheta \mid \mathcal{G}^{(1:n)})$ denotes the posterior distribution of $\vartheta = (\mathcal{C}, \alpha)$. Although the above integral is not analytically available, it is straightforward to simulate networks from the posterior predictive distribution exploiting MCMC samples for ϑ along with the predictive distribution structure of the underlying DP. On the other hand, conditionally on an estimated partition of the observed graphs, the cluster-specific one-step-ahead posterior predictive distribution implied by our model is available in closed-form. Predicting a graph $\mathcal{G}^\star$ from the posterior predictive distribution specific to the cluster of observations with indices in $\mathcal{D}_k$ translates into sampling M independent Bernoulli distributions. Specifically,

$$A_{\mathcal{G}^\star[ij]} \mid \mathcal{G}^{(\mathcal{D}_k)} \overset{\text{ind}}{\sim} \text{Bern}(\tilde{p}_{kij}), \qquad\qquad i < j. \tag{22}$$

After introducing the quantities $T^\star_{kij} = \sum_{\{u,v\}\neq\{i,j\}} \max\{n^{(k)}_{uv}, n_k + 1 - n^{(k)}_{uv}\}$ and $t^\star_{kij} = \sum_{\{u,v\}\neq\{i,j\}} \min\{n^{(k)}_{uv}, n_k + 1 - n^{(k)}_{uv}\}$ where $n^{(k)}_{uv} = \sum_{l\in\mathcal{D}^\dagger_k} A_{\mathcal{G}_l[uv]}$, the Bernoulli parameters in (22) are given by:

$$\tilde{p}_{kij} = \mathbb{E}\left[A_{\mathcal{G}^\star[ij]} \mid \mathcal{G}^{(\mathcal{D}_k)}\right] = \frac{1}{p(\mathcal{G}^{(\mathcal{D}_k)})} \sum_{r=0}^{T^\star_{kij}-t^\star_{kij}} \tilde{w}_{kr} \frac{\left[\mathcal{B}(1/2; \tilde{a}^{(r)}_{kij}, \tilde{b}^{(r)}_{kij}) + \mathcal{B}(1/2; \tilde{c}^{(r)}_{kij}, \tilde{d}^{(r)}_{kij})\right]}{\mathcal{B}\left(1/2; a, b\right)}, \tag{23}$$

where $\tilde{a}^{(r)}_{kij} = a + n^{(k)}_{ij} + t^\star_{kij} + r + 1$, $\tilde{b}^{(r)}_{kij} = b + (n_k + 1)\, M - (n^{(k)}_{ij} + t^\star_{kij} + r)$, $\tilde{c}^{(r)}_{kij} = a + n_k + 1 - n^{(k)}_{ij} + t^\star_{kij} + r$ and $\tilde{d}^{(r)}_{kij} = b + (n_k + 1)\, M - (n_k + 1 - n^{(k)}_{ij} + t^\star_{kij} + r) + 1$. Thus, cluster-specific one-step-ahead edge prediction probability is a linear combination of two incomplete beta functions whose parameters reflect edge presence and edge absence, respectively. In turn, the expression of the coefficient $\tilde{w}_{kr}$ in (23) can be retrieved from (15) with minor modifications. Specifically, it suffices to replace: D^*_k with $T^\star_{kij}$, d^*_k with $t^\star_{kij}$, m_{kh} with $q_{khij} = \#\left\{\{u,v\} \in \mathcal{V}^2 : \{u,v\} \neq \{i,j\}, n^{(k)}_{uv} = h\right\}$ and M_{kh} with $Q_{khij} = q_{khij} + q_{k(n_k+1-h)ij}$. Finally, the marginal likelihood $p(\mathcal{G}^{(\mathcal{D}_k)})$ in (23) is analytically available and results from the joint distribution of $\mathcal{G}^{(\mathcal{D}_k)}$ and $A_{\mathcal{G}^\star[ij]}$ by marginalizing the latter. It is given by:

$$p(\mathcal{G}^{(\mathcal{D}_k)}) = \sum_{r=0}^{D^*_k - d^*_k} w^*_{kr} \frac{\mathcal{B}(1/2; a^*_{kr}, b^*_{kr})}{\mathcal{B}\left(1/2; a, b\right)}, \tag{24}$$

with a^*_{kr} and b^*_{kr} defined in (14) and w^*_{kr} in (15). It is worth noting that the marginal prior $p\left(\mathcal{G}^\star\right) = \int \psi\left(\mathcal{G}^\star; \vartheta\right) \mathrm{d}P_0(\vartheta)$, appearing in π_{l0} in (6), can be easily retrieved for

