

Bayesian Nonparametric Modeling of Heterogeneous Populations of Networks

Francesco Barile^{*}, Simón Lunagómez[‡], and Bernardo Nipoti[†]

Abstract. The increasing availability of multiple network data has highlighted the need for statistical models for heterogeneous populations of networks. A convenient framework makes use of metrics to measure similarity between networks. In this context, we propose a novel Bayesian nonparametric model that identifies clusters of networks characterized by similar connectivity patterns. Our approach relies on a location-scale Dirichlet process mixture of centered Erdős–Rényi kernels, with components parametrized by a unique network representative, or mode, and a univariate measure of dispersion around the mode. We demonstrate that this model has full support in the Kullback–Leibler sense and is strongly consistent. An efficient Markov chain Monte Carlo scheme is proposed for posterior inference and clustering of multiple network data. The performance of the model is validated through extensive simulation studies, showing improvements over state-of-the-art methods. Additionally, we present an heuristic strategy to extend the application of the proposed model to datasets with a large number of nodes. We illustrate our approach with the analysis of human brain network data.

Keywords: centered Erdős–Rényi distribution, consensus subgraph clustering, Dirichlet process, multiple network data.

MSC2020 subject classifications: Primary62F15, 62H30, 62G05.

1 Introduction

In recent years, multiple network link relations on the same set of nodes have become prominent in many fields of application. For instance, in neuroscience, interconnections among brain regions are collected to characterize a population of individuals suffering from a neurological disorder (Nelson et al., 2017); in computer science, human mobility is studied by tracking individuals’ movements in relation to intelligent displays (Shaw et al., 2018). Multiple networks are also known as multiplex networks (Mucha et al., 2010) and can be intended either as multiple link relations among the nodes of the network (replicated networks) or as a single link relation observed over different conditions, such as one network evolving over time (longitudinal networks). The developments proposed in this work are motivated by the problem of modeling heterogeneous populations of networks, with a key application of our modeling approach being the clustering of multiple network data. For instance, Figure 1 illustrates a sample of observations from

arXiv: 2410.10354

^{*}Department of Economics, Management and Statistics, University of Milano-Bicocca, Italy, francesco.barile@unimib.it

[†]Department of Economics, Management and Statistics, University of Milano-Bicocca, Italy, bernardo.nipoti@unimib.it

[‡]Instituto Tecnológico Autónomo de México, CDMX 01080, Mexico, simon.lunagomez@itam.mx

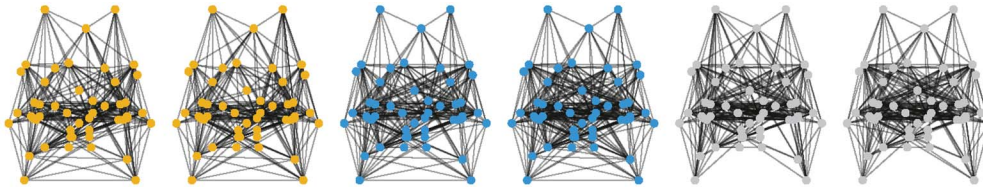


Figure 1: Top-down projection of a sample of six network observations extracted from the Human Brain Networks dataset (see Section 5 for details). The nodes of each network are colored according to the network cluster assignments, as inferred by the proposed method.

the Human Brain Network dataset analyzed in Section 5, with networks colored based on the clusters inferred by our method. Notably, the approach successfully identifies clusters of similar networks, even when differences between networks are not visually evident.

Recent studies have extended modeling strategies originally designed for a single network observation to multiple network data. Three different, not necessarily disjoint, frameworks for multiple network data have been considered in the literature. The first one is defined through latent space models, with the occurrence of an edge between two nodes depending on the positions of the nodes in a latent space (Gollini and Murphy, 2016; Durante et al., 2017; Nielsen and Witten, 2018; D’Angelo et al., 2019; Arroyo et al., 2021; Jing et al., 2021; Wang et al., 2021; D’Angelo et al., 2023). A second one is based on the use of metrics to measure similarities among networks (Donnat and Holmes, 2018). Central to this framework is the notion of representative network, defined with respect to a probabilistic model induced by a suitable choice of metric on the space of networks, as a reference for a population (Ginestet et al., 2017; Kolaczyk et al., 2017; Lunagómez et al., 2021). Third, measurement error models account for noise in network data by addressing the uncertainty associated with falsely observing edges or non-edges in a network (Young et al., 2022; Mantziou et al., 2024). In this setting, Le et al. (2018) propose a two-stage algorithm that assumes the network population can be represented by a single network with community structure, while accounting for a measurement error process that prevents an accurate observation of the representative network. Stanley et al. (2016) and Mukherjee et al. (2017) are among the first works to address the problem of clustering multiple network data. The former propose a method that clusters networks into groups sharing a common community structure and then detects node communities within each group using a shared stochastic block model, whereas the latter define a graphon mixture model. Diquigiovanni and Scarpa (2019) propose a clustering approach that involves specifying an ad hoc measure of similarity between networks and implementing an agglomerative method to cluster the networks based on these similarities. Signorelli and Wit (2020) provide a model-based approach for clustering multiple network data with a predefined number of clusters, accounting for the presence of covariates. In line with Le et al. (2018), Mantziou et al. (2024) relax the unimodality assumption to accommodate heterogeneous network data by defining a mixture of measurement error models with a fixed and known number of components and

a stochastic block model structure for the representative networks. A similar approach is proposed by Young et al. (2022), with the key difference being that the representative networks are not required to form communities in their structural topology. Yin et al. (2019) propose a finite mixture with components assumed to belong to the exponential random graph class of models. Finally, Durante et al. (2025) recently introduced a comprehensive framework for a broad class of multidimensional networks, including the multiple network data considered here, to accommodate block connectivity structures both within and across networks.

We propose a structure-free modeling approach for multiple network data, where no constraints are imposed on the topology characterizing the data generative process. Our strategy accounts for the heterogeneity that may exist in a population of networks, without imposing rigid assumptions on the number of network subgroups driving the heterogeneity. Such flexibility is achieved by adopting a Bayesian nonparametric approach. The literature on nonparametric inferential methods for network-valued observations in heterogeneous populations is relatively young and rapidly evolving. Durante et al. (2017) introduce a Bayesian nonparametric approach for the joint modeling of edge distributions through a flexible mixture representation. Guha and Guhaniyogi (2024) propose a model for clustering networks via covariate-dependent coefficients. Josephs et al. (2025) develop a framework that simultaneously clusters nodes and networks in the case of unlabeled graphs. Focusing exclusively on the structure of the nodes, Amini et al. (2024) propose a method for community detection in multiplex networks. Other relevant contributions to network clustering include Reyes and Rodriguez (2016) and Ren et al. (2023), which are based on stochastic block models and exponential random graph models, respectively. In a supervised setting, Josephs et al. (2023) and Guha et al. (2024) address the problem of network classification. Guha and Rodriguez (2021, 2023) propose methodologies in which networks serve as covariates. To our knowledge, no existing nonparametric methods have been devised for clustering multiple networks without imposing structural assumptions on the generative process. We address this gap by proposing a Bayesian nonparametric distance-based model that combines flexibility and tractability.

1.1 Our contribution

We propose a location-scale Dirichlet process mixture of centered Erdős–Rényi kernels to model heterogeneous populations of networks. This kernel choice favors both analytical and computational tractability. The flexibility of the model is ensured by its full topological support, in the Kullback–Leibler sense, over the space of network distributions on a given set of nodes. Additionally, the model’s location-scale structure aids in the interpretability of posterior inference. We investigate the properties of the proposed model, present a strategy for posterior computation, demonstrate its effectiveness in various inferential tasks, and explore its applicability to large-dimensional data. The key contributions of this work can be summarized as follows:

- i) We prove that the proposed model possesses desirable theoretical features, including full support on the space of probability distributions on networks and posterior consistency.

- ii) We develop an efficient Gibbs sampler, relying on the availability of closed-form full conditional distributions for the model’s elements.
- iii) We show the model is effective for various inferential tasks, such as clustering networks, estimating probability mass functions, and making predictions.
- iv) Through simulations, we evaluate our method’s performance against existing approaches from the literature in tasks such as clustering and probability mass function estimation.
- v) We apply our method to a dataset from the HNU1 study (Zuo et al., 2014), which includes diffusion magnetic resonance imaging (dMRI) from multiple subjects, to demonstrate its practical application.
- vi) To handle clustering in populations of networks with many nodes, we introduce an approximate solution called consensus subgraph clustering.

The remainder of the article is organized as follows. In [Section 2](#), we introduce a new model for multiple network data and study its main theoretical properties. [Section 3](#) outlines a strategy for posterior computation and discusses its implementation. [Section 4](#) presents extensive simulation studies comparing our model with existing methods. In [Section 5](#), our method is applied to the analysis of the HNU1 human brain network dataset. [Section 6](#) introduces an approximate computational strategy for clustering large-dimensional datasets. Concluding remarks are presented in [Section 7](#). Proofs and additional results are provided as Supplementary Material (Barile et al., 2026).

