Graph-Aligned Random Partition Model (GARP)

Graph-Aligned Random Partition Model (GARP)

Giovanni Rebaudo[a] (giovanni.rebaudo@unito.it)
Peter Müller[b] (pmueller@math.utexas.edu)

[a]University of Torino, IT

[b]University of Texas at Austin, USA

Abstract

Bayesian nonparametric mixtures and random partition models are powerful tools for probabilistic clustering. However, standard independent mixture models can be restrictive in some applications such as inference on cell lineage due to the biological relations of the clusters. The increasing availability of large genomic data requires new statistical tools to perform model-based clustering and infer the relationship between homogeneous subgroups of units. Motivated by single-cell RNA data we develop a novel dependent mixture model to jointly perform cluster analysis and align the clusters on a graph. Our flexible graph-aligned random partition model (GARP) exploits Gibbs-type priors as building blocks, allowing us to derive analytical results for the probability mass function (pmf) on the graph-aligned random partition. We derive a generalization of the Chinese restaurant process from the pmf and a related efficient and neat MCMC algorithm to implement Bayesian inference. We illustrate posterior inference under the GARP using single-cell RNA-seq data from mice stem cells. We further investigate the performance of the model in recovering the underlying clustering structure as well as the underlying graph by means of simulation studies.

Keywords: Bayesian Nonparametrics, Random Partition Model, Gibbs-Type Prior, Dependent Mixture Model, Exchangeability, Single-Cell RNA

1 Introduction

We introduce a graph-aligned random partition model with one set of clusters being identified as vertices of a graph and other clusters being interpreted as edges between those. The model construction is motivated by the increasing availability of genomic data that requires new statistical tools to perform inference and uncertainty quantification on homogeneous subgroups of units (e.g., single-cells) and hypothesized relationships between the subgroups (e.g., transitions between the subgroups). In the present article, we deal with single-cell RNA sequencing experiments (scRNA-seq) that provide an unprecedented opportunity to study cellular heterogeneity and the evolution of complex tissues. The interest is to identify the main homogeneous cell subpopulations (i.e., clusters) in terms of gene expressions and jointly infer transitions of cells between these.

Dirichlet process (DP) mixtures (Lo, 1984) are well-established Bayesian nonparametric (BNP) models to infer homogeneous subgroups of observations via probabilistic clustering. However, the law of the random partition induced by the DP, related to the so-called Chinese restaurant process (CRP), is controlled by a single parameter. This leaves DP mixture models too restrictive for many applications and several alternative models were introduced in the literature to allow more flexible clustering. This includes the symmetric finite Dirichlet prior (Green and Richardson, 2001), the Pitman-Yor process (PYP) (Pitman and Yor, 1997), the normalized inverse Gaussian (NIG) (Lijoi *et al.*, 2005), the normalized generalized gamma process (NGGP) (Lijoi *et al.*, 2007b), mixture of finite mixtures (MFM) (Nobile, 1994; Richardson and Green, 1997; Nobile and Fearnside, 2007; Miller and Harrison, 2018) and the mixture of DP (MDP) models (Antoniak, 1974). All these belong to the wider family of Gibbs-type priors (Gnedin and Pitman, 2006) that can be seen as a natural, flexible generalization of the DP (De Blasi *et al.*, 2015).

However, Gibbs-type processes entail independent cluster-specific parameters not allowing us to infer the relationship between clusters as needed in our motivating example. Recently, repulsive priors that allow for dependent cluster-specific parameters were successfully introduced to favor more parsimonious and well-separated clusters (Petralia *et al.*, 2012; Xu *et al.*, 2016; Beraha *et al.*, 2022). Repulsive mixtures introduce (negative) dependence between cluster-specific values to better separate clusters. However, these models still stop short of inferring a biological relationship between the clusters, such as aligning the clusters on a graph, as desired in our framework.

In this article, we propose a graph-aligned random partition model (GARP) that exploits the flexible, but tractable, building blocks of Gibbs-type priors to build a random partition aligned on a graph. The desired interpretation of clusters as vertices and edges in a graph naturally gives rise to dependent priors on cluster-specific parameters. In the motivating example with single-cell RNA-seq data, vertex-clusters represent homogeneous cell subpopulations and edge-clusters correspond to cells that are transitioning between those. See Figure 1 for a scatter plot of single-cell RNA data in a two-dimensional space that captures most of the recorded genetic expressions of mice stem cell data.

The remainder of the article is as follows. In Section 2 we introduce a model for graph-aligned probabilistic clustering. In Section 3 we introduce special examples. In Sections 4, 5 and 6 we study a useful approximation, implied homogeneity assumptions, and identifiability of vertices versus edges. Section 7 applies the model to single-cell RNA-seq data of mice stem cells and Section 8 concludes with final comments. Substantive additional details, including proofs, validations on simulated data, a characterization in terms of discrete probabilities, a discussion of hyperparameter choices, and details on the strategy to obtain point estimates from posterior samples are available as an online supplement. The code is available at `https://github.com/GiovanniRebaudo/GARP`.

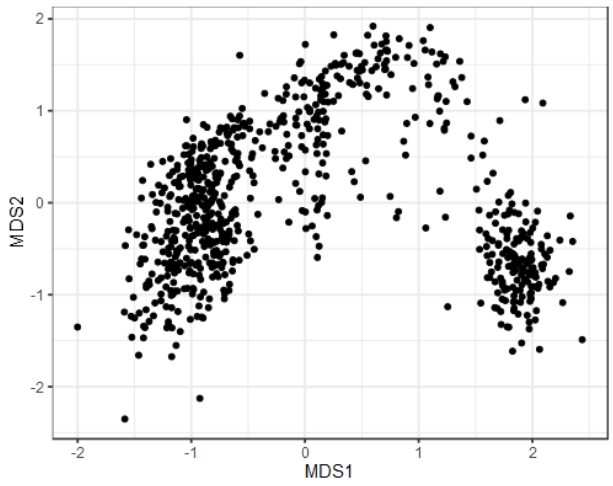


Figure 1: Two-dimensional representation of genetic expressions of the RNA mice single-cells data.

2 Graph-Aligned Random Partition Model

We introduce a graph-aligned random partition model (GARP) for $\boldsymbol{y} = \{\boldsymbol{y}_i : i = 1, \ldots, N\}$, $\boldsymbol{y}_i \in \mathbb{R}^d$. The two main features of the model are a two-level random partition structure that assigns observations into vertex-clusters and edge-clusters, and a mixture of normal sampling models with cluster-specific parameters that reflect this split into vertex and edge-clusters. That is, the mixture of normal models is set up such that observations in vertex-clusters form homogeneous subsets in Euclidean space, and observations in edge-clusters are located between the adjacent vertices. We characterize the model in three different representations that are minor variations of representations that are traditionally used for infinitely exchangeable random partition models (Pitman, 1996), including (1) the probability mass function (pmf) of the graph-aligned random partition via the introduction of exchangeable partition probability functions (EPPF); (2) a composition of Pólya urn schemes, i.e., predictive probability functions, using a generalized CRP (gCRP); and (3) the configuration of ties that is implied by sampling from a composition of discrete random probability measures, similar to the construction of species sampling processes (SSP). See Pitman (1996) and Lee *et al.* (2013a) for details on these three characterizations for infinitely exchangeable random partitions (without alignment on a graph).

2.1 A Gaussian Mixture over Vertices and Edges

We start the model construction with a sampling model given the latent graph-aligned partition. We need some notation. Let V_i be an indicator for observation i being placed into a vertex-cluster and let Z_i denote a cluster membership indicator. We write $\boldsymbol{V} = (V_1, \ldots, V_N)$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_N)$ (throughout $\boldsymbol{x}$ denotes the collection of all previously

defined elements x_a). We denote with $N_{v,N} = \sum_{i=1}^{N} V_i$ the number of observations in vertex-clusters, and with $N_{e,N} = N - N_{v,N}$ the implied number in edge-clusters. For notational simplicity, we drop the subscript $_N$ when implied by the context. If i belongs to a vertex (i.e., $V_i = 1$), then $Z_i \in [K_v] \equiv \{1, \ldots, K_v\}$, where K_v is the random number of vertex-clusters. If i belongs to an edge (i.e., $V_i = 0$), then $Z_i = (k, k')$, with $k < k'$ indicating the adjacent vertex-clusters. Let K_e denote the number of edge-clusters. Clearly, an edge must connect two vertices, implying $K_e \leq \frac{K_v(K_v-1)}{2} \equiv M_e$. Finally, let $\boldsymbol{Z_v} = (Z_i : V_i = 1)$ and $\boldsymbol{Z_e} = (Z_i : V_i = 0)$ denote the set of cluster membership indicators for vertices and edges, respectively.

Given a graph-aligned random partition, we assume normal sampling

$$\boldsymbol{y}_i \mid Z_i, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^* \overset{\text{ind}}{\sim} \mathrm{N}(\boldsymbol{y}_i \mid \boldsymbol{\mu}^*_{Z_i}, \boldsymbol{\Sigma}^*_{Z_i}), \quad (i = 1, \ldots, N), \tag{1}$$

keeping in mind that $Z_i = k$ for $V_i = 1$ and $Z_i = (k, k')$ for $V_i = 0$. The cluster-specific parameters are defined as follows. For the vertex-parameters $\boldsymbol{\theta}^*_k = (\boldsymbol{\mu}^*_k, \boldsymbol{\Sigma}^*_k)$ we assume (conditionally) conjugate normal-inverse Wishart priors

$$\boldsymbol{\theta}^*_k \mid K_v \overset{\text{iid}}{\sim} \mathrm{NIW}(\boldsymbol{\mu}_0, \lambda_0, \kappa_0, \boldsymbol{\Sigma}_0), \quad (k = 1, \ldots, K_v). \tag{2}$$

For edge-clusters, cluster-specific parameters $\boldsymbol{\theta}^*_{k,k'} = (\boldsymbol{\mu}^*_{k,k'}, \boldsymbol{\Sigma}^*_{k,k'})$ are defined as functions of the adjacent vertex-clusters,

$$\boldsymbol{\mu}^*_{k,k'} = \frac{\boldsymbol{\mu}^*_k + \boldsymbol{\mu}^*_{k'}}{2}, \quad \boldsymbol{\Sigma}^*_{k,k'} = f(\boldsymbol{\mu}^*_k, \boldsymbol{\mu}^*_{k'}, r_0, r_1). \tag{3}$$

Here f is such that the $\alpha\%$-level contour of the $N(\boldsymbol{\mu}^*_{k,k'}, \boldsymbol{\Sigma}^*_{k,k'})$ density is stretched around the line $L_{k,k'}$ connecting $\boldsymbol{\mu}^*_k$ and $\boldsymbol{\mu}^*_{k'}$, the Gaussian component projected onto $L_{k,k'}$ has standard deviation $r_0\,||\boldsymbol{\mu}^*_k - \boldsymbol{\mu}^*_{k'}||$, and the projection onto the orthogonal complement $L^{\perp}_{k,k'}$ are $d-1$ independent Gaussian distributions with variances r_1^2. Figure S.1 in Section S.1 of the supplemental materials shows the contour plot of an edge-cluster in $\mathbb{R}^2$. See the same section and Section S.3 of the supplementary materials for more discussion of $\boldsymbol{\Sigma}^*_{k,k'}$, and comments on the choice of hyperparameters r_0, r_1.