generative purposes as an instance of (22) when $n_k = 0$, that is $\mathcal{D}_k = \varnothing$. In this case, $p(\mathcal{G}^{(\varnothing)}) = 1$ and (23) boils down to the prior expectation:

$$\mathbb{E}\left[A_{\mathcal{G}^\star[ij]}\right] = \frac{\mathcal{B}(1/2; a+1+A_{\mathcal{G}_0[ij]}, b+1-A_{\mathcal{G}_0[ij]}) + \mathcal{B}(1/2; a+1-A_{\mathcal{G}_0[ij]}, b+1+A_{\mathcal{G}_0[ij]})}{\mathcal{B}\left(1/2; a, b\right)}.$$

Figure 1 shows how the cluster-specific one-step-ahead posterior predictive probability in (23) varies based on how frequently the edge between node $\{i, j\}$ appears in the graphs belonging to the k-th cluster and the prior graph $\mathcal{G}_0$, with $\tilde{p}_{kij} = 1/2$ for $n_{ij}^{(k)} = (n_k + 1)/2$. Moreover, as highlighted by the mixed color of each line, $\tilde{p}_{kij}$ is symmetric with respect to the (equal) frequency of all the other edges $\{u, v\} \neq \{i, j\}$, meaning that the values $n_{uv}^{(k)} = z$ and $n_{uv}^{(k)} = n_k + 1 - z$ share the same curve. In addition, it is worth noting how $\tilde{p}_{kij}$ is an odd function with center shifted at $((n_k + 1)/2, 1/2)$, meaning that $f(x) = 1 - f(n_k + 1 - x)$ for $\tilde{p}_{kij}$ function of $n_{ij}^{(k)}$. While in Figure 1 we aim at isolating the effect of differences among the $n_{uv}^{(k)}$'s on $\tilde{p}_{kij}$, Figure 2, shows, instead, how $\tilde{p}_{kij}$ varies as a function of $n_{ij}^{(k)}$, for $z = \sum_{\{u,v\} \neq \{i,j\}} n_{uv}^{(k)}$ ranging in $\{0, 1, \ldots, (n_k + 1)(M - 1)\}$, where $n_{uv}^{(k)}$ can change across $\{u, v\}$. In this more realistic case, symmetries can only happen based on the value of q_{khij}.

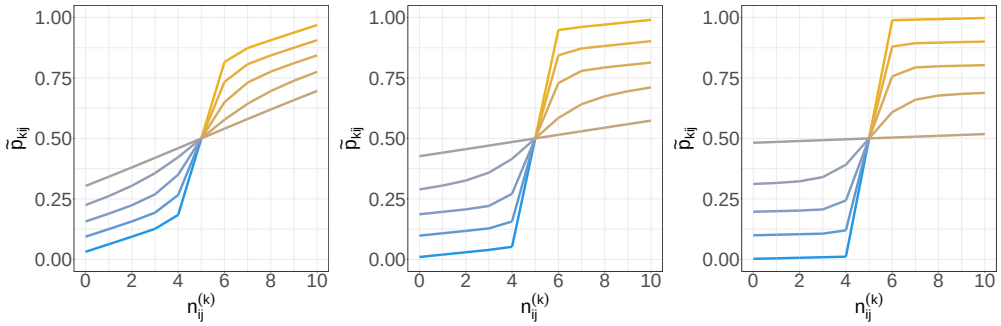

Figure 1: Probability $\tilde{p}_{kij}$ in (23), with $n_{uv}^{(k)} = z \; \forall \{u, v\} \neq \{i, j\}$, with $z \in \{0, 1, \ldots, n_k + 1\}$ (blue for low and yellow for high) and for $n_{ij}^{(k)} \in \{0, 1, \ldots, n_k + 1\}$, with $n_k + 1 = 10$, and for $N \in \{3, 5, 10\}$ (from left to right).