2 Modeling strategy

2.1 Preliminaries

A simple undirected labeled binary graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a set of labeled vertices \mathcal{V} and a set of edges $\mathcal{E} \subseteq \{(v_1, v_2) \in \mathcal{V} \times \mathcal{V} : v_1 \neq v_2\}$, that is a subset of the set of pairs of distinct nodes. Given \mathcal{V} , we let $\mathcal{G}_{\mathcal{V}}$ denote the set of all simple undirected labeled binary graphs with nodes \mathcal{V} , or graph space. While $\mathcal{G}_{\mathcal{V}}$ is the main focus of this work, it is worth noting that both the modeling and computational strategies we introduce can be readily extended to directed graphs and graphs with self-relations. Given this, and for simplicity, we will henceforth refer to the elements of $\mathcal{G}_{\mathcal{V}}$ simply as graphs or networks, with a slight abuse of terminology. We observe that, if $N = |\mathcal{V}|$, then $|\mathcal{G}_{\mathcal{V}}| = 2^M$, where $M = \binom{N}{2}$ is the maximum number of edges that a graph with N nodes may feature. A graph \mathcal{G} can be represented by an $N \times N$ adjacency matrix $A_{\mathcal{G}}$ such that $A_{\mathcal{G}[ij]} = 1$ if $\{i, j\} \in \mathcal{E}$, and $A_{\mathcal{G}[ij]} = 0$ otherwise, where the subscript $[ij]$ is used to indicate the element in position (i, j) of a matrix. The symmetry of the adjacency matrix $A_{\mathcal{G}}$ follows from the fact that \mathcal{G} is assumed undirected. Finally, we assume that, throughout this work, all random variables are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

We consider a dataset $\mathcal{G}^{(1:n)} = \{\mathcal{G}_1, \dots, \mathcal{G}_n\}$ of multiple networks, that is a collection of multiple observations of networks with nodes $\mathcal{V} = \{1, \dots, N\}$. In other terms, for any $l = 1, \dots, n$, we have $\mathcal{G}_l = (\mathcal{V}, \mathcal{E}_l)$. This type of data is common, for instance,

in medical imaging, with brain regions assigned to the nodes of a graph according to an atlas, and edges representing the connections recorded among regions. Modeling an observation \mathcal{G}_l is equivalent to modeling the M -dimensional vector $\text{vech}(A_{\mathcal{G}_l})$ defined as the half-vectorization of $A_{\mathcal{G}_l}$, whose components coincide with the elements of the lower triangular half of $A_{\mathcal{G}_l}$. Given the finite dimensionality of $\mathcal{G}_{\mathcal{V}}$, a set $\mathcal{G}^{(1:n)}$ of random graphs, each taking values in $\mathcal{G}_{\mathcal{V}}$, can be modeled using a categorical distribution. This can be achieved, for example, by employing latent class models (Goodman, 1974), which require the selection of an appropriate number of classes, or by resorting to the nonparametric tensor factorization model for multivariate unordered categorical data proposed by Dunson and Xing (2009). We pursue a different strategy by building upon the notion of a representative category to capture the presence of common underlying structures shared among different network configurations. The model we propose is based on the idea that distributions on the graph space are conveniently parameterized in terms of a mean, induced by a specified metric, and a measure of the distribution’s dispersion around this mean. To this end, the notion of Fréchet mean is particularly useful as it generalizes the first moment to non-Euclidean settings, providing a tool to identify a measure of central tendency with respect to the specified metric. Assessing similarities among networks based on their global or local characteristics naturally allows us to evaluate variability in the space of graphs. Our modeling strategy thus requires the specification of a metric on the graph space to appropriately define network structural similarity (Donnat and Holmes, 2018) and to map non-Euclidean objects to Euclidean spaces. The Hamming distance (Hamming, 1950), a special instance of the broader class of graph-edit distances, is arguably the simplest distance metric between two graphs. Defined as

$$d_{\text{H}}(\mathcal{G}_1, \mathcal{G}_2) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathbb{1}_{\{A_{\mathcal{G}_1[ij]} \neq A_{\mathcal{G}_2[ij]}\}},$$

the Hamming distance between two graphs measures the number of edge deletions and insertions necessary to transform one graph into another, thus capturing local changes. Working with the Hamming distance implies that all additions and deletions are assumed to have equivalent importance. When compared to other distances, e.g. spectral distances, the Hamming distance offers the notable advantage of allowing the use of standard combinatorial tools. This proves useful for both specifying families of probability distributions for graphs and carrying out efficient posterior inference. An example of a probability distribution for unordered categorical data defined using the Hamming distance is provided in Argiento et al. (2024). In the network literature, a flexible distribution for random graphs, defined using the Hamming distance, is the centered Erdős–Rényi (CER) (Lunagómez et al., 2021). The CER distribution arises as the product of independent but not identically distributed Bernoulli probability mass functions. The probability of an edge connecting two given nodes is either α or $1 - \alpha$, depending on whether an edge is connecting the same pair of nodes of a central graph \mathcal{C} , which can be interpreted as graph mode. As a result, the scale of variation parameter α drives the variability around \mathcal{C} . Formally, for any $i < j$, $\mathbb{P}(A_{\mathcal{G}[ij]} \neq A_{\mathcal{C}[ij]}) = 1 - \mathbb{P}(A_{\mathcal{G}[ij]} = A_{\mathcal{C}[ij]}) = \alpha$. The joint distribution of the M components of $\text{vech}(A_{\mathcal{G}})$ thus leads to an equivalent probability mass function for \mathcal{G} , that is

$$p_{\text{CER}}(\mathcal{G}; \mathcal{C}, \alpha) = \alpha^{d_{\text{H}}(\mathcal{G}, \mathcal{C})} (1 - \alpha)^{M - d_{\text{H}}(\mathcal{G}, \mathcal{C})}. \quad (1)$$

A random graph \mathcal{G} , taking values in $\mathcal{G}_{\mathcal{V}}$ and with probability mass function (1), is said to have CER distribution with location parameter $\mathcal{C} \in \mathcal{G}_{\mathcal{V}}$ and scale of variation parameter $\alpha \in (0, 1/2)$. We use the notation $\mathcal{G} \sim \text{CER}(\mathcal{C}, \alpha)$. Although (1) is well defined for any $\alpha \in (0, 1)$, restricting the set of values that α can take to $(0, 1/2)$ ensures that the resulting distribution is unimodal, with mode at \mathcal{C} . That is, if $d_{\text{H}}(\mathcal{G}_1, \mathcal{C}) > d_{\text{H}}(\mathcal{G}_2, \mathcal{C})$ then $p_{\text{CER}}(\mathcal{G}_2; \mathcal{C}, \alpha) > p_{\text{CER}}(\mathcal{G}_1; \mathcal{C}, \alpha)$, formalizing the idea that graphs closer to the graph mode are more likely. The CER distribution serves as the building block of the flexible Bayesian model that we introduce next.

2.2 A Bayesian nonparametric model

We introduce a Bayesian nonparametric model for networks, defined as a nonparametric location-scale mixture of CER kernels. The CER kernel function, denoted as $\psi(\cdot; \cdot)$, is defined on $\mathcal{G}_{\mathcal{V}} \times \Theta$, with $\Theta = \mathcal{G}_{\mathcal{V}} \times (0, 1/2)$. Specifically, $\psi(\mathcal{G}; \vartheta = (\mathcal{C}, \alpha)) = p_{\text{CER}}(\mathcal{G}; \mathcal{C}, \alpha)$, where p_{CER} is defined in (1). For simplicity, we define the nonparametric mixture model using the distribution of a Dirichlet process (DP). However, the posterior computation strategies presented in this work are easily adapted to more general mixing measures, as long as their predictive distribution is available in closed form. This is the case, for example, with the class of Gibbs-type priors (De Blasi et al., 2015), and, following the introduction of a suitable auxiliary random variable, with the class of normalized random measures with independent increments (Regazzini et al., 2003).

Definition 2.1 (Location-scale DP mixture of CER kernels). The location-scale DP mixture of CER kernels on $\mathcal{G}_{\mathcal{V}}$ is the random probability mass function \tilde{f} defined as

$$\tilde{f}(\cdot) = \int_{\Theta} \psi(\cdot; \vartheta) d\tilde{P}(\vartheta), \quad (2)$$

where $\Theta = \mathcal{G}_{\mathcal{V}} \times (0, 1/2)$ and \tilde{P} is distributed as a DP with base measure P_0 on Θ , and concentration parameter $c > 0$.

Following the introduction of the latent variables $\vartheta^{(1:n)} = \{\vartheta_1, \dots, \vartheta_n\}$, the same model can be expressed in hierarchical form as

$$\begin{aligned} \mathcal{G}_l &| \vartheta^{(1:n)} \stackrel{\text{iid}}{\sim} \psi(\mathcal{G}_l; \vartheta_l) & l = 1, \dots, n \\ \vartheta_l = (\mathcal{C}_l, \alpha_l) &| \tilde{P} \stackrel{\text{iid}}{\sim} \tilde{P} & l = 1, \dots, n \\ \tilde{P} &\sim \text{DP}(c, P_0). \end{aligned} \quad (3)$$

The DP mixture of CER kernels is completed by specifying the base measure P_0 , which we define as the joint distribution of $\vartheta = (\mathcal{C}, \alpha)$ for which

$$\begin{aligned} \alpha &\sim \text{TBeta}(1/2; a, b) \\ \mathcal{C} &| \alpha \sim \text{CER}(\mathcal{G}_0, \alpha), \end{aligned} \quad (4)$$

for some hyperparameters $a, b > 0$, and $\mathcal{G}_0 \in \mathcal{G}_{\mathcal{V}}$. For $q \in (0, 1)$, $\text{TBeta}(q; a, b)$ in (4) denotes the Truncated-Beta distribution on $(0, q)$, whose probability density function is

given by

$$f_{\text{TBeta}}(\alpha; q, a, b) = \frac{\alpha^{a-1} (1 - \alpha)^{b-1}}{\mathcal{B}(q; a, b)},$$

where $\mathcal{B}(q; a, b) = \int_0^q \alpha^{a-1} (1 - \alpha)^{b-1} d\alpha$ indicates the incomplete beta function. We emphasize that restricting the support of the component-specific dispersion parameters α_l to the interval $(0, 1/2)$ is crucial for ensuring that the CER kernel is unimodal, with its mode at \mathcal{C}_l . In the context of mixture modeling, this property is particularly appealing, as it makes the inferred clusters easily interpretable. We further note that a prior distribution can be assigned to the hyperparameters (a, b, c, \mathcal{G}_0) appearing in the Truncated-Beta/CER specification of P_0 .

Remark 1. The proposed model constitutes a non-trivial generalization of the CER model in Lunagómez et al. (2021). While the CER model is parametric and unimodal, our approach is explicitly designed to capture heterogeneity in populations of networks, without imposing a fixed number of modes. This is achieved by placing a nonparametric prior on both the location and scale parameters of the CER kernel and by introducing a Truncated-Beta/CER base measure, which differs from the prior in Lunagómez et al. (2021) and substantially improves computational tractability. Such tractability plays a central role in this work, as it allows for the analytical marginalization of the DP and the formulation of a marginal algorithm for posterior sampling. We also note a close connection between our approach and the mixture models of Young et al. (2022) and Mantziou et al. (2024). Both works propose parametric, multimodal models with kernel structures closely related to the CER kernel considered here, differing mainly by the inclusion of two component-specific dispersion parameters. Moreover, in Mantziou et al. (2024) the modes are constrained by a stochastic block model structure. In contrast, our approach neither imposes a specific structural form on the modes nor fixes their number. A distinctive feature of our proposal is that the number of mixture components is allowed to grow unboundedly with the number of observed graphs. This becomes evident when considering the infinite-sum representation of $\tilde{f}(\cdot)$ obtained by substituting the stick-breaking representation of \tilde{P} (Sethuraman, 1994) into (2) and goes beyond a mere technicality, playing a central role in establishing the desirable theoretical properties of our model, discussed in the next section.