2.2 Graph-Aligned Random Partition (GARP)

We introduce a flexible graph-aligned random partition model. In words, we first label each item as belonging to a vertex or edge cluster (with probability p_v and $(1 - p_v)$, respectively), then use a Gibbs-type prior to cluster items associated with vertices, and a Dirichlet-multinomial prior to place those associated with edges into one of the M_e possible edges, respectively. Let $(n_1, \ldots, n_{K_v})$ denote the cardinalities of the vertex-clusters, i.e., $n_k = \sum_i \mathbb{1}(\{V_i = 1\} \cap \{Z_i = k\})$, and similarly let $n_{k,k'} = \sum_i \mathbb{1}(\{V_i = 0\} \cap \{Z_i = (k, k')\})$

denote the sizes of the implied edge-clusters, with $n_{k,k'} = 0$ indicating the lack of an edge between k, k'. We define a graph-aligned random partition model via the pmf of $\boldsymbol{V}, \boldsymbol{Z}$

$$G^{(N)}(\boldsymbol{V}, \boldsymbol{Z}) \propto p_v^{N_v} \, \mathrm{EPPF}_{K_v}^{(N_v)}(n_1, \ldots, n_{K_v} \mid \alpha, \sigma)/K_v! \\ (1-p_v)^{N_e} \mathrm{DM}_{M_e}^{(N_e)}((n_{k,k'})_{k<k'} \mid \beta/M_e) \, \mathbb{1}(\underbrace{\{N_e = 0\} \cup \{M_e > 0\}}_{E_N}), \tag{4}$$

where $\mathrm{EPPF}(\cdot \mid \alpha, \sigma)$ denotes the EPPF of a Gibbs-type prior, DM is the marginal likelihood of an M_e-symmetric Dirichlet-multinomial model (for categorical realizations, and defining $\mathrm{DM}_{\cdot}^{(0)}(\cdot) = \mathrm{DM}_0^{(\cdot)}(\cdot) \equiv 1$) and $\mathbb{1}(\{N_e = 0\} \cup \{M_e > 0\})$ is an indicator that represents the constraint that edges can only be assigned if there are at least 2 vertices ($M_e > 0$, that is, $K_v > 1$), or no units are assigned to edges ($N_e = 0$). We will use E_N to refer to this truncation event. In particular, when $K_v = 1$ (and therefore $M_e = 0$) (4) reduces to $G^{(N)}(\boldsymbol{V}, \boldsymbol{Z}) \propto p_v^N \, \mathrm{EPPF}_1^{(N)}(N \mid \alpha, \sigma)$ with $V_i = Z_i = 1$, for all i, and $G^{(N)}(\boldsymbol{V}, \boldsymbol{Z}) = 0$ for any other configuration $(\boldsymbol{V}, \boldsymbol{Z})$, e.g., any configuration with $N_e > 0$ (i.e., E_N^c).

An EPPF characterizes the distribution of an exchangeable partition (Pitman, 1996), with $\mathrm{EPPF}_{K_v}^{(N_v)}(n_1, \ldots, n_{K_v})$ being the probability of observing a particular (unordered) partition of N_v observations into K_v subsets of cardinalities $\{n_1, \ldots, n_{K_v}\}$. Since an EPPF refers to unordered partitions we include the additional denominator $K_v!$ for the ordered $\boldsymbol{Z}$. See Section 4 for more discussions of the homogeneity assumptions implied by our model. We specify the EPPF as a Gibbs-type prior,

$$\mathrm{EPPF}_{K_v}^{(N_v)}(n_1, \ldots, n_{K_v} \mid \alpha, \sigma) = W_{N_v, K_v} \prod_{k=1}^{K_v} (1-\sigma)_{n_k - 1}, \tag{5}$$

where $(x)_n = x(x+1)\ldots(x+n-1)$ represents the ascending factorial, $\sigma < 1$ is a discount parameter and the set of non-negative weights $\{W_{n,k} : 1 \leq k \leq n\}$ satisfies the recursive equation $W_{n,k} = (n - \sigma k) W_{n+1,k} + W_{n+1,k+1}$. The parameter α in the conditioning set is used to define $W_{n,k}$ for some of the upcoming examples. In a second step, the observations assigned to edges are (ordered) clustered using a DM distribution.

$$G^{(N)}((Z_i : \ V_i = 0) \mid \boldsymbol{V}, K_v) = \mathrm{DM}_{M_e}^{(N_e)}((n_{k,k'})_{k<k'_v} \mid \beta/M_e) \\ = \frac{\Gamma(\beta)}{\Gamma(N_e + \beta/M_e)} \prod_{(k,k'): k<k' \leq K_v} \frac{\Gamma(n_{k,k'} + \beta/M_e)}{\Gamma(\beta/M_e)}. \tag{6}$$

Model (4) is a hierarchical constrained composition of a Gibbs-type prior and a symmetric-DM with hyperparameter β/M_e. As we shall show, the model preserves most of the analytical and computational tractability of the simpler building blocks.

2.3 Generalized Chinese Restaurant Process

In an alternative characterization of (4), the model can be defined as a truncated version of a composition of gCRP. We denote the latter, that is, the model before the truncation, as $\widetilde{G^{(N)}}$ and refer to it as the *relaxed model*.

$$G^{(N)}(\boldsymbol{V}, \boldsymbol{Z}) \propto \widetilde{G^{(N)}}(\boldsymbol{V}, \boldsymbol{Z}) \mathbb{1}(E_N). \tag{7}$$

Recall that $E_N = \{N_e = 0\} \cup \{M_e > 0\}$ is the truncation. In Section 4 we show that $\widetilde{G^{(N)}}$ assigns high probability to E_N, going to 1 with $n \to \infty$ for most Gibbs-type priors.

The relaxed model $\widetilde{G^{(N)}}(\boldsymbol{V}, \boldsymbol{Z})$ is a hierarchical composition of tractable generalized Pólya urn schemes, starting with the assignments to vertices or edges

$$V_i \overset{\text{iid}}{\sim} \text{Bern}(p_v), \quad (i = 1, \ldots, N). \tag{8}$$

Next, we sample cluster membership indicators $\boldsymbol{Z_v} = (Z_i : \ V_i = 1)$ for the vertex-clusters from the gCRP associated with Gibbs-type prior, i.e., $\boldsymbol{Z_v} \mid \boldsymbol{V} \sim \text{gCRP}(\alpha, \sigma)$, with the gCRP implied by $\widetilde{G^{(N)}}$ given as

$$\widetilde{G^{(N)}}\{Z_i = k \mid \boldsymbol{Z}^{-i}, \boldsymbol{V}^{-i}, V_i = 1\} = \begin{cases} \frac{W_{N_v, K_v^{-i}}}{W_{N_v - 1, K_v^{-i}}}(n_k^{-i} - \sigma) & k \in [K_v^{-i}] \\ \frac{W_{N_v, K_v^{-i}+1}}{W_{N_v - 1, K_v^{-i}}} & k = K_v^{-i} + 1. \end{cases} \tag{9}$$

Throughout $\mathbf{x}^{-i}$ identifies a quantity after removing the element i from $\mathbf{x}$. Moreover, we use the following notation in the manuscript: given a probability measure P we denote by $P\{E\}$ the probability measure evaluated in a set E and by $P(a)$ the corresponding probability density function (pdf) or pmf evaluated in a point a. See Section 3 for examples of different gCRP and implied prior assumptions on the number of vertices.

Finally, the cluster membership indicators $\boldsymbol{Z_e}$ for the observations in edges follow the Pólya urn scheme induced by a DM distribution

$$\widetilde{G^{(N)}}\{Z_i = (k, k') \mid V_i = 0, \boldsymbol{Z}^{-i}, E_N\} \propto n_{k,k'}^{-i} + \beta / M_e, \tag{10}$$

with $k' < k \leq K_v$. Here, β / M_e favors sparsity as the dimension of the graph increases. Note that (8) might generate $N_e > 0$, even when (9) implies $M_e = 0$. For this case we define for completeness $\widetilde{G^{(N)}}\{Z_i = (1, 2) \mid V_i = 0, \boldsymbol{Z}^{-i}, E_N^c\} \equiv 1$ (without implications for $G^{(N)}$, due to the inclusion of the truncation to E_N in (7)).

The aforementioned composition of urn schemes characterizes the GARP (4):

Proposition 1. *The random partition structure of the* GARP *model* (4) *can be characterized as the truncated composition of* gCRP *defined in* (7), (8), (9) *and* (10).

We rely on this representation to derive an MCMC algorithm that generalizes the marginal MCMC algorithms for DP mixture models and Gibbs-type priors (Neal, 2000; De Blasi *et al.*, 2015; Miller and Harrison, 2018). Moreover, as we shall see, the probability of the truncation event E_N is high and rapidly goes to 1 in most cases.

Composition of Discrete Random Probabilities. Finally, in Section S.2 of the supplementary materials we derive a third characterization of the proposed GARP. We define $\widetilde{G^{(N)}}$ as a graph-aligned random partition (with unique atoms) implied by the ties under conditionally i.i.d. sampling of $\boldsymbol{\theta}_i$. Such a characterization will be used in a lemma to prove Theorem 3 and can be used to connect with existing BNP literature to derive a conditional Gibbs sampler.

3 Specific Model Choices

Conditioning on the vertex assignments $\boldsymbol{V}$, under the relaxed model $\widetilde{G^{(N)}}$ the distribution of the clustering indicators $\boldsymbol{Z}_v$ is given by the EPPF of a Gibbs-type prior (Gnedin and Pitman, 2006; De Blasi *et al.*, 2015). We introduce four specific choices, stating the $\text{EPPF}_{K_v}^{(N_v)}(n_1, \ldots, n_{K_v})$ for partitioning N_v observations into K_v vertices. Table 1 shows the corresponding expressions for $\widetilde{G^{(N)}}\{Z_i = k \mid V_i = 1, \boldsymbol{Z}_v^{-i}, \boldsymbol{V}^{-i}\}$ in the gCRP of (9), and the weights and atoms for $P_v = \sum_{m=1}^{M_v} \pi_m \delta_{\tilde{\boldsymbol{\theta}}_m}$ in (S.1) of the supplementary materials. Throughout, the prior for cluster-specific parameters remains the NIW in (2).

Table 1: $\widetilde{G^{(N)}}\{Z_i = k \mid \ldots\}$ in the gCRP (9), and weights $(\pi_m)_{m=1}^{M_v}$ for $P_v = \sum_{m=1}^{M_v} \pi_m \delta_{\tilde{\boldsymbol{\theta}}_m}$ in (S.1). See the text for the definition of examples 1 through 4.

	$\widetilde{G^{(N)}}\{Z_i = k \mid V_i = 1, \boldsymbol{Z}_v^{-i}, \boldsymbol{V}^{-i}\} \propto$		$P(\pi_1, \pi_2, \ldots \mid M_v)$	$p(M_v = m)$
Ex.	$k \in \boldsymbol{Z}_v^{-i}$	$k = K_v^{-i} + 1$		
1	$n_k^{-i} + \rho$	$\rho(M_v - K_v^{-i})$ [a]	$\text{Dir}(\rho, \ldots, \rho)$	fixed $M_v \in \mathbb{N}$
2	$(n_k^{-i} + 1) \times (N_v^{-i} - K_v^{-i} + \gamma)$	$(K_v^{-i})^2 - K_v^{-i}\gamma$	$\text{Dir}(1, \ldots, 1)$	$\frac{\gamma(1-\gamma)_{m-1}}{m!}$
3	n_k^{-i}	α	$\text{GEM}(\alpha)$ [b]	$M_v = \infty$
4	$n_k^{-i} - \sigma$	$\alpha + K_v^{-i}\sigma$	$\text{GEM}(\alpha, \sigma)$ [b]	$M_v = \infty$

[a] subject to $K_v^{-i} < M_v$.
[b] GEM stands for the distribution of probability weights after Griffiths, Engen, and McCloskey (Ewens, 1990), using the 1-parameter version defined there and the related 2-parameters extension.

Example 1 (M_v-dimensional symmetric Dirichlet)**.** *If prior information on an upper bound M_v on the number of vertices is available we can proceed with a finite-dimensional symmetric Dirichlet prior (Green and Richardson, 2001).*

$$\text{EPPF}_{K_v}^{(N_v)}(n_1,\dots,n_{K_v}) = \frac{M_v!}{(M_v-K_v)!}\frac{\Gamma(\rho\, M_v)}{\Gamma(N_v+\rho\, M_v)\Gamma(\rho)^{K_v}}\prod_{k=1}^{K_v}\Gamma(n_k+\rho). \tag{11}$$

Allowing for unknown M_v the model becomes a mixture of symmetric Dirichlet model, that is, a mixture of finite mixtures (MFM). MFMs can be particularly interesting for allowing consistent estimation of any finite number of clusters (Nobile, 1994; Miller and Harrison, 2018). MFMs are a special case of Gibbs-type priors. A relevant example is the *Gnedin process*.

Example 2 (Gnedin process, with $\sigma = -1$)**.** *Under the Gnedin prior with parameter $\gamma \in (0,1)$ the* $\text{EPPF}_{K_v}^{(N_v)}$ *in* (4) *becomes*

$$\text{EPPF}_{K_v}^{(N_v)}(n_1,\dots,n_{K_v}) = \sum_{m=1}^{\infty}\text{EPPF}_{K_v}^{(N_v)}(n_1,\dots,n_{K_v} \mid M_v = m)\, p(M_v = m),$$

where $\text{EPPF}_{K_v}^{(N_v)}(n_1,\dots,n_{K_v} \mid M_v = m)$ *is the* EPPF *of the M_v-symmetric Dirichlet prior in* (11)*, with $\rho = 1$ and $p(M_v = m) = \frac{\gamma(1-\gamma)_{m-1}}{m!}$.*

The gCRP for the Gnedin process allows tractable analytical results and efficient algorithms. Moreover, the Gnedin process entails a distribution on the number of components M_v that has the mode at 1, a heavy tail, and infinite expectation (Gnedin, 2010). Therefore, the implied MFM favors a small number of vertices, while also being robust due to the heavy tail distribution of M_v.