Cluster-specific m-step-ahead posterior predictive distribution

The distribution given in (22)–(24) can be generalized to make predictions on, say, m graphs jointly, conditionally on the estimated partition and on the fact that they belong to the same cluster, say the k-th one. Such distribution, denoted by $p(\mathcal{G}_1^\star, \ldots, \mathcal{G}_m^\star \mid \mathcal{G}^{(\mathcal{D}_k)})$, is defined on the m-dimensional cartesian product $\mathscr{G}_\mathcal{V} \times \ldots \times \mathscr{G}_\mathcal{V}$ and it turns out that we can make edge-specific predictions independently, conditionally on $\mathcal{G}^{(\mathcal{D}_k)}$. It is thus sufficient to study the conditional distribution of $\mathcal{H}_{[ij]}^{(1:m)} = \sum_{l=1}^m A_{\mathcal{G}_l^\star[ij]}$, given

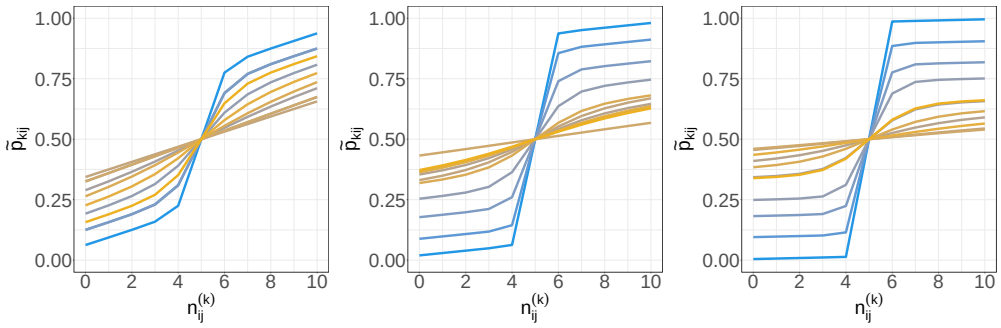

Figure 2: Probability $\tilde{p}_{kij}$ in (23), with $\sum_{\{u,v\}\neq\{i,j\}} n_{uv}^{(k)} = z$, with z taking 11 equally-spaced values in $\{0, 1, \ldots, (n_k+1)(M-1)\}$ (blue for low and yellow for high) and for $n_{ij}^{(k)} \in \{0, 1, \ldots, n_k+1\}$, with $n_k+1 = 10$, and for $N \in \{3, 5, 10\}$ (from left to right).

$\mathcal{G}^{(\mathcal{D}_k)}$, for which we get

$$\mathcal{H}_{[ij]}^{(1:m)} \mid \mathcal{G}^{(\mathcal{D}_k)} \overset{\text{ind}}{\sim} \text{Cat}(\tilde{p}_{kij0}, \tilde{p}_{kij1}, \tilde{p}_{kij2}, \ldots, \tilde{p}_{kijm}), \qquad i < j$$

where, for $h = 0, \ldots, m$,

$$\begin{aligned}\tilde{p}_{kijh} &= \mathbb{P}\left(\mathcal{H}_{[ij]}^{(1:m)} = h \mid \mathcal{G}^{(\mathcal{D}_k)}\right) \\ &= \frac{\binom{m}{h}}{p(\mathcal{G}^{(\mathcal{D}_k)})} \sum_{r=0}^{T_{kij}^\star - t_{kij}^\star} \tilde{w}_{kr} \frac{\left[\mathcal{B}(1/2; \tilde{a}_{kijh}^{(r)}, \tilde{b}_{kijh}^{(r)}) + \mathcal{B}(1/2; \tilde{c}_{kijh}^{(r)}, \tilde{d}_{kijh}^{(r)})\right]}{\mathcal{B}(1/2; a, b)}\end{aligned} \tag{25}$$

where $\tilde{a}_{kijh}^{(r)} = a + n_{ij}^{(k)} + t_{kij}^\star + r + h$, $\tilde{b}_{kijh}^{(r)} = b + (n_k+1)\,M - (n_{ij}^{(k)} + t_{kij}^\star + r) + m - h$, $\tilde{c}_{kijh}^{(r)} = a + n_k + 1 - n_{ij}^{(k)} + t_{kij}^\star + r + m - h$ and $\tilde{d}_{kijh}^{(r)} = b + (n_k+1)\,M - (n_k + 1 - n_{ij}^{(k)} + t_{kij}^\star + r) + h$, with $t_{kij}^\star$, $T_{kij}^\star$, $\tilde{w}_{kr}$ and $p(\mathcal{G}^{(\mathcal{D}_k)})$ defined in Section 3.1. Figure 3 shows how $\tilde{p}_{kijh}$ varies as function of h, for different values of $n_{ij}^{(k)}$ and $n_{uv}^{(k)}$.