2.3 Kullback–Leibler property and posterior consistency

We show that the location-scale DP mixture of CER kernels, introduced in Definition 2.1 and with base measure (4), has full support in the Kullback–Leibler sense. Specifically, for any $\varepsilon > 0$, the prior induced by \tilde{f} assigns positive probability to the Kullback–Leibler neighborhood $\mathbb{B}_\varepsilon(p_*) = \{p \in \mathcal{P}_{\mathcal{G}_V} : \text{KL}(p_*; p) \leq \varepsilon\}$ of any probability mass function $p_* \in \mathcal{P}_{\mathcal{G}_V}$, where $\text{KL}(p_*; p)$ denotes the Kullback–Leibler divergence between the probability mass functions p_* and p , and $\mathcal{P}_{\mathcal{G}_V}$ denotes the space of all probability distributions on \mathcal{G}_V . This property, also known as the Kullback–Leibler property, is appealing as it formalizes the idea that any distribution in $\mathcal{P}_{\mathcal{G}_V}$ can be approximated arbitrarily well by a set of realizations of \tilde{f} with positive prior probability. For Bayesian modeling, it

is convenient to use priors with the Kullback–Leibler property, hence nonparametric, especially when there is no conclusive prior information about the parametric shape of the distribution generating the data (Walker et al., 2004). The Kullback–Leibler property is also key in studying the large n behavior of the posterior distribution of \tilde{f} , given $\mathcal{G}^{(1:n)}$. Considering the finiteness of $\mathcal{G}_{\mathcal{V}}$, the Kullback–Leibler property implies that the posterior distribution of \tilde{f} is strongly consistent at any $p_* \in \mathcal{P}_{\mathcal{G}_{\mathcal{V}}}$. These properties are formalized in [Theorem 2.1](#) and [Corollary 2.1](#), with proofs provided in the Supplementary Material (Barile et al., 2026).

Theorem 2.1. *The prior Π induced by a location-scale DP mixture of CER kernels with base measure as in (4) has the Kullback–Leibler property. That is, for any $p_* \in \mathcal{P}_{\mathcal{G}_{\mathcal{V}}}$ and any $\varepsilon > 0$, $\Pi(\mathbb{B}_{\varepsilon}(p_*)) > 0$.*

Although [Theorem 2.1](#) explicitly refers to the base measure in (4), its proof only relies on the fact that P_0 has full support on Θ . Therefore, the Kullback–Leibler property extends to any specification of the base measure with full support on Θ .

Corollary 2.1. *The posterior distribution $\Pi_n(\cdot \mid \mathcal{G}^{(1:n)})$ of a location-scale DP mixture of CER kernels with base measure as in (4), given $\mathcal{G}^{(1:n)}$, is strongly consistent at any $p_* \in \mathcal{P}_{\mathcal{G}_{\mathcal{V}}}$. That is, for any $p_* \in \mathcal{P}_{\mathcal{G}_{\mathcal{V}}}$, $\Pi_n(U_{p_*}^c \mid \mathcal{G}^{(1:n)}) \rightarrow 0$ almost surely, as $n \rightarrow \infty$, for any neighborhood U_{p_*} of p_* .*

It should be noted that, while the literature on the asymptotic properties of network models has traditionally focused on the regime where the number of nodes $N \rightarrow \infty$ within a single network (see, e.g., Tang et al., 2013), another interesting question arises when N is fixed and $n \rightarrow \infty$. In this setting, Josephs et al. (2023) recently established posterior consistency for their graph classifier. [Corollary 2.1](#) contributes to this line of research by addressing the consistent estimation of the entire population distribution for a network-valued random variable. This theoretical result is complemented by the simulation study presented in [Section 4.2](#), which investigates how the posterior estimate concentrates around its true value as a function of sample size n .

3 Posterior computation

We adapt the marginal algorithm from Escobar and West (1995), originally introduced for location-scale DP mixtures of univariate Gaussian kernels, to mixtures of CER kernels, as specified in [Definition 2.1](#) and (4). Posterior sampling is achieved through a Gibbs sampler, following the analytical marginalization of the DP \tilde{P} . Despite the inherently complex structure of $\mathcal{G}_{\mathcal{V}}$, the distributions involved in the algorithm are available in closed form, which conveniently simplifies posterior sampling. The algorithm consists of sequential Gibbs updates of the individual location-scale parameters $\vartheta_l = (\mathcal{C}_l, \alpha_l)$, for $l = 1, \dots, n$, from their full conditional distribution

$$\mathbb{P}\left(\vartheta_l \in \cdot \mid \vartheta_{(-l)}^{(1:n)}, \mathcal{G}^{(1:n)}\right) = \pi_{l0} P_l(\cdot) + \sum_{k=1}^{K_{(-l)}} \pi_{lk} \delta_{\vartheta_{k(-l)}^*}(\cdot), \quad (5)$$

where the subscript $(-l)$ denotes quantities computed after removing ϑ_l from $\vartheta^{(1:n)}$. Equation 5 represents the celebrated generalized Pólya urn scheme of Blackwell and MacQueen (1973), formalizing the idea that ϑ_l can either coincide with any of the distinct values in $\vartheta_{(-l)}^{(1:n)}$, that is $\vartheta_{k(-l)}^*$, with probability π_{lk} , for $k = 1, \dots, K_{(-l)}$, or take a new value with probability π_{l0} . The probabilities in (5) are given, up to a proportionality constant, by

$$\pi_{l0} \propto c \sum_{r=0}^{M-d_l} w_{lr} \frac{\mathcal{B}(1/2; a_{lr}, b_{lr})}{\mathcal{B}(1/2; a, b)} \quad (6)$$

$$\pi_{lk} \propto n_{k(-l)} \psi(\mathcal{G}_l; \vartheta_{k(-l)}^*) \quad k = 1, \dots, K_{(-l)}, \quad (7)$$

where $d_l = d_{\mathbb{H}}(\mathcal{G}_0, \mathcal{G}_l)$, $w_{lr} = 2^{d_l} \binom{M-d_l}{r}$, $a_{lr} = a + 2r + d_l$ and $b_{lr} = b + 2M - 2r - d_l$. See Section S2.1 in the Supplementary Material (Barile et al., 2026) for the derivation of (6). The distribution P_l of new values for ϑ_l , conditionally on \mathcal{G}_l , is proportional to $\psi(\mathcal{G}_l; \vartheta) dP_0(\vartheta)$. Sampling from P_l translates into sampling from a mixture of Truncated-Beta distributions, and from M independent Bernoulli distributions. Specifically, $\vartheta_l = (\mathcal{C}_l, \alpha_l) \mid \mathcal{G}_l \sim P_l$ can be expressed as

$$\alpha_l \mid \mathcal{G}_l \sim \sum_{r=0}^{M-d_l} \varphi_{lr} \text{TBeta}(1/2; a_{lr}, b_{lr}) \quad (8)$$

$$A_{\mathcal{C}_l[ij]} \mid \alpha_l, \mathcal{G}_l \stackrel{\text{ind}}{\sim} \text{Bern}(p_{lij}) \quad i < j, \quad (9)$$

where $\varphi_{lr} \propto w_{lr} \mathcal{B}(1/2; a_{lr}, b_{lr})$ and

$$p_{lij} = \left(1 + \left(\frac{\alpha_l}{1 - \alpha_l} \right)^{2(A_{\mathcal{G}_0[ij]} + A_{\mathcal{G}_l[ij]} - 1)} \right)^{-1}. \quad (10)$$

The probability p_{lij} of generating a graph mode with an edge connecting the nodes $\{i, j\}$ thus depends on whether \mathcal{G}_0 and \mathcal{G}_l display such an edge. The left panel of Figure 2 illustrates the dependence of p_{lij} on $A_{\mathcal{G}_0[ij]} + A_{\mathcal{G}_l[ij]}$ and α_l . The derivation of the characterization of P_l in (8)–(10) is reported in Section S2.3 of the Supplementary Material (Barile et al., 2026).

To improve the mixing of the algorithm, it is key to introduce a reshuffling step that independently updates the cluster labels, i.e., the values $\vartheta_k^* = (\mathcal{C}_k^*, \alpha_k^*)$ for $k = 1, \dots, K$, after each Gibbs sampling iteration (Bush and MacEachern, 1996). We let $\mathcal{D}_k = \{l \in \{1, \dots, n\} : \vartheta_l = \vartheta_k^*\}$ be 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 $\mathcal{D} \subseteq \{0, 1, \dots, n\}$, we introduce the notation $\mathcal{G}^{(\mathcal{D})} = \{\mathcal{G}_l : l \in \mathcal{D}\}$, 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 $\{i, j\}$. Similarly to P_l , updating ϑ_k^* from its full conditional distribution translates into sampling from a mixture of Truncated-Beta distributions, and from M independent Bernoulli distributions. Specifically, after introducing the quantities $D_k^* = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \max\{n_{ij}^{(k)}, n_k + 1 - n_{ij}^{(k)}\}$ and

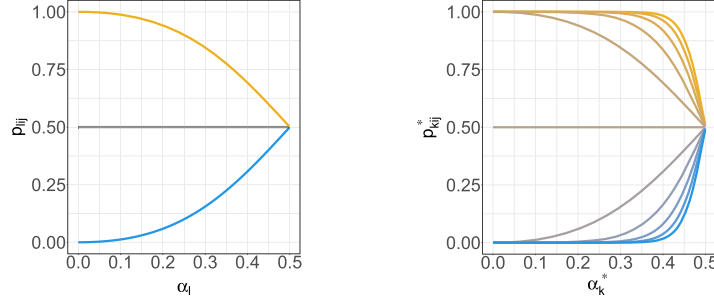


Figure 2: Left panel: probability p_{lij} in (10), with $A_{\mathcal{G}_0[ij]} + A_{\mathcal{G}_l[ij]} \in \{0, 1, 2\}$ (blue for 0, gray for 1, and yellow for 2) and for α_l ranging in $(0, 1/2)$. Right panel: probability p_{kij}^* in (13), with $n_{ij}^{(k)} \in \{0, 1, \dots, n_k + 1\}$, $n_k + 1 = 10$, (blue for low, yellow for high) and for α_k^* ranging in $(0, 1/2)$.