Note that one can use $M_v = \infty$ to let the number of vertices (i.e., K_v) grow to infinity with N_v. Examples are the DP which entails a logarithmic growth of the number of vertices and the PYP which entails a polynomial growth of the number of vertices.

Example 3 (DP)**.** *Under the DP prior with parameter $\alpha > 0$ the* $\text{EPPF}_{K_v}^{(N_v)}$ *in* (4) *becomes*

$$\text{EPPF}_{K_v}^{(N_v)}(n_1,\dots,n_{K_v}) = \frac{\alpha^{K_v}\Gamma(\alpha)}{\Gamma(\alpha+N_v)}\prod_{k=1}^{K_v}(n_k-1)!$$

Example 4 (PYP)**.** *Under a PYP prior with parameters $\sigma \in [0,1)$ and $\alpha > 0$ the* $\text{EPPF}_{K_v}^{(N_v)}$ *in* (4) *becomes*

$$\text{EPPF}_{K_v}^{(N_v)}(n_1,\dots,n_{K_v}) = \frac{\Gamma(\alpha+1)\prod_{k=1}^{K_v-1}(\alpha+k\sigma)}{\Gamma(\alpha+N_v)}\prod_{k=1}^{K_v}(1-\sigma)_{n_k-1}.$$

With $\sigma = 0$ the PYP reduces to the DP. Other popular sub-classes of Gibbs-type priors include the NGPP (Lijoi *et al.*, 2007b), the NIG (Lijoi *et al.*, 2005, 2007a), and the MFM (Nobile and Fearnside, 2007; Miller and Harrison, 2018). See De Blasi *et al.* (2015) for a comprehensive review of Gibbs-type priors.

Finally, we note that here we focus on prior elicitation of the Gibbs-type random partition that controls the vertex-clusters and the number of vertices (i.e., $K_v \leq \min(M_v, N_v) \leq \min(M_v, N)$). Given K_v the possible number of edges is finite. The only Gibbs-type prior with a finite fixed number of components M_e is the symmetric Dirichlet (see e.g, De Blasi *et al.*, 2015), that is the DM_{M_e} in (4). Although the preceding discussion focuses on the Gibbs-type partition that controls the vertices assignment, it entails (thanks to the hierarchical definition e.g., in Section 2.3) similar flexibility in the joint prior elicitation of the vertices assignments.

4 Goodness of the Approximation

We discuss properties of the approximation of the GARP model in (4) by the relaxed model $\widetilde{G^{(N)}}$, and why it is a good approximation of $G^{(N)}$, justifying the prior elicitation of $G^{(N)}$ via $\widetilde{G^{(N)}}$. Importantly, the results allow us to effectively sample from the GARP via rejection sampling, using proposals from $\widetilde{G^{(N)}}$.

Proposition 2. *The probability of the truncation event E_N under the relaxed model is*

$$\widetilde{G^{(N)}}\{E_N\} = p_v^N + \sum_{n_v=2}^{N-1} \binom{N}{n_v} p_v^{n_v} (1-p_v)^{(N-n_v)} \big[1 - (1-\sigma)_{n_v-1} W_{n_v,1}\big]. \tag{12}$$

Here $p_v^N = \widetilde{G^{(N)}}\{N_v = N\}$, and $(1-\sigma)_{n_v-1} W_{n_v,1}$ in the second term arises from (5) as the probability given $\{N_v = n_v\}$ of having a single vertex, i.e., $\widetilde{G^{(N)}}\{K_v = 1 \mid N_v = n_v\} = \mathrm{EPPF}_1^{(n_v)}(n_v)$. For the Gibbs-type priors in the following examples, the latter reduces to simple analytical expressions.

In the upcoming discussion, we introduce several closely related distributions. To avoid confusion we provide a summary and list of defined distributions in Table S.1 in the supplementary materials. Let $\widetilde{G_{\mathrm{VZ}}^{(N)}}$ denote the law of V_i, $i = 1, \ldots, N$ and $\boldsymbol{Z_v} = (Z_i : \ i \in [N], V_i = 1)$ under the relaxed model. More precisely, $\widetilde{G_{\mathrm{VZ}}^{(N)}}$ is the joint law of the random variables $(T_1, \ldots, T_N)$, where $T_i = V_i$ if $V_i = 0$ and $T_i = (V_i, Z_i)$ if $V_i = 1$. Let $\widetilde{G_{\mathrm{VZ}}}$ denote the law of the stochastic process with Kolmogorov consistent finite dimensional $(\widetilde{G_{\mathrm{VZ}}^{(N)}})_{N\in\mathbb{N}}$. Such a process exists due to the i.i.d. nature of V_i and the exchangeable nature of the Gibbs-type prior that defines $\boldsymbol{Z}_v$ given $\boldsymbol{V}$. We therefore have by the strong law of large numbers $\lim_{N\to\infty} N_v/N = p_v$, $\widetilde{G_{\mathrm{VZ}}}$-a.s. Also, note that the truncation event E_N is a

function of $(\boldsymbol{V}, \boldsymbol{Z}_v)$ (thus $\mathbf{T}$) only, allowing us to evaluate $\widetilde{G^{(N)}}\{E_N\}$ in (12) as probabilities under $\widetilde{G_{\text{VZ}}}$.

We are now ready to analyze (12). First, note that E_N^c can be decomposed as $E_N^c = (\{K_v = 1\} \cap \{N_v \neq N\}) \cup \{N_v = 0\}$ and therefore

$$\widetilde{G_{\text{VZ}}}\{E_N^c\} = \widetilde{G_{\text{VZ}}}\{K_v = 1\} - p_v^N \widetilde{G_{\text{VZ}}}\{K_v = 1 \mid N_v = N\} + (1 - p_v)^N, \tag{13}$$

with the last term corresponding to $\widetilde{G_{\text{VZ}}}\{N_v = 0\}$ and the sum of the first two terms corresponding to $\widetilde{G_{\text{VZ}}}\{\{K_v = 1\} \cap \{N_v \neq N\}\}$. Note that $(\widetilde{G^{(N)}}\{K_v = 1\})_{N \in N}$ and $(\widetilde{G^{(N)}}\{K_v = 1 \mid N_v = n_v\})_{n_v \in \mathbb{N}}$ (well defined for any $N = f(n) \geq n$) are non-increasing sequences of elements in $[0, 1]$. This is the case since they can be seen as the probability $\widetilde{G_{\text{VZ}}}$ of non-increasing sequences of events. The two sequences are thus convergent.

For any $p_v \in (0, 1)$, $(\widetilde{G_{\text{VZ}}}\{E_N^c\})_{N \in \mathbb{N}}$ in (13) has limit equal to $\lim_{N \to \infty} \widetilde{G^{(N)}}\{K_v = 1\}$ (since p_V^N and $(1 - p_v)^N$ go to 0). Let then $g^\infty = \lim_{N \to \infty} \widetilde{G^{(N)}}\{K_v = 1\}$, and let $g_v^\infty = \lim_{n_v \to \infty} \widetilde{G^{(N)}}\{K_v = 1 \mid N_v = n_v\}$. Since K_v depends on $Z_1, \ldots, Z_N$ only indirectly through the N_v units allocated in $\mathbf{Z}_v$ and $N_v/N \to p_v$ a.s. (see the proof of Theorem 1 for more discussion), the two limits are equal, i.e., $g^\infty = g_v^\infty$. We shall show that they equal 0 for several Gibbs-type priors, implying that the GARP will go to the relaxed model, that is, $\widetilde{G^{(N)}}\{E_N\} \to 1$ as $N \to \infty$. Table 2 summarizes the results for the earlier four examples. We use $n_v \leq N$ and for any sequences a_n and b_n, we write $a_n \asymp b_n$ if and only $\lim_n a_n/b_n = 1$.

Table 2: $\widetilde{G^{(N)}}\{K_v = 1 \mid N_v = n_v\}$, limit g_v^∞ and asymptotic rate as $n_v \to \infty$ for Examples 1 (M_v-dimensional symmetric DM, with $M_v > 1$), 2 (Gnedin), 3 (DP) and 4 (PYP).

	$g_{n_v} \equiv \widetilde{G^{(N)}}\{K_v = 1 \mid N_v = n_v\}$		
Ex.	$g_{n_v} =$	$g_{n_v} \asymp$	$g_v^\infty \equiv \lim_{n_v \to \infty} g_{n_v}$
1	$\frac{(\rho)_{n_v}}{(\rho M_v)_{n_v}} M_v$	$\frac{\Gamma(\rho M_v) M_v}{\Gamma(\rho)} n_v^{\rho(1-M_v)}$	0
2	$\frac{\gamma n_v}{\gamma + n_v - 1}$	γ	$\gamma \in (0, 1)$
3	$\frac{\Gamma(\alpha+1)(n_v-1)!}{\Gamma(\alpha+n_v)}$	$\Gamma(\alpha+1) n_v^{-\alpha}$	0
4	$\frac{(1-\sigma)_{n_v-1}}{(\alpha+1)_{n_v-1}}$	$\frac{\Gamma(\alpha+1)}{\Gamma(1-\sigma)} n_v^{-(\alpha+\sigma)}$	0

Theorem 1. *Under the relaxed model $\widetilde{G^{(N)}}$ we have $g^\infty = g_v^\infty = \lim_{N \to \infty} \widetilde{G^{(N)}}\{E_N^c\}$ with $g^\infty = 0$ under the symmetric Dirichlet, the DP, the PYP, and $g^\infty = \gamma \in (0, 1)$ under the Gnedin process. The asymptotic rates of g_{n_v} are given in the second column of Table 2.*

Theorem 1 and (7) show that performing prior elicitation and posterior simulation based on the (analytically and computationally) simpler relaxed model $\widetilde{G^{(N)}}$ becomes practically

attractive. Table 2 also provides the rate at which $\widetilde{G^{(N)}}(E_N^c)$ (where the two models differ) converges. For instance, when $\widetilde{G^{(N)}}(E_N^c) \approx 0$ (in Theorem 1), it is immediate to consider p_v as the prior proportion of observations assigned to vertex clusters under $\widetilde{G^{(N)}}$ for any sample size N. Another important consequence of Theorem 1 and (7) is that we can effectively sample from the prior GARP model with an acceptance-rejection method that proposes a realization from the simple relaxed model $\widetilde{G^{(N)}}$ having theoretical guarantees that the acceptance probability is around 1 in most of the cases. Also with the convergence of $\widetilde{G^{(N)}}(E_N^c)$ to $\gamma > 0$ under the Gnedin process, the approximation remains attractive, as rejection sampling remains practically feasible with known acceptance probability $\widetilde{G^{(N)}}(E_N)$ going to $1 - \gamma$ (instead of 1, under the other models), where γ is a hyperparameter that we can control.

Finally, in most examples, the relaxed model $\widetilde{G^{(N)}}$ approaches the GARP $G^{(N)}$ as the sample N increases in an even stronger way.

Theorem 2. *Under $\widetilde{G^{(N)}}$ with symmetric Dirichlet, DP or PYP ($\sigma \geq 0$) in* (4)

$$G_{VZ}\{E_N \text{ eventually}\} = 1. \tag{14}$$

Thus, for any $k \in \mathbb{N}$ and any possible set of points $a_k = (\mathbf{v}_{1:N+k}, \mathbf{z}_{1:N+k})$

$$\widetilde{G}_{VZ}\left\{\left\{G^{(N+k)}(a_k \mid \mathbf{V}_{1:N}, \mathbf{Z}_{v,N}) = \widetilde{G^{(N+k)}}(a_k \mid \mathbf{V}_{1:N}, \mathbf{Z}_{v,N})\right\} \text{ eventually}\right\} = 1. \tag{15}$$

Under $\widetilde{G^{(N)}}$ with the Gnedin process we have $\widetilde{G_{VZ}}\{E_N \cup \{M_v = 1\} \text{ eventually}\} = 1$ and $\widetilde{G}_{VZ}\left\{\left\{G^{(N+k)}(a_k \mid \mathbf{V}_{1:N}, \mathbf{Z}_{v,N}) = \widetilde{G^{(N+k)}}(a_k \mid \mathbf{V}_{1:N}, \mathbf{Z}_{v,N})\right\} \cup \{M_v = 1\} \text{ eventually}\right\} = 1.$

In words, almost surely either the predictive pmf under the GARP and the relaxed will eventually coincide or (under $\widetilde{G^{(N)}}$ with the Gnedin process) there is only one possible vertex-cluster for any $N \in \mathbb{N}$. The latter has a positive probability $\widetilde{G_{VZ}}\{M_v = 1\} = \gamma \in (0, 1)$ for the Gnedin process.