3.2 Cluster-specific posterior distribution of $\mathcal{C}_k^*$

For each $k = 1, \ldots, \hat{K}$, the conditional distribution of $\mathcal{C}_k^*$ given α_k^* and $\mathcal{G}^{(\mathcal{D}_k)}$ is given in (12) and (13). Here we study the conditional distribution of $\mathcal{C}_k^*$ given $\mathcal{G}^{(\mathcal{D}_k)}$, that we obtain from the latter by marginalizing with respect to α_k^*. It turns out that:

$$A_{\mathcal{C}_k^*[ij]} \mid \mathcal{G}^{(\mathcal{D}_k)} \overset{\text{ind}}{\sim} \text{Bern}(p_{kij}^m), \qquad i < j \tag{26}$$

where the Bernoulli parameters in (26) are given by:

$$p_{kij}^m = \mathbb{E}\left[A_{\mathcal{C}_k^*[ij]} \mid \mathcal{G}^{(\mathcal{D}_k)}\right] = \frac{1}{p(\mathcal{G}^{(\mathcal{D}_k)})} \sum_{r=0}^{T_{kij}^\star - t_{kij}^\star} \tilde{w}_{kr} \frac{\mathcal{B}(1/2; a_{kr}^m, b_{kr}^m)}{\mathcal{B}\,(1/2; a, b)} \tag{27}$$

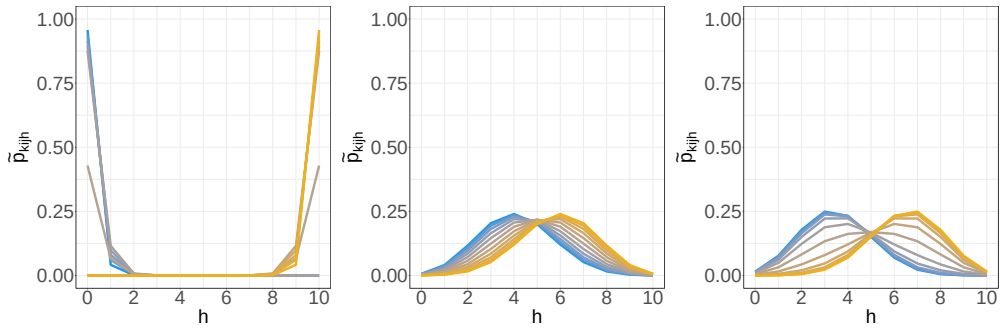

Figure 3: Probability $\tilde{p}_{kijh}$ in (25), for $h \in \{0, \ldots, m\}$, with $m = 10$, and $n_{ij}^{(k)} \in \{0, 1, \ldots, n_k + 1\}$, with $n_k + 1 = 10$ (blue for low and yellow for high), with $N = 10$, and $\sum_{\{u,v\} \neq \{i,j\}} n_{uv}^{(k)} = z$, with $z \in \{0, (\frac{n_k+1}{2})(M-1), (n_k+1)(M-1)\}$ (from left to right).

and $a_{kr}^m = a + n_k + 1 - n_{ij}^{(k)} + t_{kij}^\star + r$, $b_{kr}^m = b + (n_k + 1)(M-1) + n_{ij}^{(k)} - t_{kij}^\star - r$, with $t_{kij}^\star$, $T_{kij}^\star$ and $\tilde{w}_{kr}$ defined in Section 3.1, and $p(\mathcal{G}^{(\mathcal{D}_k)})$ is given in (24).

Figure 4 shows how p_{kij}^m varies as a function of $n_{ij}^{(k)}$, for $z = \sum_{\{u,v\} \neq \{i,j\}} n_{uv}^{(k)}$ ranging in $\{0, 1, \ldots, (n_k + 1)(M-1)\}$, where $n_{uv}^{(k)}$ can change across $\{u, v\}$ and can serve as comparison to the probability in Figure 2.