$d_k^* = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \min\{n_{ij}^{(k)}, n_k + 1 - n_{ij}^{(k)}\}$, we can write

$$\alpha_k^* | \mathcal{G}^{(\mathcal{D}_k)} \sim \sum_{r=0}^{D_k^* - d_k^*} \varphi_{kr}^* \text{TBeta}(\alpha_k^*; 1/2, a_{kr}^*, b_{kr}^*) \quad (11)$$

$$A_{\mathcal{C}_k^*[ij]} | \alpha_k^*, \mathcal{G}^{(\mathcal{D}_k)} \stackrel{\text{ind}}{\sim} \text{Bern}(p_{kij}^*), \quad i < j. \quad (12)$$

The Bernoulli parameters in (12) are defined as

$$p_{kij}^* = \left(1 + \left(\frac{\alpha_k^*}{1 - \alpha_k^*} \right)^{2(n_{ij}^{(k)} - (n_k + 1)/2)} \right)^{-1}. \quad (13)$$

The probability p_{kij}^* to generate a graph mode with an edge connecting the nodes $\{i, j\}$ is increasing in the number $n_{ij}^{(k)}$ of graphs in \mathcal{D}_k^\dagger displaying such an edge. Moreover, p_{kij}^* is increasing in α_k^* if $n_{ij}^{(k)} < (n_k + 1)/2$ and decreasing if $n_{ij}^{(k)} > (n_k + 1)/2$. The right panel of Figure 2 shows that as $n_{ij}^{(k)}$ increases, the probability of $A_{\mathcal{C}_k^*[ij]} = 1$ also rises, with this effect being more pronounced for small values of α_k^* . The mixture weights in (11) are given, up to a proportionality constant, by

$$\varphi_{kr}^* \propto w_{kr}^* \mathcal{B}(1/2; a_{kr}^*, b_{kr}^*), \quad (14)$$

where $a_{kr}^* = a + d_k^* + r$ and $b_{kr}^* = b + (n_k + 1)M - d_k^* - r$. In turn, the coefficients w_{kr}^* in (14) result from a generating function defined by a product of polynomials, embedding a subset-sum problem. Specifically, we have

$$w_{kr}^* = \begin{cases} \sum_{\mathcal{S}_{kr}} \prod_{h=0}^{n_k/2} \binom{M_{kh}}{s_h} & \text{if } n_k \text{ is even} \\ \sum_{\mathcal{R}_{kr}} 2^{m_k(\lfloor n_k/2 \rfloor + 1)} \prod_{h=0}^{\lfloor n_k/2 \rfloor} \binom{M_{kh}}{s_h} & \text{if } n_k \text{ is odd,} \end{cases} \quad (15)$$

where $\lfloor x \rfloor$ denotes the integer part of x , the sums in (15) are taken over the sets

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

$m_{kh} = \#\{\{i, j\} \in \mathcal{V}^2 : n_{ij}^{(k)} = h\}$ indicates the number of pairs of nodes that are connected by an edge in exactly h graphs in $\mathcal{G}(\mathcal{D}_k^{\dagger})$, $M_{kh} = m_{kh} + m_{k(n_k+1-h)}$, and $\gamma_{kh}(s_h) = (n_k + 1 - 2h)s_h + hM_{kh}$. We note that the conditions defining the sets \mathcal{S}_{kr} and \mathcal{R}_{kr} , for any $k = 1, \dots, K$ and $r = 0, \dots, D_k^* - d_k^*$, represent linear Diophantine equations with the decision variables $\{s_0, \dots, s_{\lfloor n_k/2 \rfloor}\}$ constrained by $s_h \leq M_{kh}$, for $h = 1, \dots, \lfloor n_k/2 \rfloor$. Moreover, as expected, when $n_k = 1$, that is $\mathcal{D}_k = \{l\}$ for some $l = 1, \dots, n$, the distribution in (11) and (12) simplifies to the distribution P_l , in (8) and (9), for ϑ_l , conditionally on \mathcal{G}_l and given that it takes a new value. The steps of the algorithm are summarized in Algorithm 1. See Section S2.2 in the Supplementary Material (Barile et al., 2026) for the derivation of the full conditional distribution of ϑ_k^* involved in the reshuffling step. Additionally, Section S3 of the Supplementary Material (Barile et al., 2026) provides closed-form expressions for the cluster-specific posterior distribution of \mathcal{C}_k^* and the cluster-specific m -step-ahead posterior predictive distribution, both obtained by building on the conditional distribution of ϑ_k^* in (11) and (12).

4 Simulation study

We explore the behavior of the DP mixture of CER kernels through the analysis of synthetic data. The study has two objectives: (i) assessing the model’s ability to cluster multiple network data with a known partition structure, under data-generating processes characterized by varying levels of variability; and (ii) investigating the impact of sample size on the accuracy of posterior estimates. To facilitate graphical presentation, we focus on networks with $N = 20$ nodes. Observations are sampled from a mixture of four CER components $p_*(\cdot) = \sum_{k=1}^4 0.25 p_{\text{CER}}(\cdot; \mathcal{C}_{0k}, \alpha_{0k})$, where well-defined component-specific network structures are defined through the modes \mathcal{C}_{0k} . Specifically, and in the same spirit as Durante et al. (2017), each component is centered around a network configuration, or centroid, with distinct structures: scale-free (Barabási and Albert, 1999) for \mathcal{C}_{01} , small-world (Watts and Strogatz, 1998) for \mathcal{C}_{02} , stochastic block model (Nowicki and Snijders, 2001) for \mathcal{C}_{03} , and Erdős–Rényi (Erdős and Rényi, 1960) for \mathcal{C}_{04} . This choice is designed to assess whether the proposed model can effectively cluster and estimate the distribution of a collection of multiple network data with heterogeneous underlying structures. The generated centroids are displayed in the first row of Figure 3. The specification of the parameters for the four models used to generate the centroids is summarized in Table S1 in the Supplementary Material (Barile et al., 2026). The study

Algorithm 1: Gibbs sampler for DP mixture of CER kernels

Input: Data $\mathcal{G}^{(1:n)}$; hyperparameters a, b, c, \mathcal{G}_0 ; number of iterations T ; number of burn-in iterations T_0 ;

Output: Sample from the posterior of location-scale parameters: $\{\vartheta_{[t]}^{(1:n)}\}_{t=T_0+1}^T$

Initialise $\vartheta^{(1:n)}$ randomly;

for $t \leftarrow 1$ **to** T **do**

for $l \leftarrow 1$ **to** n **do**

Remove the l th component of $\vartheta^{(1:n)}$ to obtain $\vartheta_{(-l)}^{(1:n)}$;

Let $K_{(-l)}$ be the number of distinct values in $\vartheta_{(-l)}^{(1:n)}$;

Let $(\vartheta_{1(-l)}^*, \dots, \vartheta_{K_{(-l)}(-l)}^*)$ be the set of distinct values in $\vartheta_{(-l)}^{(1:n)}$;

Compute π_{l0} , up to a constant, as in (6);

for $k \leftarrow 1$ **to** $K_{(-l)}$ **do**

Compute π_{lk} , up to a constant, as in (7);

Normalize all probabilities to get $\pi_l = (\pi_{l0}, \pi_{l1}, \dots, \pi_{lK_{(-l)}})$;

Sample *category* from Categorical($K_{(-l)} + 1, \pi_l$);

if *category* == 1 **then**

Draw parameter ϑ_l from P_l , as in (8)–(10);

else if *category* == $k \in \{2, \dots, K_{(-l)} + 1\}$ **then**

Set ϑ_l equal to $\vartheta_{k-1(-l)}^*$;

Let K be the number of distinct values in $\vartheta^{(1:n)}$;

Let $(\vartheta_1^*, \dots, \vartheta_K^*)$ be the set of distinct values in $\vartheta^{(1:n)}$;

for $k \leftarrow 1$ **to** K **do**

Update ϑ_k^* , sampling from (11)–(13);

Set $\vartheta_{[t]}^{(1:n)}$ equal to $\vartheta^{(1:n)}$;

return $\{\vartheta_{[t]}^{(1:n)}\}_{t=T_0+1}^T$;

consists of two main parts: in the first one, multiple network data are generated by considering various component-specific scales of variation while keeping the sample size fixed; in the second one, the sample size varies while the scale parameters of the data-generating models remain fixed. For both parts, we specify the parameters of the base measure (4) as follows. The centroid \mathcal{G}_0 is set, by using an empirical Bayes approach, as an element of the sample Fréchet mean set (see Lunagómez et al., 2021). Specifically, \mathcal{G}_0 is the network that has an edge between the nodes $\{i, j\}$ if and only if that edge is present in at least 50% of the networks in the dataset. The parameters a and b are both set equal to one, thus implying the prior model for the scale of variation parameter is centered at a uniform distribution on $(0, 1/2)$. Finally, the concentration parameter c is set equal to one.

For each scenario in the two simulation studies, 100 datasets are generated. Each

dataset is analyzed by running 1,200 Gibbs iterations, with the first 200 discarded as burn-in. The results produced by our model are compared with those from Durante et al. (2017), Mantziou et al. (2024), Signorelli and Wit (2020), and Josephs et al. (2025), with the latter included only in the first part of the study. Although not originally intended for this purpose, the method of Durante et al. (2017) is readily extended to address clustering problems for multiple network data. Further details on the implementation of the competing methods are provided in Section S5 of the Supplementary Material (Barile et al., 2026). This study suggests that, overall, our model performs comparably to or better than state-of-the-art methods in two key aspects: effectively modeling a population of networks with heterogeneous characteristics and accurately clustering the elements of a network population. An additional simulation experiment is reported in Section S4 of the Supplementary Material (Barile et al., 2026), where the data-generating process is again specified as a mixture model, but with one component characterized by a core-periphery structure. This setting induces more intricate connectivity patterns and may arise, for instance, from a non-assortative stochastic block model. The performance of our model appears robust to this more complex scenario.