5 Finite Exchangeability and Projectivity

Under the GARP the distribution of the sample is (finitely) exchangeable, that is the marginal law of $(\boldsymbol{y}_i)_{i=1}^N$ from (1)–(4) is invariant with respect to permutations of the labels $1, \ldots, N$. This homogeneity assumption entails that the order in which we look at the observations does not affect the prior and the inferential results, as it should. The same homogeneity assumption is true for the graph-aligned random partition induced by $(V_i, Z_i)_{i=1}^N$. We discuss some more details of homogeneity assumptions in the model. We will write $G^{(N)}$ for different distributions implied by the GARP model (1)–(4), with the specific distribution being clear from the argument of $G^{(N)}(\cdot)$.

Finite EPPF. Let Ψ_N denote the random partition of observations $[N]$ defined by clustering i and j together if and only if $\boldsymbol{\theta}_i = \boldsymbol{\theta}_j$ (recall that $\boldsymbol{\theta}_i = \boldsymbol{\theta}^*_{Z_i}$). Under the GARP model Ψ_N is an exchangeable random partition with dependent cluster-specific parameters. We introduce the notion of finite EPPF (fEPPF) to characterize the distribution of such random partitions: $G^{(N)}\{\Psi_N = \{C_1, \ldots, C_K\}\} = \text{fEPPF}_K^{(N)}(c_1, \ldots, c_K)$, where $(c_1, \ldots, c_K) = (|C_1|, \ldots, |C_K|)$ are the cluster sizes (in a given arbitrary order). Note that $\{c_1, \ldots, c_K\}$ is a sufficient statistic for an exchangeable random partition. Here K denotes the number of clusters, i.e., $K = K_v + K_e$. The fEPPF is a symmetric function of a composition of N (positive integers that sum up to N). The fEPPF induced by the GARP can be obtained via marginalization of the probability function (4) of the graph-aligned random partition. Several expressions can be aggregated via probabilistic invariance.

Proposition 3. *Under the* GARP

$$\text{fEPPF}_K^{(N)}(|C_1|, \ldots, |C_K|) \propto \sum_{N_v=1}^{N} \left\{ \binom{N}{N_v} p_v^{N_v} (1-p_v)^{N-N_v} \right.$$
$$\left. \sum_{K_v=1}^{M_v} \left[\binom{M_e}{K-K_v} \sum_{(n_1,\ldots,n_{K_v})} \text{EPPF}_{K_v}^{(N_v)}(n_1, \ldots, n_{K_v}) \text{DM}_{M_e}^{(N-N_v)}((n_{k,k_{v'}})_{k<k'_v}). \right] \right\} \quad (16)$$

In the last sum, for given $(n_1, \ldots, n_{K_v})$ the cardinalities $n_{k,k'}$ of edge-clusters are implied by the remaining elements of $(|C_1|, \ldots, |C_K|)$ that are not matched with the vertex-cluster cardinalities n_k. The exact range of the sums is stated in Section S.6.5 of the supplementary materials. Essentially, $\{n_1, \ldots, n_{K_v}\} \cup \{n_{k,k'} : k < k'\} = \{c_1, \ldots, c_K\}$. Moreover, the normalization constant in (16) is $1/\widetilde{G^{(N)}}\{E_N\}$, which we studied in detail before.

A common stronger assumption in the literature on random partitions is that the observed data $(\boldsymbol{y}_i)_{i=1}^N$ are a subset of an infinite (thus unobservable) sequence of exchangeable random variables. This assumption does not apply to the GARP – see below. However, if the assumption applies then the exchangeable random partition of the sample can be seen as a projection of an exchangeable random partition of the natural numbers $\mathbb{N}$ to the set $[N]$. Formally, this is equivalent to assuming:

(a) each random partition Ψ_N is exchangeable over $[N]$;

(b) the sequence of random partitions $(\Psi_N)_{N=1}^{\infty}$ is Kolmogorov consistent, that is, Ψ_n is equal in distribution to the restriction of Ψ_N to $[n]$ for any $1 \leq n \leq N$.

Note that, although we stated the properties for the random partition, the same definitions hold for other sequences of random variables, such as the sample $(\boldsymbol{y}_i)_{i=1}^N$. As done in, e.g., Betancourt *et al.* (2022) we refer to (a) as *finite exchangeability*, (b) as *projectivity*, and to their combination as *infinite exchangeability*.

Proposition 4. *The graph-aligned random partition induced by* $(V_i, Z_i)_{i=1}^N$*, the sample* $(\boldsymbol{y}_i)_{i=1}^N$ *and the random partition* Ψ_N *are finitely exchangeable but they are not a projection of infinite exchangeable processes.*

From a modeling perspective, infinite exchangeability is a natural requirement only to address prediction problems in the most general framework, i.e., prediction for an unbounded number of future observations. In general, it is a desirable property for mathematical convenience to ease prior elicitation (e.g., via de Finetti's representation theorem), to simplify posterior inference, and to study the properties of the model across sample sizes. While the GARP is not infinitely exchangeable, as stated in the previous result, in some cases it turns out to be very close to infinite exchangeability, in the sense that the model is equivalent to an infinitely exchangeable model for large enough N, as discussed next. See also Diaconis and Freedman (1980) for general results and probabilistic characterizations of finite exchangeability and approximate projectivity. The next result shows that in some cases the predictive distribution of the GARP model eventually (i.e., for a large enough sample size N) can be characterized as a projection of the predictive of a limiting infinitely exchangeable model, thus where projectivity holds.

We also characterize the limit via the directing measure, i.e., the law of the random probability in de Finetti's representation theorem. See Table S.1 for a recap of the notation for different distributions.

Theorem 3. *Under the* GARP *model with the* M_v*-symmetric Dirichlet (Example 1) in* (4) *there exists a finite random sample size* $\bar{N}$ *and an infinite dimensional law* $G^{(\infty)}$*, such that for any* $N > \bar{N}$ *the predictive distributions under the GARP model, are* $\widetilde{G_{VZ}}$*-almost surely equal to the predictive distributions under the (Kolmogorov consistent) marginal laws* $\left(G_N^{(\infty)}\right)_{N\in\mathbb{N}}$ *of the infinite-dimensional law* $G^{(\infty)}$*.*

That is, for any possible sequence of sets of points $(a_k)_{k\in\mathbb{N}}$*, with* $a_k = (\mathbf{v}_{1:N+k}, \mathbf{z}_{1:N+k})$

$$\widetilde{G_{VZ}}\Big\{\big\{G_{N+k}^{(\infty)}(a_k \mid \mathbf{V}_{1:N}, \mathbf{Z}_{v,N}) = G^{(N+k)}(a_k \mid \mathbf{V}_{1:N}, \mathbf{Z}_{v,N})\ \forall\, k\big\}\ \textit{eventually}\Big\} = 1. \tag{17}$$

Here $G^{(\infty)}$ *can be characterized by the following gCRP. Let* $M_e^+ = M_v(M_v-1)/2$*.*

$$G^{(\infty)}\{V_i = v, Z_i = z \mid \cdots, \mathbf{V}_{1:N}, \mathbf{Z}_{1:N}\} \propto \begin{cases} p_v \frac{n_k^{-i}+\gamma}{N_v^{-i}+\gamma M_v} & \textit{if } v = 1, \quad z \in [M_v] \\ (1-p_v)\frac{\beta/M_e^+ + n_{k,k'}^{-i}}{\beta/M_e^+ + N_v^{-i}} & \textit{if } v = 0, \quad z = (k,k'). \end{cases} \tag{18}$$

The directing measure characterizing the infinitely exchangeable random parameters that

imply $G^{(\infty)}$ *is defined as*

$$(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \mid P \overset{iid}{\sim} P, \qquad P = p_v \sum_{m=1}^{M_v} \pi_m \delta_{\tilde{\boldsymbol{\theta}}_m} + (1 - p_v) \sum_{k<k' \leq M_v} \pi_{k,k'} \delta_{\tilde{\boldsymbol{\theta}}_{k,k'}} \tag{19}$$

where $(\pi_1, \ldots, \pi_{M_v}) \sim Dir(\rho, \ldots, \rho)$, $(\pi_{k,k'})_{k<k'} \sim Dir(\beta/M_e^+, \ldots, \beta/M_e^+)$, *and* $\tilde{\boldsymbol{\theta}}_m$ *and* $\tilde{\boldsymbol{\theta}}_{k,k'}$ *follow the same distributions as in* (2) *and* (3).

Let $G^{(\infty)}(\boldsymbol{V}, \boldsymbol{Z})$ *denote the pmf of* $\boldsymbol{V}, \boldsymbol{Z}$ *implied by* (19). *It can also be characterized by the projective pmfs for any* $N \in \mathbb{N}$ *(we omit the sub-index* $_N$ *for the finite projections of* $G^{(\infty)}$ *when it is clear from the context):*

$$G^{(\infty)}((V_i, Z_i)_{1:N}) = p_v^{N_v} \text{EPPF}_{M_v}^{(N_v)}(n_1, \ldots, n_{K_v} \mid \alpha, \sigma)/K_v! \\ \times (1 - p_v)^{N_e} DM_{M_e^+}^{(N_e)}((n_{k,k'})_{k<k'_v} \mid \beta/M_e^+). \tag{20}$$

Corollary 1. *Conditional on a given* M_v, *Theorem 3 remains true also under the* GARP *with a Gnedin process (Example 2), with* $G^{(\infty)}(M_v = m) = \widetilde{G_{VZ}}(M_v = m) = \frac{\gamma(1-\gamma)_{m-1}}{m!}$.

See S.6 for the explicit statement of Corollary 1 and the proofs.

Analogous results hold for any MFM. We state it for the special case of the Gnedin process which we introduced and discussed in Section 3.

Note that even if projectivity is not strictly needed to carry out inference under the GARP, approximate projectivity is still a useful property. Without any form of approximate projectivity (i.e., coherence), inference on the partition structure for N observed units would depend on whether or not an investigator plans to collect more data in the future. This would greatly complicate the understanding of model assumption and learning mechanisms.

6 Posterior Inference

Building on the earlier results we develop MCMC algorithms for posterior simulation under the GARP. The algorithms generalize the posterior sampling scheme for the CRP under a DP mixture (Neal, 2000) and under Gibbs-type mixtures. To derive tractable full conditional distributions that are easy to sample from, we exploit the representation of the GARP as a truncated composition of Gibbs-type priors derived in Section 2.3.

In this way, we can exploit the product partition form of the pmf under the relaxed model to simplify the expressions of the conditional probability in the predictive (i.e., the composition of gCRPs) and full conditional distributions. Expressions reduce to simple ratios.

In general, without projectivity and composition of product partition EPPF, it is not possible to generalize a priori (and a posteriori) tractable Pólya urn schemes and thus

tractable marginal algorithms such as the ones in Neal (2000). Projectivity allows us to evaluate conditional probabilities (of cluster membership) as ratios of the same EPPFs over different N. Under the specific product form of the EPPF for Gibbs-type priors, this ratio reduces to a simple expression (De Blasi *et al.*, 2015).

Specifically, the relaxed model $\widetilde{G^{(N)}}$ is a hierarchical composition of Kolmogorov consistent EPPFs with product partition forms (Sections 2.2 and 3) that thus induce a tractable (a priori) composition of gCRPs (Section 2.3). This allows us to derive the following efficient marginal sampler. See Section S.4.1 in the supplementary materials for details.

For an explicit statement of Gibbs sampling transition probabilities, we introduce the notation $I_i = \mathbb{1}(\{n_k^{-i} = 0\} \cap \{\sum_{k' \neq k} \left(n_{k,k'} + n_{k',k}\right) > 0\})$ as an indicator for violating the support of the GARP in (4). That is, $I_i = 1$ if removing i from its current cluster removes the last unit in a vertex-cluster k (for some k) and it leaves an edge-cluster (k, k') (for some $k' \neq k$) without adjacent vertex-cluster k.