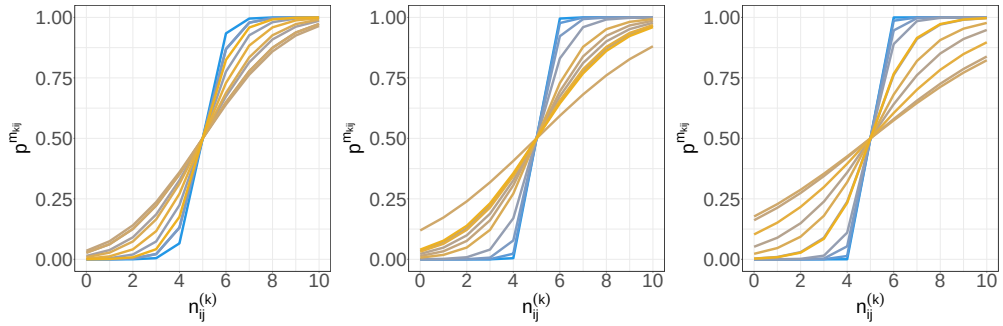

Figure 4: Probability p_{kij}^m in (27), with $\sum_{\{u,v\} \neq \{i,j\}} n_{uv}^{(k)} = z$, with z taking 11 equally-spaced values in $\{0, 1, \ldots, (n_k + 1)(M-1)\}$ (blue for low and yellow for high) and for $n_{ij}^{(k)} \in \{0, 1, \ldots, n_k + 1\}$, with $n_k + 1 = 10$, and for $N \in \{3, 5, 10\}$ (from left panel to right panel).

4 Further details on the simulation study of Section 4

In Table 1, we report the specification of the parameters for the four data-generating processes used in Section 4 to generate the centroids $\mathcal{C}_{0k}$.

Graphical structure	Specification
Scale-free	We set the power law exponent of the degree distribution to 2 and the sparsity to 0.2.
Small-world	We set the degree of the lattice to 10 and the probability of rewiring to 0.2.
Stochastic Block Model	We set the number of blocks to 2, with membership probabilities equal to 1/2; the inclusion probabilities were set as 0.9 and 0.1 for diagonal and nondiagonal blocks, respectively.
Erdős–Rényi	Probability of inclusion was set to 0.3.

Table 1: Graphical structures and the corresponding parameter specification used to define the distribution centroids.

4.1 Varying sample size with $\mathbb{L}^1$ distance

We present additional simulation experiments investigating how the posterior mean

$$\hat{f}(\cdot) = \mathbb{E}[\tilde{f}(\cdot) \mid \mathcal{G}^{(1:n)}] = \frac{1}{c+n} \int_{\Theta} \psi(\cdot;\vartheta)\, \mathrm{d}P_0(\vartheta) + \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{K^{(t)}} \frac{n_k^{(t)}}{c+n}\, \psi(\cdot;\vartheta_k^{*(t)})$$

concentrates around its true value $p_*(\cdot) = \sum_{k=1}^{4} 0.25 p_{\mathrm{CER}}(\cdot;\mathcal{C}_{0k},\alpha_{0k})$, as a function of the sample size n, with $\hat{f}$ evaluated based on the posterior sample generated from Algorithm 1. Unlike the study presented in Section 4.2, we focus here on the $\mathbb{L}^1$ distance as a metric on $\mathcal{P}_{\mathscr{G}_\mathcal{V}}$, and study the distribution of the distance between p_* and $\hat{f}$ for finite samples of size $n \in \{40, 80, 120, 200\}$. The evaluation of $\mathbb{L}^1(p_*;\hat{f})$ requires summation over the graph space $\mathscr{G}_\mathcal{V}$, which is prohibitive even for moderate N. Thus, we propose an importance-sampling approximation of $\mathbb{L}^1(p_*;\hat{f})$. Namely,

$$\begin{aligned}\mathbb{L}^1(p_*;\hat{f}) &= \sum_{\mathcal{G}\in\mathscr{G}_\mathcal{V}} |p_*(\mathcal{G}) - \hat{f}(\mathcal{G})| = \sum_{\mathcal{G}\in\mathcal{G}_N} \frac{|p_*(\mathcal{G}) - \hat{f}(\mathcal{G})|}{p_*(\mathcal{G})} p_*(\mathcal{G}) = \mathbb{E}_{p_*}\left[\frac{|p_*(\mathcal{G}) - \hat{f}(\mathcal{G})|}{p_*(\mathcal{G})}\right] \\ &\approx \frac{1}{L}\sum_{l=1}^{L} \frac{|p_*(\mathcal{G}) - \hat{f}(\mathcal{G})|}{p_*(\mathcal{G})},\end{aligned}$$

with $\mathcal{G}_l \overset{\text{iid}}{\sim} p_*$, for $l = 1, \ldots, L$. The results are presented in Figure 5, which shows that the posterior estimate $\hat{f}$ gets closer to p_* as the sample size increases. Our model appears to converge to p_* faster than the models proposed by Durante et al. (2017), Mantziou et al. (2024) and Signorelli and Wit (2020). This additional study gives credibility to the robustness of our model with respect to the choice of the metric d on $\mathcal{P}_{\mathscr{G}_\mathcal{V}}$.