4.1 Data-generating models with varying scales of variation

To assess the ability of a method to cluster multiple network data, we compare the estimated partition to the true partition, which reflects the four-component mixture structure of the data-generating model. We resort to three metrics: the adjusted Rand index (ARI), clustering entropy and clustering purity. A point estimate for the data partition is obtained from the posterior samples produced by Algorithm 1, by minimizing the posterior expected Variation of Information, as implemented in the `Salso` R package (Dahl et al., 2022). We investigate the robustness of our method in clustering multiple network data generated from models characterized by different levels of variability. To this end, we fix a sample size of $n = 40$ and focus on four scenarios with increasing scale of variation parameters shared across all components. We also consider a more realistic scenario with different scales of variation for each of the four components of the data-generating model. The values of the component-specific scales of variation for these scenarios are reported in Table 1. The results of our investigations are displayed in Figure 4. According to the considered metrics, our model outperforms all the competing methods across all scenarios, showing higher values for ARI and clustering

Level of variability	α_{01}	α_{02}	α_{03}	α_{04}
low	0.25	0.25	0.25	0.25
medium-low	0.30	0.30	0.30	0.30
medium	0.35	0.35	0.35	0.35
high	0.40	0.40	0.40	0.40
mixed	0.25	0.35	0.30	0.40

Table 1: Definition of five simulation scenarios through the specification of the scale of variation parameters $\{\alpha_{01}, \dots, \alpha_{04}\}$ of the four CER components of p_* . See Section 4.

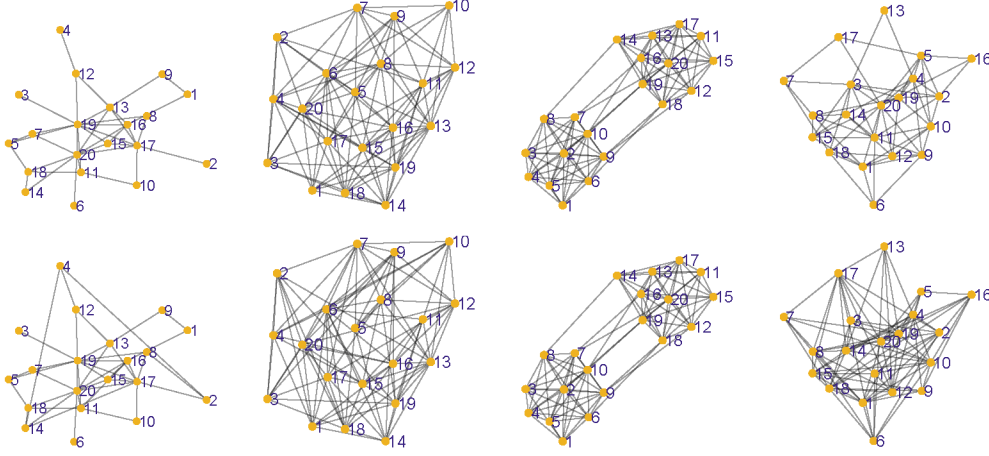


Figure 3: Top row: centroids with Scale-free (\mathcal{C}_{01}), Small-world (\mathcal{C}_{02}), Stochastic Block Model (\mathcal{C}_{03}), and Erdős-Rényi (\mathcal{C}_{04}) structures (from left to right). Bottom row: posterior Fréchet means for the four clusters estimated based on a dataset generated from the mixed level of variability scenario, with sample size $n = 40$. See [Section 4](#).

purity and lower values for clustering entropy. As expected, scenarios characterized by higher levels of variability are more challenging for all methods. We also observe the unsatisfactory performance of Josephs et al. (2025), even in relatively simple scenarios. This is likely due to the method overlooking node correspondence across layers, as it is designed for multiplex networks in the unlabeled setting. Consequently, we chose not to include this method as a competitor in the subsequent illustrations. It is also instructive to explore the properties of the clusters identified by our method. Given the estimated partition and denoting by \hat{K} the corresponding number of clusters, we can produce cluster-specific point estimates for the centroids by looking at the Fréchet mean of the posterior distribution of \mathcal{C}_k^* , for $k = 1, \dots, \hat{K}$. Specifically, we sample from the posterior distribution of ϑ_k^* in (11) and (12), and consider the sample Fréchet mean of \mathcal{C}_k^* . We henceforth refer to this as to the cluster-specific posterior Fréchet mean. Alternatively, one can resort to Equation (S27) in the Supplementary Material (Barile et al., 2026). For illustrative purposes, we focus on a randomly selected dataset generated from the mixed variability scenario. The second row of [Figure 3](#) shows the posterior Fréchet mean for centroids of the four estimated clusters. The topological structures of these estimates align with those of the four-component data-generating mixture model, with the correspondence between true and estimated components identified based on the frequency of observations generated from a given component in the estimated clusters. Moreover, the point estimates of the cluster-specific scale parameters, that is $\{0.262, 0.337, 0.295, 0.397\}$, reflect the heterogeneity in variability levels that characterize the data-generating model.

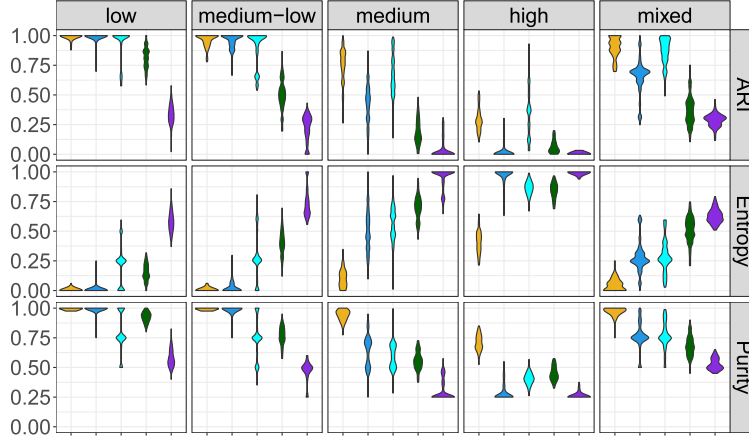


Figure 4: Adjusted Rand index, entropy and purity, for our method (yellow violins) and the methods of Durante et al. (2017) (blue violins), Mantziou et al. (2024) (cyan violins), Signorelli and Wit (2020) (green violins) and Josephs et al. (2025) (violet violins). Columns refer to the scenarios of Table 1. Distributions are estimated based on the analysis of 100 datasets. See Section 4.1.

4.2 Varying sample size

We study the accuracy of the posterior mean $\hat{f} = \mathbb{E}[\tilde{f} \mid \mathcal{G}^{(1:n)}]$ as an estimator of the true data-generating distribution p_* , with \hat{f} evaluated based on the posterior sample generated from Algorithm 1. Specifically, we investigate how this accuracy changes for different sample sizes n . This study aims to provide a finite-sample analogue to the strong consistency property of the DP mixture of CER kernels, as reported in Corollary 2.1, which states that for any $\varepsilon > 0$ and a given metric d on $\mathcal{P}_{\mathcal{G}_V}$, $\mathbb{P}(d(p_*, \hat{f}) > \varepsilon \mid \mathcal{G}^{(1:n)}) \rightarrow 0$, almost surely, as $n \rightarrow \infty$. Using the Kullback–Leibler divergence, we 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 $\text{KL}(p_*; \hat{f})$ requires summation over the graph space \mathcal{G}_V , which is prohibitive even for moderate N . Thus, we propose an importance-sampling approximation of $\text{KL}(p_*; \hat{f})$:

$$\text{KL}(p_*; \hat{f}) = \sum_{\mathcal{G} \in \mathcal{G}_V} p_*(\mathcal{G}) \log \left(\frac{p_*(\mathcal{G})}{\hat{f}(\mathcal{G})} \right) = \mathbb{E}_{p_*} \left[\log \left(\frac{p_*(\mathcal{G})}{\hat{f}(\mathcal{G})} \right) \right] \approx \frac{1}{L} \sum_{l=1}^L \log \left(\frac{p_*(\mathcal{G}_l)}{\hat{f}(\mathcal{G}_l)} \right),$$

with $\mathcal{G}_l \stackrel{\text{iid}}{\sim} p_*$, for $l = 1, \dots, L$. The results of our study are illustrated 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). Similar results, focusing on the \mathbb{L}^1 distance between p_* and \hat{f} , are presented in Section S4 of the Supplementary Material (Barile et al., 2026).

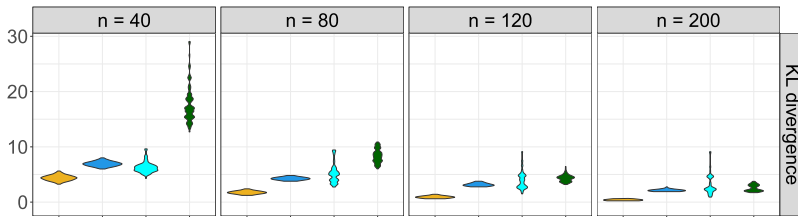


Figure 5: Importance-sampling approximate distributions of $\text{KL}(p_*; \hat{f})$ for our method (yellow violins), and the methods of Durante et al. (2017) (blue violins), Mantziou et al. (2024) (cyan violins) and Signorelli and Wit (2020) (green violins). Distributions are estimated based on the analysis of 100 datasets. See Section 4.2.

5 Analysis of human brain networks data

We analyze the popular HNU1 human brain dataset, publicly available at https://networks.skewed.de/net/human_brains (Peixoto, 2023), from the Consortium for Reliability and Reproducibility (CoRR) repository (Zuo et al., 2014). Connectivity patterns across different brain regions were measured for 30 healthy individuals at rest. Up to 10 measurements per individual were taken using diffusion magnetic resonance imaging (dMRI) over a month, totaling $n = 266$ network observations. These measurements are represented as labeled networks with $N = 48$ nodes corresponding to fixed brain regions of interest (ROI), defined by the JHU ICBM DTI-81 atlas (Mori et al., 2005), and edges denoting connections among these regions. Two regions are considered connected if at least one white matter fiber links them. Importantly, fiber-tracking pipelines are subject to measurement errors. Figure 1 displays a sample of six observations from this dataset. The same dataset, though with different node granularity, has been discussed by Zuo et al. (2014), Arroyo et al. (2021), Lunagómez et al. (2021), and Mantziou et al. (2024). The latter three studies analyze the data from a modeling perspective. Arroyo et al. (2021) investigate their method’s ability to identify individual differences based on network communities. Lunagómez et al. (2021) assume unimodality in the network generation process and infer a representative network for the population, while Mantziou et al. (2024) focus on detecting outlier networks. Our analysis aims to characterize differences in brain connectivity between subjects in the dataset. We compare our model’s results with those obtained using the methods of Durante et al. (2017) and Mantziou et al. (2024). We do not include the method of Signorelli and Wit (2020) here, as the code made available by the authors restricts mixtures to a maximum of 7 components, making it unsuitable for the current problem. Methods are first compared using posterior predictive checks, to assess their ability to recover the generative mechanism underlying the observed graphs for selected network summary measures. We simulate networks from the posterior predictive distribution, which for our model is given in Equation (S21) in the Supplementary Material (Barile et al., 2026), and compute network summary measures for these simulations. If the model lacks flexibility, we expect the observed data’s network measures to fall in the tails of their corresponding posterior predictive distributions. Figure 6 shows that the DP mixture of CER kernels,