We then have the following full conditional probabilities.
`(1)` Sample (V_i, Z_i) form $G^{(N)}(V_i, Z_i \mid \cdots)$. If $I_i = 1$ we do not move. Otherwise sample from $G^{(N)}\{V_i = v, Z_i = z \mid \cdots\} \propto$

$$\begin{cases} p_v \frac{W_{N_v, K_v^{-i}}}{W_{N_v - 1, K_v^{-i}}} (n_k^{-i} - \sigma) \mathrm{N}(\boldsymbol{y_i} \mid \boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*) & \text{if } v = 1, \quad z \in [K_v^{-i}] \\ p_v \frac{W_{N_v, K_v^{-i}+1}}{W_{N_v - 1, K_v^{-i}}} g_{\text{new}}(\boldsymbol{y}_i) & \text{if } v = 1, \quad z = K_v^{-i} + 1 \\ (1 - p_v) \frac{\beta/M_e + n_{k,k'}^{-i}}{\beta/M_e + N_v^{-i}} \mathrm{N}(\boldsymbol{y_i} \mid \boldsymbol{\mu}_{k,k'}^*, \boldsymbol{\Sigma}_{k,k'}^*) & \text{if } v = 0, \quad z = (k, k'), \end{cases}$$

where

$$g_{\text{new}}(\boldsymbol{y}_i) = \int \mathrm{N}(\boldsymbol{y}_i \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \mathrm{dNIW}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \boldsymbol{\mu}_0, \lambda_0, \kappa_0, \boldsymbol{\Sigma}_0) = \mathrm{T}_{\lambda_0 - 1}\left(\boldsymbol{y}_i \mid \boldsymbol{\mu}_0, \frac{\kappa_0 + 1}{\kappa_0(\lambda_0 - 1)}\right)$$

is the pdf of a generalized Student-T distribution of degree $\lambda_0 - 1$.
`(2)` Sample the vertices parameters $(\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*)$ from

$$G^{(N)}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \mid \cdots) \propto \underbrace{\mathrm{NIW}\big(\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^* \mid \hat{\boldsymbol{\mu}}, \hat{\nu}, \hat{\kappa}, \hat{\boldsymbol{\Sigma}}\big)}_{p^0(\boldsymbol{\theta}_k^*)} \times \prod_{k' \neq k} \mathrm{N}(\boldsymbol{y}_i \mid \boldsymbol{\mu}_{k,k'}^*, \boldsymbol{\Sigma}_{k,k'}^*)$$

where in the last product for $k' < k$ we interpret $\boldsymbol{\theta}_{k,k'}^*$ as $\boldsymbol{\theta}_{k,k'}^* \equiv \boldsymbol{\theta}_{k',k}^*$, and $\hat{\nu} = \nu_0 + n_k$, $\hat{\kappa}_k = \kappa_0 + n_k$, $\hat{\boldsymbol{\mu}} = \frac{\kappa_0 \boldsymbol{\mu}_0 + n_k \bar{\boldsymbol{y}}_k}{\hat{\kappa}_k}$ and $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_0 + \boldsymbol{S}_k + \frac{\kappa_0 n_k}{\hat{\kappa}_k}(\bar{\boldsymbol{y}}_k - \boldsymbol{\mu}_0)(\bar{\boldsymbol{y}}_k - \boldsymbol{\mu}_0)^\intercal$, with $\bar{\boldsymbol{y}}_k = \frac{\sum_{i:Z_i=k} \boldsymbol{y}_i}{n_k}$ and $\boldsymbol{S}_k = \sum_{i:Z_i=k}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_k)(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_k)^\intercal$.

If a vertex is isolated, that is, no observations are assigned to any of the possible edges associated with the vertex, then the full conditional in `(2)` reduces to the conjugate NIW posterior distribution $p^0(\boldsymbol{\theta}_k^*)$. In general, the density of the full conditional is proportional to p^0 times the likelihood of the observations assigned to corresponding edges. An effective

transition probability is a Metropolis-Hasting step exploiting p^0 as a proposal.

In step `(1)`, when we create a new vertex-cluster, i.e., if $v = 1$ and $k = K_v^{-i} + 1$, we follow up with a transition probability (2) for the new cluster parameters, that reduces to the conjugate NIW for $\boldsymbol{\theta}_k^*$. Throughout, edge-parameters $\boldsymbol{\theta}_{k,k'}^*$ are always evaluated using the currently imputed adjacent vertex parameters $\boldsymbol{\theta}_k^*, \boldsymbol{\theta}_{k'}^*$.

Note that it is also possible to add an extra transition probability to update Z_i as in (1), but leaving V_i unchanged. Such transition probabilities could lead to a better mixing Markov chain and are analogous to the ones used, for example, in Teh *et al.* (2006) exploiting the Chinese restaurant franchise representation of the hierarchical DP.

In principle, all posterior inference is implemented by appropriate summaries of the posterior Monte Carlo sample. However, how to report point estimates for a random partition or graph is not trivial. There are several proposals in the recent literature, including Wade and Ghahramani (2018), Dahl *et al.* (2022), and Franzolini and Rebaudo (2024). They are based on casting the selection of the reported summary as a decision problem. In Section S.4.2 of the supplemental materials, we discuss an implementation for the GARP.

Finally, like in any mixture model, posterior inference about specific clusters must consider label switching. See, for example, Green (2018) for a discussion. An additional challenge that arises in the proposed model is the distinction between vertex versus edge clusters. Consider, for example, a configuration (A) with 2 vertices and a connecting edge, with cluster-specific parameters $(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\theta}_{1,2}^* = f(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*))$ (as in (3)), versus an alternative configuration (B) with 3 vertices and $(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\theta}_3^*)$ and $\boldsymbol{\theta}_3^* = f(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*)$. While the sampling model (1) remains unchanged under (A) versus (B), we argue that the prior implements a strong preference for (the more parsimonious) model (A).

For given vertex parameters $\boldsymbol{\theta}_1^*$ and $\boldsymbol{\theta}_2^*$, the edge parameter $\boldsymbol{\theta}_{1,2}^*$ in (A) can assume just one value, i.e., its parameter space is the single point $f(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*)$ in the parameter space of the third vertex $\boldsymbol{\theta}_3^*$ in the latter model. In other words, when we consider the joint parameter space $\boldsymbol{\Theta}_0$ of the atoms $\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\theta}_{1,2}^*$ (two vertices and one edge) it is a lower dimensional sub-space of the parameter space $\boldsymbol{\Theta}$ for the three vertices $\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\theta}_3^*$. The NIW prior in (2) on $\boldsymbol{\theta}_j^*$ assigns prior probability 0 to $\boldsymbol{\Theta}_0$, and thus also zero posterior probability. The issue is similar to identifiability related to the replication of terms in a standard mixture model with independent priors on cluster-specific parameters Green (2018).

7 Application to Single-Cell RNA Data

We fit the GARP model for the RNA-seq data shown in Figure 1. Single-cell RNA-seq experiments record cell-specific transcriptional profiles that allow us to infer, for example, cell differentiation or cancer progression. Inference under the GARP model for the data shown in Figure 1 reconstructs transitions of stem cells into fully differentiated cells in a

scRNA-seq experiment on horizontal basal cells from the adult mouse olfactory epithelium. The original data is available on GEO in GSE95601.

The transcriptional profiles map differences in gene expressions due to the development phases of the cells. Stem cells evolve into fully differentiated cells by gradual transcriptional changes, passing through a small number of homogeneous subpopulations of cells. The primary inferential goal is to find these homogeneous subpopulations of cells (i.e., vertex-clusters) and understand the relationships between them aligning such subpopulations on a biologically interpretable graph.

7.1 ScRNAseq Data and Pre-Processing

The raw data is a count matrix with rows corresponding to cells and columns representing different genes. Most of the counts in the matrix are zeros, usually about 90% (the percentage can vary according to the scRNA-seq technology used).

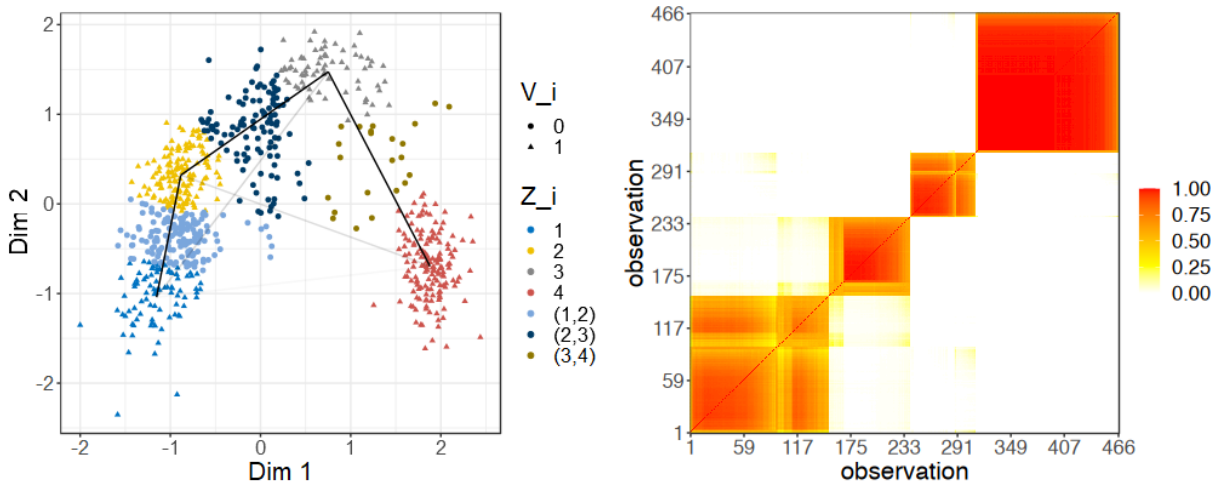
The data set originally contains measurements for 28284 genes in 849 cells with 84% zeros. To extract a lower dimensional signal we implemented pre-processing following the pipeline described in Perraudeau *et al.* (2017) and available in `Bioconductor` (Gentleman *et al.*, 2004). We briefly describe the pipeline. We first discard around 100 low-quality cells and retain the 1000 most variables genes. Next, we normalize the data matrix and extract 50-dimensional biomarkers from the count data, accounting for zero-inflation and over-dispersion of the scRNA-seq data via "Zero-Inflated Negative Binomial Wanted Variation Extraction" (ZINB-WaVE) (Risso *et al.*, 2018). Finally, we reduced the dimensionality to the 2 most relevant markers via multidimensional scaling analysis. The data matrix obtained after pre-processing is denoted by $\boldsymbol{y} = (\boldsymbol{y}_{i,j} : i = 1, \ldots, N, j = 1, 2)$, where the rows represent 747 cells, and the columns record the two final biomarkers. The data is shown in Figure 1.

7.2 Results

We implement inference under the GARP model using the Gnedin process (Example 2) to control the vertex-clustering. We choose the Gnedin process because one of the goals is inference on K_v. The Gnedin process is a particularly attractive Gibbs-type prior for clustering from both, a Bayesian modeling perspective as well as for its frequentist properties of the posterior distribution, as discussed in Section 3.

The posterior estimated GARP places 466 cells into vertex-clusters (main phases) and 281 into ordered edge-clusters (transition phases). Figure 2a summarizes inference. The heat-map in Figure 2b shows the posterior probabilities of co-clustering of pairs of observations, suggesting low posterior uncertainty around the estimated main phases, making the point estimate under the GARP a meaningful posterior summary. The conditional

uncertainty of the graph-alignment of the vertices given the point estimates of the main phases is low. Visual inspection of the results suggests that the model is effectively working as expected. Once we have identified the main phases (vertex-clusters) we find the



(a) (b)

Figure 2: Left Panel: Scatter-plot of the scRNA data. Triangular plot symbols indicate cells assigned to vertices ($V_i = 1$) while the remaining cells are assigned to edges ($V_i = 0$) and are represented with a circular shape. Cells are colored according to the different phases (i.e., Z_i) in the point estimate. The segments denote the edges of the graph and the color is darker if the probability of assigning observations to the edge is greater. Right panel: Posterior probabilities of co-clustering of observations assigned to vertices.

biomarkers that best characterize such clusters, i.e., the most differently expressed genes (DE genes). We rely on the function `findMarkers` of the `Bioconductor` package `scran` (Lun *et al.*, 2016). More precisely, we first perform an exact binomial test to identify DE genes between pairs of groups of cells (vertex-clusters). From that, we identify the 6 most significant biomarkers for each pairwise comparison. For each gene then a combined p-value is computed using Simes multiplicity adjustment applied to all p-values obtained by the pairwise comparisons (Simes, 1986). Note that these p-values are not directly used for ranking and are only used to find the DE genes. Finally, the p-values are consolidated across all genes using the BH method of Benjamini and Hochberg (1995) to implement multiple comparisons under a restriction on false discovery rate (FDR) (Benjamini *et al.*, 2009). The adjusted p-values are reported in Table 3. The reported FDRs are intended only as a rough measure of significance. Note that properly correcting for multiple testing is not generally possible when clusters are based on the same data that is used for the DE testing. Nonetheless, a small FDR remains desirable. Table 3 shows the average within vertex-cluster gene expressions for the selected top 6 biomarkers and corresponding FDRs. The log means expression in the different biomarkers and vertices are also shown in Figure 3. On average the main phases obtained (vertex-clusters) have very different expressions of

the selected biomarkers. Finally, we show the entire distribution of the cells in the different biomarkers and main phases in Figure 4.