4.2 Additional simulation experiment

We present an additional experiment to assess the behavior of the DP mixture of CER kernels when more intricate connectivity patterns than those presented in Section 4 of

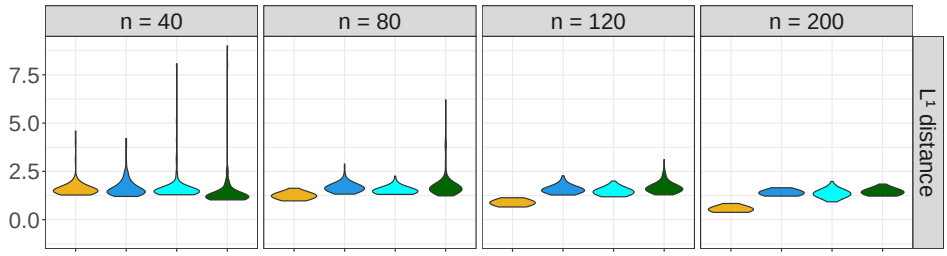

Figure 5: Importance-sampling approximate distributions of $\mathbb{L}^1(p_*; \hat{f})$ distance for our method (yellow violins), and the methods of Durante et al. (2017) (blue violins), Mantziou et al. (2024) (cyan violins) and Signorelli and Wit (2020) (green violins). Distributions are estimated based on the analysis of 100 datasets.

the main paper are considered. Specifically, we consider the core-periphery structure, which may result, for example, from a non-assortative stochastic block model generative process, in which nodes in the core are densely linked to each other and often to the periphery, and peripheral nodes are typically linked to the core but weakly connected with each other. As in Section 4 of the main paper, we focus on networks with $N = 20$ nodes. A set of $n = 40$ observations are sampled from a two-component mixture of CER $p_*(\cdot) = 0.5p_{\text{CER}}(\cdot; \mathcal{C}_{01}, \alpha_{01}) + 0.5p_{\text{CER}}(\cdot; \mathcal{C}_{02}, \alpha_{02})$, where the centroids $\mathcal{C}_{01}$ and $\mathcal{C}_{02}$ have a core-periphery (Borgatti and Everett, 2000) and a Erdős–Rényi (Erdős and Rényi, 1960) structure, respectively, and the component-specific scales of variation are set equal to $\alpha_{01} = 0.4$ and $\alpha_{02} = 0.3$. Figure 6 illustrates the generated centroids. To assess the ability of our method to cluster multiple network data, we compare the estimated partition to the true partition, which reflects the two-component mixture structure of the data-generating model. We resort to three metrics: the adjusted Rand index, clustering entropy and clustering purity. The results of our investigation are displayed in Figure 7. The performance of our model appears robust to this more complex scenario.

5 On the implementation of the competing methods

We detail here the specification of the hyperparameters for all models included in the comparison.

Table 2 reports the hyperparameter settings adopted for the implementation of the method of Durante et al. (2017) in the studies of Section 4, corresponding to the default options of the authors' code. In the study of Section 5, the upper bound on the number of classes H is set to twice the number of individuals, i.e. $H = 60$.

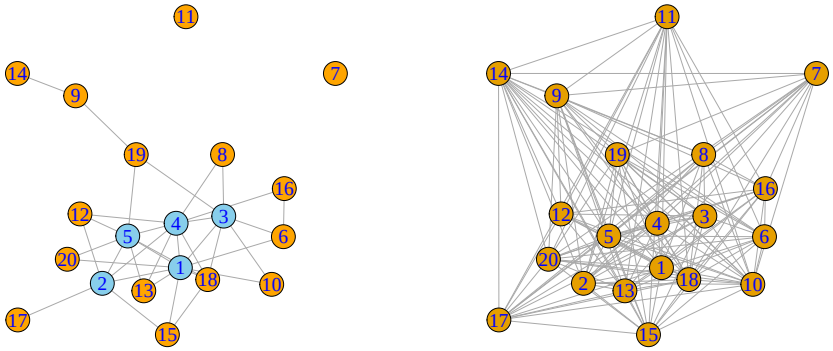

Figure 6: Left: centroid $\mathcal{C}_{01}$ with a core–periphery structure, core nodes shown in blue. Right: centroid $\mathcal{C}_{02}$ with an Erdős–Rényi structure. See Section 4.2.

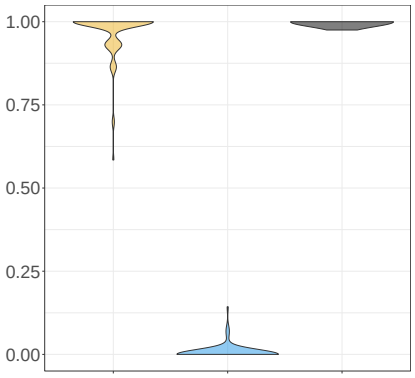

Figure 7: Adjusted Rand index (left), clustering entropy (center), and clustering purity(right). Distributions are estimated based on the analysis of 100 datasets. See Section 4.2.