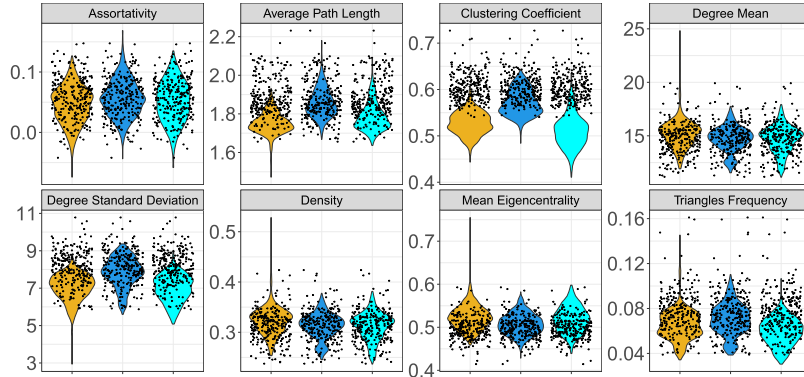


Figure 6: Posterior predictive checks. Posterior predictive distribution for selected network summary measures, for our method (yellow violins) and the methods of Durante et al. (2017) (blue violins) and Mantziou et al. (2024) (cyan violins). Jittered dots represent the network summary measures computed for the observed brain network data. See Section 5.

the method of Durante et al. (2017), and the approach of Mantziou et al. (2024) are all sufficiently flexible to capture the variability of functionals of the posterior predictive distribution. Some network summary measures, however, such as transitivity and average path length, indicate a comparatively better fit for the model proposed by Durante et al. (2017).

Since edges represent physical connections between ROIs, the partition of observations based on the 30 subjects in the study can serve as a proxy for the true clustering structure of the data. We thus investigate whether brain scans of the same subject tend to be assigned to the same cluster and thus can be considered similar, a relevant question for researchers in neuroscience. We do this by comparing the estimated partition of the sample of network data, with the partition implied by the presence of 30 subjects. Clustering metrics reported in Table 2 clearly indicate that our model detects similarities among the brain scans of the same individual, further validating its effectiveness in consistently clustering networks. Our model outperforms the methods of Durante et al. (2017) and Mantziou et al. (2024) in terms of clustering accuracy. The optimal partition

Model	\hat{K}	Adjusted Rand Index	Entropy	Purity
DPM-CER	50	0.8065	0.0065	0.9925
Durante et al.	23	0.6822	0.1418	0.7143
Mantziou et al.	57	0.7508	0.0278	0.9511

Table 2: Clustering results of the Human Brain data set in terms of number of inferred clusters and clustering metrics with respect to the natural partition implied by individuals. Findings are compared with the models of Durante et al. (2017) and Mantziou et al. (2024). See Section 5.

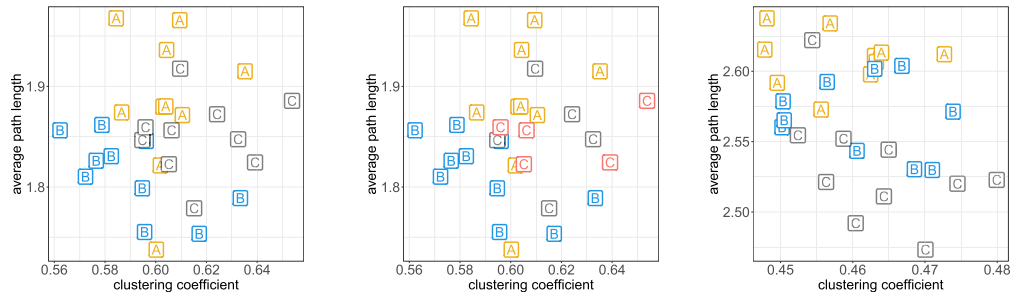


Figure 7: Scatter plots for the small-world properties of brain networks for three subjects in the dataset. Colors indicate the cluster membership, letters refer to the subject ID in the dataset, namely 0025443 (A), 0025445 (B) and 0025446 (C). Left panel: $N = 48$, partition estimated via DP mixture of CER kernels (see Section 5). Central and right panels: $N = 48$ and $N = 200$ respectively, partition estimated via consensus subgraph clustering (see Section 6).

identified by our model consists of 50 clusters, thus exceeding the number of subjects in the study. Only two clusters contain networks from different individuals, while for four subjects, the networks corresponding to the same subject are distributed across multiple clusters. These results hold potential biological significance, offering valuable insights for further investigation.

We conclude our analysis by evaluating whether the clusters identified by our model display features with neuroscientific interpretability. A similar question is addressed by Mantziou et al. (2024), who identify a subgroup of individuals with brain connectivity patterns distinct from the majority, based on network summary measures of interest to neuroscientists. We focus on average path length and clustering coefficient, two metrics of neuroscientific significance as human brains are known to typically exhibit a small-world structure, characterized by short average path lengths and high clustering coefficients (Bassett and Bullmore, 2006).

The left panel of Figure 7 shows the average shortest path length and clustering coefficient for the brain networks, for the three largest clusters of the estimated partition. The three clusters have size 10 and coincide with the scans of three subjects in the dataset. There is a clear distinction between the brain scans in each cluster with respect to these two network properties. This confirms our model’s ability to identify differences among data characterized by topological structures indicative of small-world behavior.

6 Consensus subgraph clustering for large networks

Probabilistic models for network data often become computationally infeasible when the number of nodes is large (Jordan, 2004). This issue is exacerbated when dealing with multiple network data. Our method, for instance, becomes computationally intensive as N grows, requiring numerical evaluations that significantly slow down Algorithm 1.

In addition, the computational burden increases with n . To address the challenge of clustering elements within a population of networks when the number of nodes is large, we propose a heuristic approach inspired by consensus clustering techniques (Strehl and Ghosh, 2002), which we call *consensus subgraph clustering*. Similar to variational methods, our approach breaks down some dependencies between nodes and measures similarity among networks based on their local characteristics. This is achieved by running our model-based clustering method in parallel on subgraph observations. Subgraphs are created by partitioning the N nodes into blocks of at most N_{sub} nodes. This step is akin to assuming a block structure at the vertex level, with block memberships possibly assigned based on available information, e.g. spatial, on the nodes. Each Gibbs sampler produces a sample of partitions of the multiple network data from the posterior distribution of the model, conditional on the subgraphs obtained by restricting the original data to specific node blocks. The subgraph-specific posterior samples are then pooled into a unique sample, from which we identify a representative partition by minimizing the posterior expected Variation of Information. We illustrate this strategy through the analysis of a version of the human brain network data analyzed in Section 5, with $N = 200$ and thus characterized by finer node granularity.

6.1 Brain network data with finer granularity

Constructed from the same $n = 266$ dMRI scans of 30 healthy individuals already analyzed in Section 5, the dataset we consider is based on the CC200 human brain atlas (Craddock et al., 2012), which includes $N = 200$ ROIs, and is available in the same repository. Compared to the 48 ROIs considered in Section 5, this version presents a substantially higher dimensionality, posing a greater computational challenge. We implement the described consensus subgraph clustering approach and start by investigating the effect of N_{sub} on the estimated data partition. We consider subgraphs defined on mutually exclusive vertex sets with cardinality at most equal to $N_{\text{sub}} \in \{5, 10, 15, 20, 25, 30, 35, 40, 50\}$, which leads, respectively, to $m_{\text{sub}} \in \{40, 20, 14, 10, 8, 7, 6, 5, 4\}$ distinct datasets of multiple network data of dimension at most N_{sub} . Nodes are divided into m_{sub} blocks based on the physical distance between ROIs, using a balanced clustering technique (Grötschel and Wakabayashi, 1989), as implemented in the `antclust` R package (Papenberg and Klau, 2021). This approach reduces dependencies between the most distant nodes, as illustrated in Figure S8 in the Supplementary Material (Barile et al., 2026). For each subgraph we thus have $n = 266$ observations, corresponding to the restriction of the original network data to a subset of nodes. Conditionally on each sample of subgraphs, we run Algorithm 1 for 1,200 iterations, of which the first 200 are discarded as burn-in.

As in Section 5, we assess the accuracy of the estimated partition by comparing it to the one implied by the presence of 30 subjects in the study. The three panels on the left of Figure 8 show the values of three summary metrics for the considered values of N_{sub} . The consensus subgraph clustering approach struggles to correctly discriminate individual brain scans based on local characteristics when subgraphs with 5 nodes are used. However, when the analysis is conducted using subgraphs with at least 10 nodes, the method successfully captures the heterogeneity inherent in the brain networks of the 30

individuals. This experiment suggests that, as long as the subgraphs are not too small, examining local differences in connectivity patterns of subregions of the brain may be sufficient to detect overall similarities and differences across brain network data. Selecting an optimal N_{sub} involves balancing accuracy and computational time. The three right panels of Figure 8 display the ratios of the three clustering summary metrics already considered, relative to the computational time. To improve interpretability, we used $1 - \text{Entropy}$ instead of rescaling the Entropy directly. For each value of N_{sub} , the computational time is defined as the maximum time taken to analyze any of the m_{sub} datasets. Since these datasets can be analyzed in parallel, this definition of computational time represents the total time needed by a machine with unlimited cores. It is apparent that, for $N_{\text{sub}} \geq 15$, the extra computational cost is not rewarded in terms of accuracy. Therefore, it seems reasonable to select a value for N_{sub} by looking at where the time-rescaled clustering metrics are maximized, which for this dataset is $N_{\text{sub}} = 10$. For comparison, we conducted a similar study on the effectiveness of the consensus subgraph clustering approach using the Human Brain dataset based on 48 ROIs, dataset for which we can compare the approximate posterior distribution with the exact one, as studied in Section 5. The results, displayed in Figure S9 in the Supplementary Material (Barile et al., 2026), lead to selecting $N_{\text{sub}} = 15$.