DE Genes	Vertex 1	Vertex 2	Vertex 3	Vertex 4	FDR
Slc26a7	408.68	120.96	0.24	0.05	1.30e-23
Pik3c2b	14.15	231.82	105.74	98.38	1.38e-08
Hes6	3.10	21.16	691.19	41.62	4.17e-13
Stmn3	0.43	0.09	23.90	320.38	1.58e-20
Abca13	10.15	337.45	4.25	0.54	8.29e-08
Ccp110	5.75	15.19	1008.68	105.50	7.47e-13

Table 3: Average within vertex-cluster gene expressions and FDRs in the selected top 6 biomarkers.

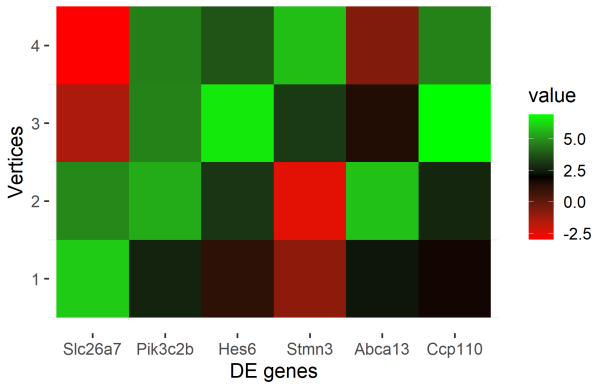


Figure 3: Heatmap of the log mean expressions in top 6 DE genes in the main phases.

7.3 Comparison with Independent Gaussian Mixtures

For comparison, we estimate an independent Gaussian mixture model without edges and cluster alignment (implemented as the GARP model with $p_v = 1$). The posterior distribution of the number of clusters (see Table 4) shows more uncertainty since the model fails to find well-separated clusters, due to the noise that is introduced by the presence of the cells transitioning between the main phases. In other words, including cells in transition in the clustering has reduced the statistical power in detecting homogeneous subpopulations. This is illustrated in Figure 5. Recall that we are using variation of information (VI) loss to summarize the posterior random partition. As a consequence of the increased uncertainty, the point estimate of the clustering of the main phases becomes sensitive to the choice of the loss function. For instance, both the point estimate and the maximum a posteriori estimate of the number of main phases is 4 under GARP, while the earlier is 5 and the latter is 6 under the independent Gaussian mixture model. In the figures, we show the

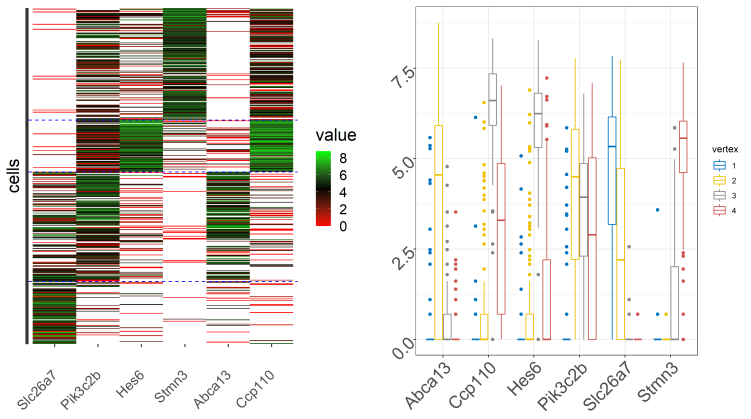
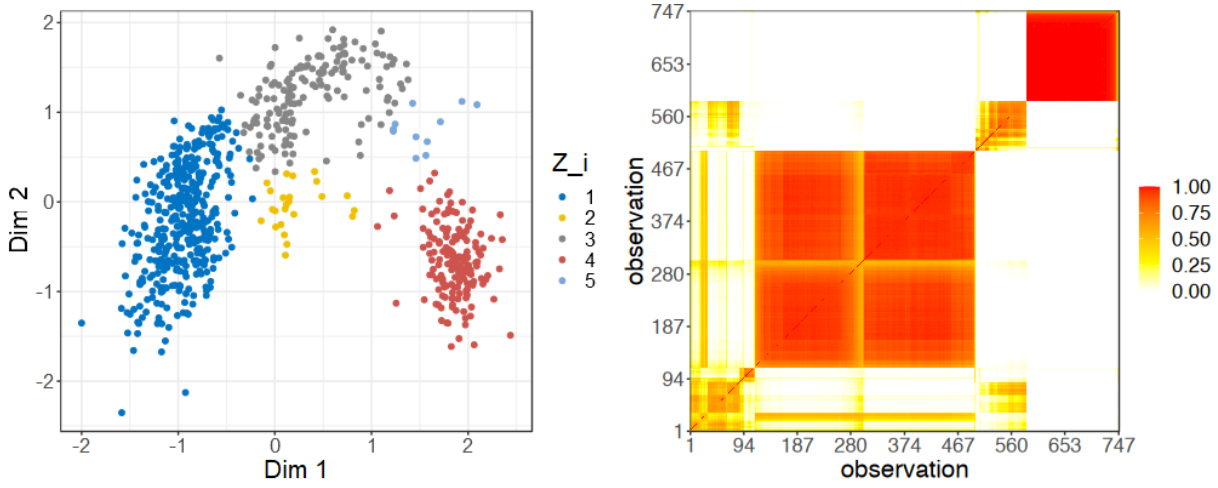


Figure 4: Left panel: Heat-map log genetic expressions in top 6 DE genes in all cells ordered by main phases. The cells are sorted by vertex-cluster memberships and the dashed blue lines separate the cells in the different clusters. Right panel: Boxplot genetic expressions (after $\log(\cdot + 1)$ transformation) in the top 6 DE genes in all cells in the different main phases (vertex-clusters).

estimated cluster arrangement that minimizes the VI loss for coherency in the comparison.



(a) (b)

Figure 5: Results with independent mixtures. Left Panel: Scatter-plot of the scRNA data. Cells are colored according to the different phases in the point estimate. Right panel: Co-clustering posterior probabilities.

Panel A: GARP

k	4	5	6	7	8
$\hat{P}(K_v = k)$	0.7801	0.1951	0.0240	0.0004	0.0004

Panel B: Independent Mixture Model

k	5	6	7	8	9	10	11	12	13
$\hat{P}(K = k)$	0.0692	0.4362	0.3115	0.114	0.0516	0.0128	0.002	0.0012	0.0016

Table 4: Panel A: Estimated posterior of the number of main phases under GARP. Panel B: Estimated posterior of the number of main phases under the independent Gaussian mixture model.

8 Discussion

We proposed a graph-aligned random partition model to infer homogeneous subgroups of observations aligned on a graph, explicitly allowing for units transitioning between the clusters. The motivating applications are single-cell RNA experiments where scientists are interested in understanding fundamental biological processes such as cell differentiation and tumor evolution. Interesting future applications include inference for cell type transitions in a tumor microenvironment. Other extensions could include data integration with other modalities, such as histology data.

Methodological extensions include jointly clustering similar cells *and* genes, via separately exchangeable nested random partition models (Lee *et al.*, 2013b; Lin *et al.*, 2021). Another interesting extension is to combine the results of partially exchangeable random partition models that arise from the compositions of Gibbs-type and species sampling priors (Teh *et al.*, 2006; Camerlenghi *et al.*, 2019; Argiento *et al.*, 2020; Bassetti *et al.*, 2020; Lijoi *et al.*, 2023) to the GARP model with dependent locations. In the context of the scRNA-seq experiment, this would allow inference on multiple single-cell RNA-seq data matrices. In such a way one could borrow information across different measurements while accounting for relevant heterogeneity. Finally, including unit-specific spatial information, the model can be used for spatial clustering with transitions between the clusters.

Acknowledgment

The authors are grateful to the Editor, the Associate Editor, and the anonymous referees, whose feedback have led to a significant improvement in the manuscript. Both authors have been partially supported by NSF/DMS 1952679. Most of the paper was completed while G. R. was a Postdoc at UT Austin. G. R. is also affiliated to "de Castro" Statistics Initiative, Collegio Carlo Alberto, Torino and acknowledges support of MUR - Prin 2022 - Grant no. 2022CLTYP4, funded by the European Union – Next Generation EU.

References

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Stat.*, **2**, 1152–1174.

Argiento, R., Cremaschi, A., and Vannucci, M. (2020). Hierarchical normalized completely random measures to cluster grouped data. *J. Am. Stat. Assoc.*, **115**, 318–333.

Bassetti, F., Casarin, R., and Rossini, L. (2020). Hierarchical species sampling models. *Bayesian Anal.*, **15**, 809–838.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.*, **57**, 289–300.

Benjamini, Y., Heller, R., and Yekutieli, D. (2009). Selective inference in complex research. *Philos. Trans. Royal Soc. A*, **367**, 4255–4271.

Beraha, M., Argiento, R., Möller, J., and Guglielmi, A. (2022). MCMC computations for Bayesian mixture models using repulsive point processes. *J. Comput. Graph. Stat.*, **31**, 422–435.

Betancourt, B., Zanella, G., and Steorts, R. C. (2022). Random partition models for microclustering tasks. *J. Am. Stat. Assoc.*, **117**, 1215–1227.

Camerlenghi, F., Lijoi, A., Orbanz, P., and Prünster, I. (2019). Distribution theory for hierarchical processes. *Ann. Stat.*, **47**, 67–92.

Dahl, D. B., Johnson, D. J., and Müller, P. (2022). Search algorithms and loss functions for Bayesian clustering. *J. Comput. Graph. Stat.*, **31**, 1189–1201.

De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prünster, I., and Ruggiero, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE Trans. Pattern Anal. Mach. Intell.*, **37**, 212–229.

Diaconis, P. and Freedman, D. (1980). Finite exchangeable sequences. *Ann. Probab.*, **8**, 745–764.

Ewens, W. J. (1990). Population genetics theory - the past and the future. In S. Lessard, editor, *Mathematical and Statistical Developments of Evolutionary Theory*, volume 299, pages 177–227. Springer.

Franzolini, B. and Rebaudo, G. (2024). Entropy regularization in probabilistic clustering. *Stat. Methods. Appt.*, **in press**.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, 1–16.

Gnedin, A. V. (2010). A species sampling model with finitely many types. *Electron. Commun. Probab.*, **15**, 79–88.

Gnedin, A. V. and Pitman, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *J. Math. Sci.*, **138**, 5674–5685.

Green, P. J. (2018). Introduction to finite mixtures. In S. Fruhwirth-Schnatter, G. Celeux, and C. P. Robert, editors, *Handbook of Mixture Analysis*, pages 3–20. Chapman and Hall/CRC.

Green, P. J. and Richardson, S. (2001). Modelling heterogeneity with and without the Dirichlet process. *Scand. J. Stat.*, **28**, 355–375.

Lee, J., Quintana, F. A., Müller, P., and Trippa, L. (2013a). Defining predictive probability functions for species sampling models. *Stat. Sci.*, **28**, 209–222.

Lee, J., Müller, P., Zhu, Y., and Ji, Y. (2013b). A nonparametric Bayesian model for local clustering with application to proteomics. *J. Am. Stat. Assoc.*, **108**, 775–788.

Lijoi, A., Mena, R. H., and Prünster, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *J. Am. Stat. Assoc.*, **100**, 1278–1291.

Lijoi, A., Mena, R. H., and Prünster, I. (2007a). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika*, **94**, 769–786.

Lijoi, A., Mena, R. H., and Prünster, I. (2007b). Controlling the reinforcement in Bayesian non-parametric mixture models. *J. R. Stat. Soc. Series B Stat. Methodol.*, **69**, 715–740.

Lijoi, A., Prünster, I., and Rebaudo, G. (2023). Flexible clustering via hidden hierarchical Dirichlet priors. *Scand. J. Stat.*, **50**, 213–234.

Lin, Q., Rebaudo, G., and Müller, P. (2021). Separate exchangeability as modeling principle in Bayesian nonparametrics. *Preprint at arXiv: 2112.07755*.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. density estimates. *Ann. Stat.*, **12**, 351–357.