For the method of Josephs et al. (2025) in the simulation study of Section 4.1, we set the truncation levels for the number of classes K and for the number of within-class node clusters L to 15, following the default configuration of the authors' code. We employ the Incompatible Blocked Gibbs (IBG) sampler, which the authors report as yielding the best clustering performance.
For the method of Mantziou et al. (2024), uniform priors are assigned to all component-specific parameters, and representative networks with two node blocks are used, as in the default setting of the authors' code. In the simulation studies of Section 4, the number of components is fixed to match the number of mixture components in the data-generating process, i.e. $K = 4$. In the study of Section 5, we adopt the Sparse Finite Mixture extension of Mantziou et al. (2024), setting the upper bound on the number of clusters to $C_{\max} = 60$ and placing a Gamma$(a_e = 1, b_e = 400)$ hyperprior on the hyperparameter e_0 of the symmetric Dirichlet prior on the mixture weights

$\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{C_{\max}})$, to favour values of e_0 close to zero, as recommended by the authors. For the method of Signorelli and Wit (2020), the number of components is set equal to the number of mixture components in the simulation studies of Section 4, i.e. $K = 4$, and the unconstrained network model is adopted for the specification of the mixture components, where the number of parameters equals the number of edge pairs.

Section	R	H	a_1	a_2	μ_l	σ_l^2
4	10	30	2.5	3.5	0	10
5	10	60	2.5	3.5	0	10

Table 2: Hyperparameter specification for the method of Durante et al. (2017). The table reports: the upper bound on the latent space dimension R; the upper bound on the number of classes H; the hyperparameters a_1 and a_2 of the multiplicative inverse-Gamma prior; and the Gaussian prior mean μ_l and variance σ_l^2 for Z, for all $l = 1, \ldots, N(N-1)/2$.

6 Further details on illustrations in Section 6

We present additional plots related to the application in Section 6.

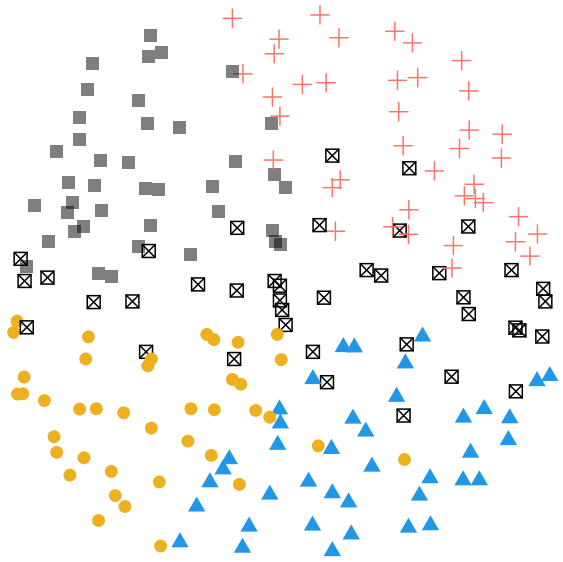

Figure 8: A 2-D visualization (top-down projection) of the atlas with 200 ROIs, where colors and shapes represent the $m_{\text{sub}} = 5$ node cluster memberships identified through balanced clustering with $N_{\text{sub}} = 40$.

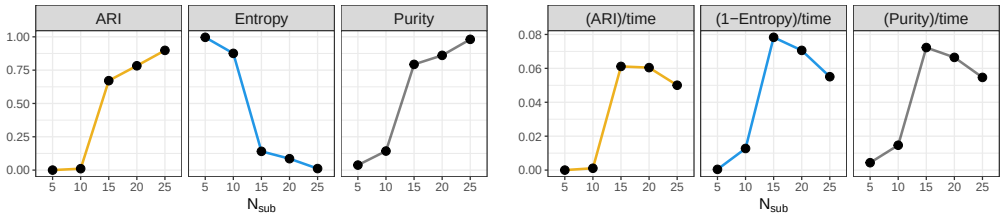

Figure 9: Clustering metrics comparing the partition estimated based on the consensus subgraph approach, with the one estimated with exact method on 48 ROIs, for the Human Brain dataset based on 48 ROIs, for N_{sub} ranging in $\{5, 10, \ldots, 25\}$.