Table 3 compares the results of the consensus subgraph clustering method applied to the human brain datasets with 48 and 200 ROIs, where N_{sub} was set equal to 15 and 10 respectively, and the summary metrics computed by comparing the estimated partition with that one implied by the presence of 30 subjects in the study. It can be appreciated that the summary metrics indicate a better performance of our method when analyzing the dataset with finer granularity. The results referring to the case $N = 48$ appear slightly worse than those obtained by applying our model on the entire graph observations, presented in Table 2. While it is clear that exploring only local properties of the graphs might reduce the ability to detect global properties of the graphs, these results indicate that the consensus subgraph clustering approach might be considered a valid alternative to cluster multiple network data when the number of nodes is large. Finally and for simplicity of illustration, we focus on the cluster allocation of the 30 observations referring to the three subjects assigned to the three largest clusters in the analysis run in Section 5. The central and right panels of Figure 7 display the cluster

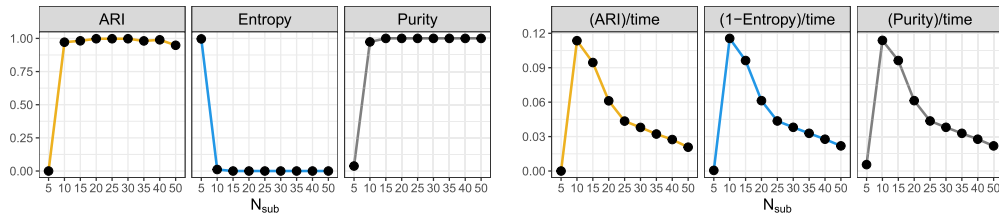


Figure 8: Clustering metrics comparing the partition estimated based on the consensus subgraph approach, with the one implied by the 30 individuals in the study, for the Human Brain dataset based on 200 ROIs, for N_{sub} ranging in $\{5, 10, \dots, 50\}$. See Section 6.1.

N (N_{sub})	\hat{K}	Adjusted Rand Index	Entropy	Purity
48 (15)	34	0.6710	0.1407	0.7932
200 (10)	30	0.9714	0.0115	0.9737

Table 3: Human brain dataset. Estimated number of clusters and clustering metrics with respect to the partition implied by the presence of 30 individuals in the study. See [Section 6.1](#).

allocation of these 30 observations, obtained by resorting to the consensus subgraph clustering approach to analyze the human brain datasets with $N = 48$ and $N = 200$, respectively, and highlight the topological properties of the identified clusters. Cluster allocation for the two cases resembles the results obtained in [Section 5](#) when analyzing the complete dataset with $N = 48$ ROIs, as shown in the left panel of [Figure 7](#). The only notable difference is that the consensus subgraph clustering applied to the dataset with 48 ROIs separates the scans of the subject labeled “C” into two distinct clusters.

7 Discussion

We introduced a novel Bayesian nonparametric approach to model heterogeneous populations of networks. The model’s location-scale structure favors interpretability while offering appealing theoretical properties, such as full support in the space of labeled graphs and posterior consistency. A key feature of our approach is that the proposed algorithm samples from distributions that are available in closed form. These distributions are derived using standard combinatorial arguments, which is made possible by the use of the Hamming distance to detect structural similarities among networks. As shown in [Section 4](#) and the brain network data analysis in [Section 5](#), our model offers notable flexibility and demonstrates overall performance improvements over existing methods in the literature. The implementation of the model results in a per-iteration computational cost that is quadratic in the number of nodes N . As a result, implementing our model can become computationally intensive when large values for N are considered. To address this challenge, we proposed a heuristic approach, named consensus subgraph clustering, designed to handle large graphs efficiently. Our analysis of human brain data with finer node granularity, as presented in [Section 6](#), demonstrates that this method performs well with larger networks and shows promise for scaling to the analysis of massive networks. Interestingly, while the method was presented using the spatial locations of the ROIs represented by network nodes, promising results were also obtained when the nodes were randomly partitioned, disregarding the available spatial information, as illustrated in the Supplementary Material (Barile et al., 2026). In addition, alternative strategies for node partitioning could be explored to implement the preliminary step of the consensus subgraph clustering approach. For instance, the concept of signal-subgraphs introduced by Vogelstein et al. (2012) provides a promising option.

The modeling strategy we presented offers multiple avenues for extension, opening up new research directions. An intriguing one involves exploring alternatives to the Hamming distance. Although the Hamming distance stands out for its tractability, it

falls short in capturing the broader structural changes within a graph (Donnat and Holmes, 2018). For instance, in brain networks, spectral distances can better assess global changes in connectivity. A natural extension of our work in this direction would be to employ the diffusion distance, which is based on the graph Laplacian and treats networks in a functional manner, focusing on changes that affect the global structure (Lunagómez et al., 2021). This approach, however, may not admit closed-form expressions, thus requiring alternative computational strategies. In this context, a particularly interesting research question is whether posterior consistency holds for kernels based on distances other than the Hamming distance, and how the choice of kernel affects the corresponding rates of convergence. Another interesting direction is to explore the possibility of introducing a stochastic block model structure through an appropriate specification of the base measure P_0 , thereby extending the scope of our model to allow for the simultaneous clustering of networks and nodes within each network. Finally, our model could serve as the building block for modeling related populations of networks in a partially exchangeable setting, in line with the approach of Durante et al. (2025), first proposed in Gaffi (2023). This could be achieved by inducing dependence at the level of population-specific mixing measures, such as through the Dependent Dirichlet process (MacEachern, 2000), thereby facilitating the sharing of information across models (see Quintana et al., 2022, for a review).

Acknowledgments

The authors are grateful to the Associate Editor and the two anonymous Referees for their valuable comments and suggestions. Bernardo Nipoti gratefully acknowledges the Department of Statistics at ITAM, Mexico City, for its hospitality during his visit, when this work was initiated.

Funding

Bernardo Nipoti acknowledges support of MUR – Prin 2022 – Grant no. 2022CLTYP4, funded by the European Union – Next Generation EU.

Supplementary Material

Supplementary Material for: Bayesian nonparametric modeling of heterogeneous populations of networks (DOI: [10.1214/26-BA1588SUPP](https://doi.org/10.1214/26-BA1588SUPP); .pdf). The Supplementary Material includes proof of theoretical results, additional details on posterior computation and prediction, and further information on simulation studies and real data analyses.

References

- Amini, A., Paez, M., and Lin, L. (2024). “Hierarchical Stochastic Block Model for Community Detection in Multiplex Networks.” *Bayesian Analysis*, 19(1): 319–345. MR4692550. doi: <https://doi.org/10.1214/22-ba1355>. 3
- Argiento, R., Filippi-Mazzola, E., and Paci, L. (2024). “Model-Based Clustering of Categorical Data Based on the Hamming Distance.” *Journal of the American Statistical*

- Association*, 1–23. MR4917287. doi: <https://doi.org/10.1080/01621459.2024.2402568>. 5
- Arroyo, J., Athreya, A., Cape, J., Chen, G., Priebe, C., and Vogelstein, J. (2021). “Inference for Multiple Heterogeneous Networks with a Common Invariant Subspace.” *Journal of Machine Learning Research*, 22: 1–49. MR4318498. 2, 16
- Barabási, A.-L. and Albert, R. (1999). “Emergence of Scaling in Random Networks.” *Science*, 286(5439): 509–512. MR2091634. doi: <https://doi.org/10.1126/science.286.5439.509>. 11
- Barile, F., Lunagómez, S., and Nipoti, B. (2026). “Supplementary Material for: Bayesian nonparametric modeling of heterogeneous populations of networks” *Bayesian Analysis*. doi: <https://doi.org/10.1214/26-BA1588SUPP>. 4, 8, 9, 11, 13, 14, 15, 16, 19, 20, 21
- Bassett, D. S. and Bullmore, E. (2006). “Small-World Brain Networks.” *The Neuroscientist*, 12(6): 512–523. 18
- Blackwell, D. and MacQueen, J. B. (1973). “Ferguson Distributions Via Polya Urn Schemes.” *The Annals of Statistics*, 1(2): 353–355. MR0362614. 9
- Bush, C. A. and MacEachern, S. N. (1996). “A Semiparametric Bayesian Model for Randomised Block Designs.” *Biometrika*, 83(2): 275–285. 9
- Craddock, R. C., James, G., Holtzheimer III, P. E., Hu, X. P., and Mayberg, H. S. (2012). “A whole brain fMRI atlas generated via spatially constrained spectral clustering.” *Human Brain Mapping*, 33(8): 1914–1928. 19
- Dahl, D. B., Johnson, D. J., and Müller, P. (2022). “Search Algorithms and Loss Functions for Bayesian Clustering.” *Journal of Computational and Graphical Statistics*, 31(4): 1189–1201. MR4513380. doi: <https://doi.org/10.1080/10618600.2022.2069779>. 13
- De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prunster, I., and Ruggiero, M. (2015). “Are Gibbs-Type Priors the Most Natural Generalization of the Dirichlet Process?” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 37(02): 212–229. 6
- Diquigiovanni, J. and Scarpa, B. (2019). “Analysis of association football playing styles: An innovative method to cluster networks.” *Statistical Modelling*, 19(1): 28–54. MR3903741. doi: <https://doi.org/10.1177/1471082X18808628>. 2
- Donnat, C. and Holmes, S. (2018). “Tracking network dynamics: A survey using graph distances.” *The Annals of Applied Statistics*, 12(2): 971–1012. MR3834292. doi: <https://doi.org/10.1214/18-AOAS1176>. 2, 5, 22
- Dunson, D. B. and Xing, C. (2009). “Nonparametric Bayes Modeling of Multivariate Categorical Data.” *Journal of the American Statistical Association*, 104(487): 1042–1051. MR2562004. doi: <https://doi.org/10.1198/jasa.2009.tm08439>. 5
- Durante, D., Dunson, D. B., and Vogelstein, J. T. (2017). “Nonparametric Bayes modeling of populations of networks.” *Journal of the American Statistical Association*.