Lun, A. T., McCarthy, D. J., and Marioni, J. C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research*, **5**, 1–64.

Miller, J. W. and Harrison, M. T. (2018). Mixture models with a prior on the number of components. *J. Am. Stat. Assoc.*, **113**, 340–356.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Stat.*, **9**, 249–265.

Nobile, A. (1994). *Bayesian Analysis of Finite Mixture Distributions*. Ph.D. thesis, Carnegie Mellon Univ.

Nobile, A. and Fearnside, A. T. (2007). Bayesian finite mixtures with an unknown number of components: the allocation sampler. *Stat. Comput.*, **17**, 147–162.

Perraudeau, F., Risso, D., Street, K., Purdom, E., and Dudoit, S. (2017). Bioconductor workflow for single-cell RNA sequencing: normalization, dimensionality reduction, clustering, and lineage inference. *F1000Research*, **6**, 1–28.

Petralia, F., Rao, V., and Dunson, D. B. (2012). Repulsive mixtures. In *Adv. Neural Inf. Process. Syst.*, volume 25, pages 1889–1897.

Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. *Lect. Notes-Monogr. Series*, **30**, 245–267.

Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, **25**, 855–900.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J. R. Stat. Soc. Series B Stat. Methodol.*, **59**, 731–792.

Risso, D., Perraudeau, F., Gribkova, S., Dudoit, S., and Vert, J.-P. (2018). A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat. Commun.*, **9**, 1–17.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**, 751–754.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *J. Am. Stat. Assoc.*, **101**, 1566–1581.

Wade, S. and Ghahramani, Z. (2018). Bayesian cluster analysis: point estimation and credible balls (with discussion). *Bayesian Anal.*, **13**, 559–626.

Xu, Y., Müller, P., and Telesca, D. (2016). Bayesian inference for latent biologic structure with determinantal point processes (DPP). *Biometrics*, **72**, 955–964.

Supplementary materials of

Graph-Aligned Random Partition Model (GARP)

Giovanni Rebaudo[a] (giovanni.rebaudo@unito.it)
Peter Müller[b] (pmueller@math.utexas.edu)

[a]University of Torino, IT

[b]University of Texas at Austin, USA

S.1 Edge Multivariate Gaussian Mixtures

Figure S.1 shows the contour plot of an edge cluster in $\mathbb{R}^2$.

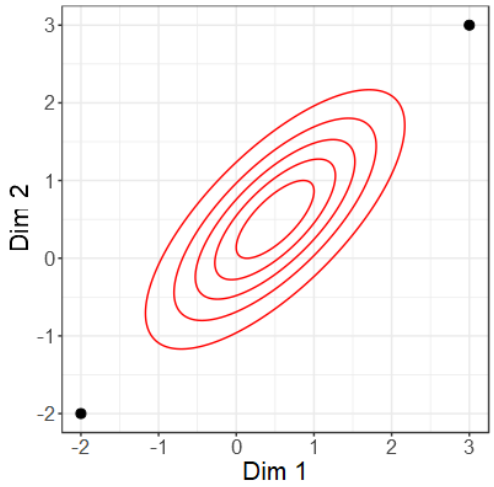


Figure S.1: Elliptical contour plot for an edge-cluster (with hyperparameters, as described in Section S.3), connecting two vertex-clusters k and k' (with locations $\boldsymbol{\mu}_1 = (-2, -2)$ and $\boldsymbol{\mu}_2 = (3, 3)$). The vertices are shown as black bullets located on the contour (not shown) line of the bivariate Gaussian such that 99% of the probability is inside such an ellipse.

Without loss of generality consider an edge connecting the two vertex-clusters, $k = 1$ and $k' = 2$, with cluster-specific parameters $\boldsymbol{\mu}_1^*$ and $\boldsymbol{\mu}_2^*$. The edge-cluster is centered around the half-point $\boldsymbol{\mu}_{1,2}^* = \frac{\boldsymbol{\mu}_1^* + \boldsymbol{\mu}_2^*}{2}$. The following construction defines $\boldsymbol{\Sigma}_{1,2}^*$ such that the edge is aligned along the connecting line $L_{1,2}$, as described in Section 2.1 of the main manuscript. Let $\boldsymbol{e} = \frac{\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^*}{||\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^*||}$, where $||\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^*||$ denotes the Euclidean distance between $\boldsymbol{\mu}_1^*$ and $\boldsymbol{\mu}_2^*$. Let $\boldsymbol{P} = \boldsymbol{e}\boldsymbol{e}^\intercal$ be the perpendicular projection matrix such that for any $\boldsymbol{y}_i \in \mathbb{R}^p$, $\boldsymbol{y}_i^{(p)} = \boldsymbol{P}\boldsymbol{y}_i$ is the perpendicular projection of $\boldsymbol{y}_i$ onto the connecting line between $\boldsymbol{\mu}_1^*$ and $\boldsymbol{\mu}_2^*$. Let $\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^\intercal$ denote a singular value decomposition (SVD) with $\boldsymbol{D} = \text{diag}(1, 0, \ldots, 0)$. Thus $\tilde{\boldsymbol{R}} = \boldsymbol{Q}^\intercal$ is the rotation matrix such that $\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{R}}\boldsymbol{y}_i$ is the rotation of $\boldsymbol{y}_i$ in the new axes where the first axis is the line connecting $\boldsymbol{\mu}_1^*$ and $\boldsymbol{\mu}_2^*$ and the others are the orthogonal directions. Now, we define $\tilde{\boldsymbol{S}} = \text{diag}(||\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^*||r_0, r_1, \ldots, r_1)$ and $\boldsymbol{\Sigma}_{1,2}^* = \tilde{\boldsymbol{R}}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{R}}^\intercal$.

Under this construction, the term in the mixture of normal sampling model (1) corresponding to the edge (k, k') is such that the Gaussian component projected onto the connecting line $L_{1,2}$ has a standard deviation $r_0 \, ||\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^*||$, implying lower likelihood for edges between distant vertices' locations. The standard deviations of the independent Gaussian distributions on the projection onto $L_{1,2}^{\perp}$ is r_1.

S.2 Composition of Discrete Random Probabilities

Let $\boldsymbol{\theta}_i = \boldsymbol{\theta}^*_{Z_i}$ denote the normal moments in the sampling model (1). As a third characterization of the proposed GARP, we define $\widetilde{G^{(N)}}$ as a graph-aligned random partition (with unique atoms) implied by the ties under conditional i.i.d. sampling of $\boldsymbol{\theta}_i$, with separate models for vertex and edge-clusters. For vertex-clusters

$$\begin{aligned} &\boldsymbol{\theta}_i \mid V_i = 1, \boldsymbol{V}, P_v \overset{\text{iid}}{\sim} P_v, \\ &P_v = \sum_{m=1}^{M_v} \pi_m \delta_{\tilde{\boldsymbol{\theta}}_m} \sim \text{Gibbs-Type Process}, \end{aligned} \tag{S.1}$$

where M_v is the number of atoms of the discrete random probability P_v that is a Gibbs-type process and can be finite, as in the finite symmetric DM case, infinite as in the DP, and PYP case, or be a random variable on $\mathbb{N}$ as in the MFM case. Thus $(\pi_m)_{m=1}^{M_v}$ are the random weights (that are sampled independently from the atoms) from the distribution on the simplex associated with the Gibbs-type process. The unique atoms $\tilde{\boldsymbol{\theta}}_m$ of P_v are i.i.d. samples from the NIW distribution in (2). Note that the unique sampled vertex parameters $\boldsymbol{\theta}_v^* = \{\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_{K_v}^*\}$ are a subset of $\{\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_{M_v}\}$.

The edge-clusters are implied by

$$\begin{aligned} &\boldsymbol{\theta}_i \mid V_i = 0, \boldsymbol{V}^{-i}, K_v, \boldsymbol{\theta}_v^*, E_N \overset{\text{iid}}{\sim} P_e, \\ &P_e = \sum_{1 \le k < k' \le K_v} \pi_{k,k'} \delta_{(\boldsymbol{\theta}^*_{k,k'})}. \end{aligned} \tag{S.2}$$

Recall that $M_e = K_v(K_v - 1)/2$. The random weights follow a symmetric M_e-dimensional Dirichlet with hyper-parameter β/M_e,

$$(\pi_{k,k'})_{1 \le k < k' \le K_v} \sim \text{Dir}(\beta/M_e, \ldots, \beta/M_e). \tag{S.3}$$

Finally, recall that (8) might generate $N_e > 0$, even when (9) implies $M_e = 0$. For this case, we define for completeness $\widetilde{G^{(N)}}\{\boldsymbol{\theta}_i = (\mathbf{0}, I_p) \mid V_i = 0, \boldsymbol{Z}^{-i}, E_N^c\} \equiv 1$, where $\mathbf{0}$ is a p-dimensional vector of 0's and I_p is a $p \times p$-dimensional identity matrix (without implications for $G^{(N)}$, due to the truncation to E_N in (7)).

From the characterizations of Gibbs-type and DM processes, it is straightforward to show that the aforementioned discrete conditional random probability models for the parameters characterize the GARP as stated in the following proposition.

Proposition S.5. *The random partition structure of the* GARP *model* (4) *and the vertex- and edge-parameters distributions can be characterized as the configuration of ties implied by the truncation sampling model in* (7), (8), *(S.1),* (S.2), *and* (S.3).

S.3 Hyperparameters Settings

In both the application and the simulation we set $\gamma = 0.5$ for the Gnedin process controlling the vertex-clusters and $\beta = 0.5$ for the symmetric DM with hyperparameter $0.5/M_v$ to favor the sparsity of the graph. Moreover, for the choice of the hyperparameters of the NIW we set $\boldsymbol{\mu}_0 = \bar{\boldsymbol{y}}$, $\kappa_0 = 0.001$, $\nu_0 = 100$, $\boldsymbol{\Lambda}_0 = \xi^2 \mathbf{I}$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Lambda}_0^{-1}$. For scenarios in which the clusters are well separated, we recommend a large value of ξ^2 (that we set equal to 150), while we recommend a smaller value of ξ^2 (that we set equal to 15) if the data are not well separated in the Euclidean space. Moreover, in both the application and the simulation we set $r_0^2 = 4(\chi^2_{2,1-\alpha})^{-1}$ and $r_1^2 = (2\chi^2_{2,1-\alpha})^{-1}$, where $\chi^2_{2,1-\alpha}$ is the quantile of order $1-\alpha$ (we set $\alpha = 1\%$) of a Chi-squared distribution with 2 degrees of freedom to have the desired eccentricity of the elliptical contour plot of the edge as well as the 99%-level of the contour plot not too spread. To obtain that, recall that c-level counter-plot of multivariate Gaussian density, such as the edge Gaussian in (1), are points $\boldsymbol{y} \in \mathbb{R}^d$ such that $(\boldsymbol{y} - \boldsymbol{\mu}_{k,k'})^\intercal \boldsymbol{\Sigma}_{k,k'}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{k,k'})$ is constant, that is the contour levels are ellipsoid centered at $\boldsymbol{\mu}_{k,k'}$. Finally, note that if $\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{\mu}_{k,k'}, \boldsymbol{\Sigma}_{k,k'})$ then,

$$(\boldsymbol{y} - \boldsymbol{\mu}_{k,k'})^\intercal \boldsymbol{\Sigma}_{k,k'}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{k,k'}) \sim \chi^2_d \,,$$

where χ^2_d denotes the Chi-square distribution with d degrees of freedom.

Visually the contour plots of such edge pdf are shown in Figure S.1 and the data sampled from such configuration looks like the one in Figure S.2.