6.1 Consensus subgraph clustering for large N with nodes partitioned at random

For comparison with the analysis in Section 6, we performed consensus subgraph clustering on the human brain datasets, with 48 and 200 ROIs, by partitioning the nodes randomly, thus without utilizing the available spatial information on the nodes. This allows us to understand the impact of incorporating spatial information when partitioning the nodes. Interestingly, when nodes are partitioned at random, the clustering metrics computed on the estimated data clustering appear only slightly worse than those obtained in Section 6 based on the available spatial information. The results of our analysis are presented in Figure 10 and Table 3. N_{sub} was set equal to 15 for both versions of the human brain datasets.

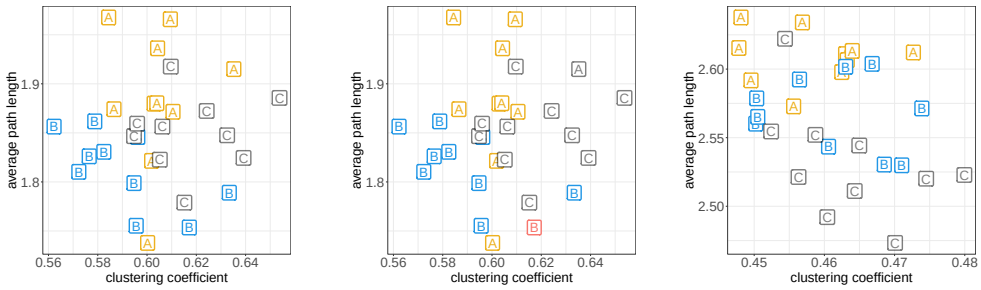

Figure 10: Scatter plots for the small-world properties of brain networks for three subjects in the dataset. Colors indicate the cluster membership, letters refer to the subject ID in the dataset, namely 0025443 (A), 0025445 (B) and 0025446 (C). Left panel: $N = 48$, partition estimated via DP mixture of CER kernels. Central panel: $N = 48$, partition estimated via consensus subgraph clustering with nodes partitioned at random. Right panel: $N = 200$, partition estimated via consensus subgraph clustering with nodes partitioned at random. See Figure 7 for a comparison.

N (N_{sub})	$\hat{K}$	Adjusted Rand Index	Entropy	Purity
48 (15)	30	0.6642	0.1420	0.7970
200 (15)	31	0.9490	0.0162	0.9699

Table 3: Human brain dataset. Estimated number of clusters and clustering metrics comparing the partition estimated based on the consensus subgraph clustering approach, with nodes partitioned at random, with the one implied by the presence of 30 individuals in the study. See Table 3 for a comparison.

References

Arnqvist, N. P., Voinov, V., Makarov, R., and Voinov, Y. (2022). *nilde: Nonnegative Integer Solutions of Linear Diophantine Equations with Applications*. 9

Borgatti, S. P. and Everett, M. G. (2000). "Models of core/periphery structures." *Social Networks*, 21(4): 375–395. 15

Durante, D., Dunson, D. B., and Vogelstein, J. T. (2017). "Nonparametric Bayes modeling of populations of networks." *Journal of the American Statistical Association*. 14, 15, 17

Erdős, P. and Rényi, A. (1960). "On the evolution of random graphs." *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 5: 17–61. 15

Ewens, W. J. (1990). "Population genetics theory-the past and the future." In *Mathematical and statistical developments of evolutionary theory*, 177–227. Springer. 1

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. 3

Hardy, G. H. and Littlewood, J. E. (1966). *Collected papers of GH Hardy: including joint papers with JE Littlewood and others*, volume 6. Oxford: Clarendon Press. 9

Josephs, N., Amini, A. A., Paez, M., and Lin, L. (2025). "Nested stochastic block model for simultaneously clustering networks and nodes." *arXiv:2307.09210*. 16

Mantziou, A., Lunagómez, S., and Mitra, R. (2024). "Bayesian model-based clustering for populations of network data." *The Annals of Applied Statistics*, 18(1): 266–302. 14, 15, 16

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors." *Statistica Sinica*, 4(2): 639–650. 1

Signorelli, M. and Wit, E. C. (2020). "Model-based clustering for populations of networks." *Statistical Modelling*, 20(1): 9–29. 14, 15, 17

Voinov, V. G. and Nikulin, M. S. (1997). *On a Subset Sum Algorithm and Its Probabilistic and Other Applications*, 153–163. Boston, MA: Birkhäuser Boston. 9