- MR3750873. doi: <https://doi.org/10.1080/01621459.2016.1219260>. 2, 3, 11, 13, 15, 16, 17
- Durante, D., Gaffi, F., Lijoi, A., and Prünster, I. (2025). “Partially Exchangeable Stochastic Block Models for (Node-Colored) Multilayer Networks.” *Journal of the American Statistical Association*, 120(551): 1811–1827. MR4973899. doi: <https://doi.org/10.1080/01621459.2025.2507825>. 3, 22
- D’Angelo, S., Alfò, M., and Fop, M. (2023). “Model-based clustering for multidimensional social networks.” *Journal of the Royal Statistical Society Series A: Statistics in Society*, 186(3): 481–507. MR4754070. doi: <https://doi.org/10.1093/jrsssa/qnac011>. 2
- D’Angelo, S., Murphy, T. B., and Alfò, M. (2019). “Latent space modelling of multi-dimensional networks with application to the exchange of votes in Eurovision song contest.” *The Annals of Applied Statistics*, 13(2): 900–930. MR3963557. doi: <https://doi.org/10.1214/18-AOAS1221>. 2
- 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. MR0125031. 11
- Escobar, M. D. and West, M. (1995). “Bayesian Density Estimation and Inference Using Mixtures.” *Journal of the American Statistical Association*, 90(430): 577–588. MR1340510. 8
- Gaffi, F. (2023). “Modelling with Discrete Random Probability Measures.” Ph.D. thesis, Bocconi University. 22
- Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S., and Kolaczyk, E. D. (2017). “Hypothesis testing for network data in functional neuroimaging.” *The Annals of Applied Statistics*, 11(2): 725–750. MR3693544. doi: <https://doi.org/10.1214/16-AOAS1015>. 2
- Gollini, I. and Murphy, T. B. (2016). “Joint Modeling of Multiple Network Views.” *Journal of Computational and Graphical Statistics*, 25(1): 246–265. MR3474046. doi: <https://doi.org/10.1080/10618600.2014.978006>. 2
- Goodman, L. A. (1974). “Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models.” *Biometrika*, 61(2): 215–231. MR0370936. doi: <https://doi.org/10.1093/biomet/61.2.215>. 5
- Grötschel, M. and Wakabayashi, Y. (1989). “A Cutting Plane Algorithm for a Clustering Problem.” *Mathematical Programming*, 45(1): 59–96. MR1017212. doi: <https://doi.org/10.1007/BF01589097>. 19
- Guha, S. and Guhaniyogi, R. (2024). “Covariate-Dependent Clustering of Undirected Networks with Brain-Imaging Data.” *Technometrics*, 66(3): 422–437. MR4783228. doi: <https://doi.org/10.1080/00401706.2024.2321930>. 3
- Guha, S. and Rodriguez, A. (2021). “Bayesian Regression With Undirected Network Predictors With an Application to Brain Connectome Data.” *Journal of the American*

- Statistical Association*, 116(534): 581–593. MR4270005. doi: <https://doi.org/10.1080/01621459.2020.1772079>. 3
- Guha, S. and Rodriguez, A. (2023). “High-Dimensional Bayesian Network Classification with Network Global-Local Shrinkage Priors.” *Bayesian Analysis*, 18(4): 1131–1160. MR4674634. doi: <https://doi.org/10.1214/23-ba1378>. 3
- Guha, S., Rodriguez-Acosta, J., and Dinov, I. D. (2024). “A Bayesian Multiplex Graph Classifier of Functional Brain Connectivity Across Diverse Tasks of Cognitive Control.” *Neuroinformatics*, 22(4): 457–472. 3
- Hamming, R. W. (1950). “Error detecting and error correcting codes.” *The Bell System Technical Journal*, 29(2): 147–160. MR0035935. doi: <https://doi.org/10.1002/j.1538-7305.1950.tb00463.x>. 5
- Jing, B.-Y., Li, T., Lyu, Z., and Xia, D. (2021). “Community detection on mixture multilayer networks via regularized tensor decomposition.” *The Annals of Statistics*, 49(6). MR4352527. doi: <https://doi.org/10.1214/21-aos2079>. 2
- Jordan, M. I. (2004). “Graphical Models.” *Statistical Science*, 19(1): 140–155. MR2082153. doi: <https://doi.org/10.1214/088342304000000026>. 18
- Josephs, N., Amini, A. A., Paez, M., and Lin, L. (2025). “Nested stochastic block model for simultaneously clustering networks and nodes.” *arXiv:2307.09210*. 3, 13, 14, 15
- Josephs, N., Lin, L., Rosenberg, S., and Kolaczyk, E. D. (2023). “Bayesian Classification, Anomaly Detection, and Survival Analysis Using Network Inputs with Application to the Microbiome.” *The Annals of Applied Statistics*, 17(1): 199–224. MR4539028. doi: <https://doi.org/10.1214/22-aos1623>. 3, 8
- Kolaczyk, E. D., Lin, L., Rosenberg, S. J., Xu, J., and Walters, J. (2017). “Averages of unlabeled networks: Geometric characterization and asymptotic behavior.” *The Annals of Statistics*. MR4065172. doi: <https://doi.org/10.1214/19-AOS1820>. 2
- Le, C. M., Levin, K., and Levina, E. (2018). “Estimating a network from multiple noisy realizations.” *Electronic Journal of Statistics*, 12(2): 4697–4740. MR3894068. doi: <https://doi.org/10.1214/18-ejs1521>. 2
- Lunagómez, S., Olhede, S. C., and Wolfe, P. J. (2021). “Modeling Network Populations via Graph Distances.” *Journal of the American Statistical Association*, 116(536): 2023–2040. MR4353730. doi: <https://doi.org/10.1080/01621459.2020.1763803>. 2, 5, 7, 12, 16, 22
- MacEachern, S. N. (2000). “Dependent dirichlet processes.” *Unpublished manuscript, Department of Statistics, The Ohio State University*, 5. 22
- 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. MR4698608. doi: <https://doi.org/10.1214/23-aos1789>. 2, 7, 13, 15, 16, 17, 18
- Mori, S., Wakana, S., van Zijl, P., and Nagae-Poetscher, L. (2005). *MRI Atlas of Human White Matter*. Elsevier Science. 16

- Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., and Onnela, J.-P. (2010). “Community Structure in Time-Dependent, Multiscale, and Multiplex Networks.” *Science*, 328(5980): 876–878. MR2662590. doi: <https://doi.org/10.1126/science.1184819>. 1
- Mukherjee, S. S., Sarkar, P., and Lin, L. (2017). “On clustering network-valued data.” *Advances in Neural Information Processing Systems*, 30. 2
- Nelson, B., Bassett, D., Camchong, J., Bullmore, E., and Lim, K. (2017). “Comparison of large-scale human brain functional and anatomical networks in schizophrenia.” *NeuroImage: Clinical*, 15. 1
- Nielsen, A. M. and Witten, D. (2018). “The multiple random dot product graph model.” *arXiv:1811.12172*. 2
- Nowicki, K. and Snijders, T. A. B. (2001). “Estimation and Prediction for Stochastic Blockstructures.” *Journal of the American Statistical Association*, 96(455): 1077–1087. MR1947255. doi: <https://doi.org/10.1198/016214501753208735>. 11
- Papenberg, M. and Klau, G. W. (2021). “Using Anticlustering to Partition Data Sets into Equivalent Parts.” *Psychological Methods*, 26(2): 161–174. 19
- Peixoto, T. P. (2023). “The Netzschleuder network catalogue and repository.” Software. 16
- Quintana, F. A., Müller, P., Jara, A., and MacEachern, S. N. (2022). “The dependent Dirichlet process and related models.” *Statistical Science*, 37(1): 24–41. MR4371095. doi: <https://doi.org/10.1214/20-sts819>. 22
- Regazzini, E., Lijoi, A., and Prünster, I. (2003). “Distributional results for means of normalized random measures with independent increments.” *The Annals of Statistics*, 31(2): 560–585. MR1983542. doi: <https://doi.org/10.1214/aos/1051027881>. 6
- Ren, S., Wang, X., Liu, P., and Zhang, J. (2023). “Bayesian nonparametric mixtures of Exponential Random Graph Models for ensembles of networks.” *Social Networks*, 74: 156–165. 3
- Reyes, P. and Rodriguez, A. (2016). “Stochastic blockmodels for exchangeable collections of networks.” *arXiv:1606.05277*. 3
- Sethuraman, J. (1994). “A constructive definition of Dirichlet priors.” *Statistica Sinica*, 4(2): 639–650. MR1309433. 7
- Shaw, P., Mikusz, M., Nurmi, P., and Davies, N. (2018). “Tacita: A privacy preserving public display personalisation service.” In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 448–451. 1
- Signorelli, M. and Wit, E. C. (2020). “Model-based clustering for populations of networks.” *Statistical Modelling*, 20(1): 9–29. MR4052400. doi: <https://doi.org/10.1177/1471082X19871128>. 2, 13, 15, 16
- Stanley, N., Shai, S., Taylor, D., and Mucha, P. J. (2016). “Clustering network layers with the strata multilayer stochastic block model.” *IEEE Transactions on Network*

- Science and Engineering*, 3(2): 95–105. MR3515211. doi: <https://doi.org/10.1109/TNSE.2016.2537545>. 2
- Strehl, A. and Ghosh, J. (2002). “Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions.” *Journal of Machine Learning Research*, 3: 583–617. MR1991087. doi: <https://doi.org/10.1162/153244303321897735>. 19
- Tang, M., Sussman, D. L., and Priebe, C. E. (2013). “Universally consistent vertex classification for latent positions graphs.” *The Annals of Statistics*, 41(3): 1406–1430. MR3113816. doi: <https://doi.org/10.1214/13-AOS1112>. 8
- Vogelstein, J. T., Roncal, W. G., Vogelstein, R. J., and Priebe, C. E. (2012). “Graph classification using signal-subgraphs: Applications in statistical connectomics.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7): 1539–1551. 21
- Walker, S., Damien, P., and Lenk, P. (2004). “On priors with a Kullback–Leibler property.” *Journal of the American Statistical Association*, 99(466): 404–408. MR2062826. doi: <https://doi.org/10.1198/016214504000000386>. 8
- Wang, S., Arroyo, J., Vogelstein, J. T., and Priebe, C. E. (2021). “Joint Embedding of Graphs.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(4): 1324–1336. 2
- Watts, D. J. and Strogatz, S. H. (1998). “Collective dynamics of ‘small-world’ networks.” *Nature*, 393(6684): 440–442. MR1716136. 11
- Yin, F., Shen, W., and Butts, C. T. (2019). “Finite Mixtures of ERGMs for Modeling Ensembles of Networks.” *Bayesian Analysis*. MR4506025. doi: <https://doi.org/10.1214/21-ba1298>. 3
- Young, J.-G., Kirkley, A., and Newman, M. E. J. (2022). “Clustering of heterogeneous populations of networks.” *Physical Review E*, 105: 014312. MR4381990. doi: <https://doi.org/10.1103/physreve.105.014312>. 2, 3, 7
- Zuo, X.-N., Anderson, J., Bellec, P., Birn, R., Biswal, B., Blautzik, J., et al. (2014). “An open science resource for establishing reliability and reproducibility in functional connectomics.” *Scientific Data*, 1: 1–13. 4, 16