S.4 Implementing Posterior Inference

S.4.1 Use of the Relaxed Model $\widetilde{G^{(N)}}$ in Posterior Simulation

We discuss in more detail the use of the projectivity property of $\widetilde{G^{(N)}}$ to define a Pólya urn scheme for a tractable marginal posterior simulation algorithm. First, recall that the relaxed model $\widetilde{G^{(N)}}$ can be seen as a hierarchical composition of a Kolmogorov consistent EPPFs with product partition forms (Sections 2.2 and 3), which implies tractable

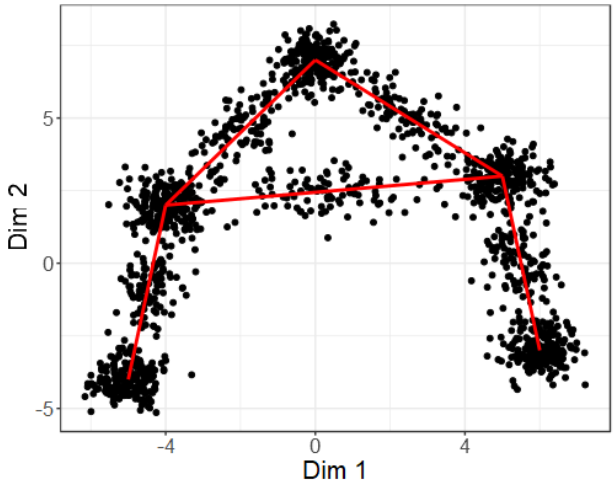


Figure S.2: Scatter-plot of the simulated data. The red segments represent the edges that connect the truth vertices.

expressions for $\widetilde{G^{(N)}}(V_i, Z_i \mid \boldsymbol{V}^{-i}, \boldsymbol{Z}^{-i}) = \frac{\widetilde{G^{(N)}}(\boldsymbol{V},\boldsymbol{Z})}{\widetilde{G^{(N)}}(\boldsymbol{V}^{-i},\boldsymbol{Z}^{-i})}$ under the relaxed model $\widetilde{G^{(N)}}$.

To derive then the desired full conditional distributions under $G^{(N)}$ we note, e.g., that if $I_i = 0$ for $(\boldsymbol{V}, \boldsymbol{Z})$ (recall the definition of I_i in Section 6), then

$$G^{(N)}\{V_i = v, Z_i = z \mid \cdots\} \propto \frac{\widetilde{G^{(N)}}(\boldsymbol{V}, \boldsymbol{Z}) \prod_{j\in[N]} \mathrm{N}(\boldsymbol{y}_j \mid \boldsymbol{\theta}^*_{Z_j})}{\widetilde{G^{(N)}}(\boldsymbol{V}^{-i}, \boldsymbol{Z}^{-i}) \prod_{j\in[N]^{-i}} \mathrm{N}(\boldsymbol{y}_j \mid \boldsymbol{\theta}^*_{Z_j})},$$

for any $v \in \{0, 1\}$ and $z \in \boldsymbol{Z}^{-i}$. Moreover, when $I_i = 0$ for $(\boldsymbol{V}, \boldsymbol{Z})$, the marginal probability in the denominator is equal to the one in a Kolmogorov consistent model (i.e., if $I_i = 0$, $\widetilde{G^{(N)}}(\boldsymbol{V}^{-i}, \boldsymbol{Z}^{-i}) = \widetilde{G^{(N-1)}}(\boldsymbol{V}^{-i}, \boldsymbol{Z}^{-i})$, up to a normalization constant) and this allows us to generalize then tractable marginal samplers such as in Neal (2000) or Teh et al. (2006) relying on the characterization of the GARP via a composition of gCRP in Section 2.3.

S.4.2 Point Estimates for the GARP Random Partition

How to choose good summaries (i.e., point estimates) for reporting posterior inference on functionals of interest can be a fundamental and nontrivial question in Bayesian analysis. It is especially challenging if the object of interest is a partition or a graph. To define a posterior point estimate and perform uncertainty quantification we build on the existing literature of posterior point estimates of random partition based on a decision-theoretic approach (Wade and Ghahramani, 2018; Dahl et al., 2022b) generalizing the results for the more challenging case of GARP. We propose a point estimate for the GARP as follows.

(1) Assign observations to vertices versus edges using the posterior mode,

$$\hat{V}_i = 1 \text{ if } \bar{V}_i \equiv \sum_t \frac{V_i^{(t)}}{T} > 0.5,$$

where T is the Monte Carlo sample size, and $V_i^{(t)}$ is the imputed value in iteration t of the MCMC simulations. The uncertainty around the point estimate is quantified using $(1 - \hat{V}_i)\bar{V}_i + \hat{V}_i(1 - \bar{V}_i)$.

(2) Given $\hat{\boldsymbol{V}}$ we find a point estimate $\hat{\boldsymbol{Z}}_v$ for the partition of vertex units by minimizing the variation of information loss (VI) (Meilă, 2007) as suggested by Wade and Ghahramani (2018) and implemented in the `R` package `salso` (Dahl et al., 2022a). Alternative loss functions can be used as needed for different applications (See e.g., Binder, 1978; Franzolini and Rebaudo, 2023). For uncertainty quantification, we report the heat-map with the posterior probabilities of co-clustering.

(3) Given $\hat{\boldsymbol{V}}$ and $\hat{\boldsymbol{Z}}_v$ we find a point estimate $\hat{\boldsymbol{Z}}_e$ and conditional uncertainty quantification for $\boldsymbol{Z_e}$ using the posterior probability of observations being assigned to the different edges. We evaluate conditional posterior probabilities of assigning the remaining observations to the possible edges,

$$G(\boldsymbol{Z_e} \mid \cdots) \propto \prod_{k<k'} \Gamma(n_{k,k'} + \beta/M_e) \prod_{C_{k,k'}} \mathrm{N}(\boldsymbol{y}_i \mid \boldsymbol{\mu}^*_{k,k'}, \boldsymbol{\Sigma}^*_{k,k'}). \tag{S.4}$$

Here the first product goes over all (k, k') with $1 \leq k < k' \leq K_v$ and the second over the $\boldsymbol{y}_i$ such that $z_i = (k, k')$, i.e., the set $C_{k,k'}$. Probabilities (S.4) are evaluated by Rao-Blackwellization (Robert and Roberts, 2021), using the full conditionals

$$G\{Z_i = (k, k') \mid V_i = 0, \cdots\} \propto (n^{-i}_{k,k'} + \beta/M_e)\mathrm{Norm}(\boldsymbol{y}_i \mid \boldsymbol{\mu}^*_{k,k'}, \boldsymbol{\Sigma}^*_{k,k'}).$$

We visualize $p(\boldsymbol{Z_e} \mid \hat{\boldsymbol{Z}}_v, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ by adding edges between vertices with color intensity proportional to the sum over observations assigned to edges (i.e., $\hat{V}_i = 0$) of the probability that such observations will be assigned to the different edges (k, k')'s.

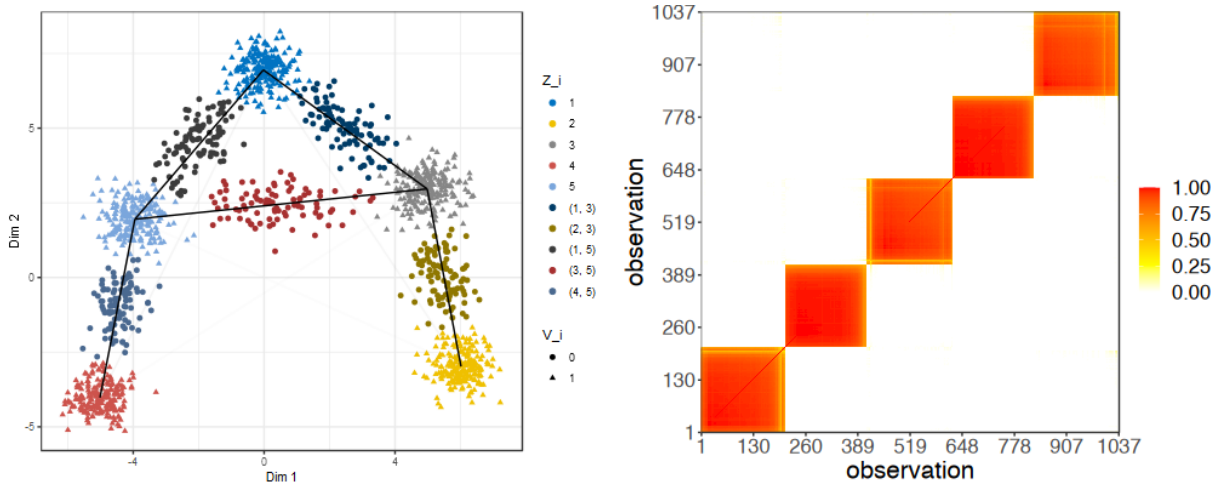
S.5 Simulation Studies

We carried out a simulation study under a well-specified and a miss-specified data generating truth to assess inference under finite sample size scenarios. We set up simulation truths close to the mouse data. The data are simulated from a 5 vertex mixture with $n_k = 200$ observations in each vertex and an additional $n_{k,k'} = 100$ observations around 5 assumed edges.

S.5.1 Well Specified Scenario

In the first simulation scenario, we assume a simulation truth with $K_v = 5$ vertex clusters with cluster-specific Gaussians with mean vectors $(-5, -4)$, $(-4, 2)$, $(0, 7)$, $(5, 3)$ and $(6, -3)$, and a common covariance matrix $\text{diag}(0.25, 0.25)$. Observations assigned to edge components are sampled from a Gaussian mixture with cluster-specific kernels as in (3). The simulated $N = 1500$ observations are shown in Figure S.3a.

Figure S.3 shows that the GARP was able to recover the simulated truth in the point estimate. Moreover, the uncertainty around the point estimate is low.



(a) (b)

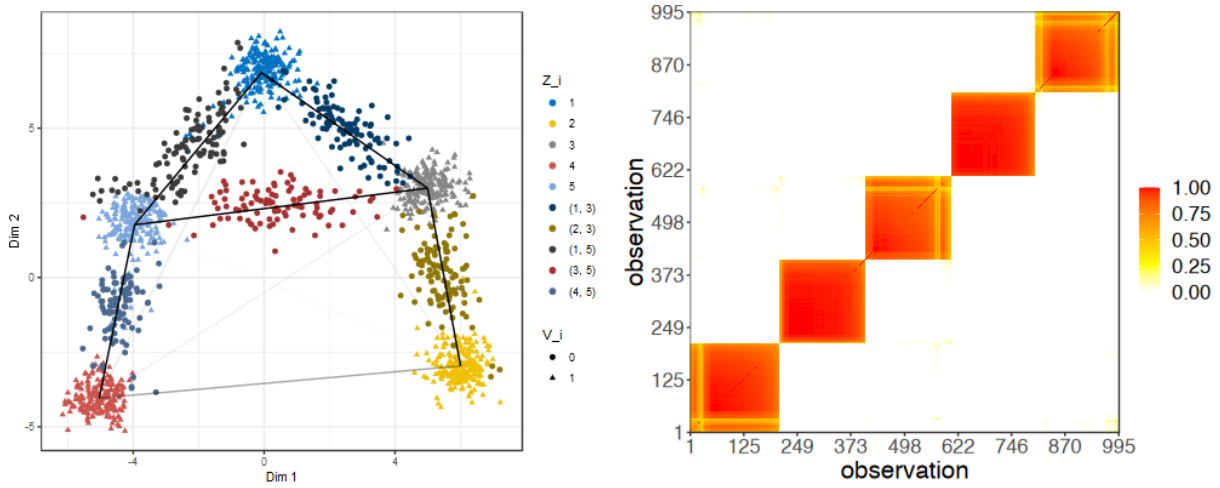
Figure S.3: Well-specified simulation scenario. Left Panel: Scatter plot of the simulated data. Observations are colored according to the estimated cluster membership. Line segments show edges of the estimated graph, with the clusters at the end of line segments being vertex clusters, and clusters along the line segments being edge clusters. The grey level of the line segments shows the estimated probability of assigning observations to the respective edge (barely varying in this case). Right panel: Posterior co-clustering probabilities for all observations assigned to vertices.

S.5.2 Misspecified Scenario

Here we consider a misspecified data-generating truth, using the same true mean vectors for five vertex clusters with cluster-specific Gaussian kernels as in the previous scenario, but inflated vertex-specific covariance matrices $\text{diag}(0.5, 0.5)$. For the edge components, we introduce two sources of misspecification. First, we center the edge components not at the midpoint of the two adjacent vertices but introduce a bias term. Instead, the edge-specific kernels are centered at $\frac{\boldsymbol{\mu}_k^* + \boldsymbol{\mu}_{k'}^*}{2}$ plus a shift of $+0.25$ in the direction of the line connecting the adjacent vertices, as well as in the perpendicular direction. Second, the observations for the edge components are generated from a uniform distribution on a rectangle centered

at the described $\boldsymbol{\mu}^*_{k,k'}$ and with the length of the side in the direction of the connecting line equal to half the length of the Euclidean distance between the adjacent vertices and the length of the other side equals 2. Under this simulation truth, the scatter plot of the simulated data still allows a meaningful definition of vertex and edge clusters, but the additional misspecification and variability with respect to the well-specified scenario make the inference with our model more challenging. The simulated $N = 1500$ observations are shown in Figure S.4a.

Figure S.4 shows that the GARP was able to recover well the simulated truth in the point estimate in this misspecified scenario. The uncertainty around the point estimate is low.



(a) (b)

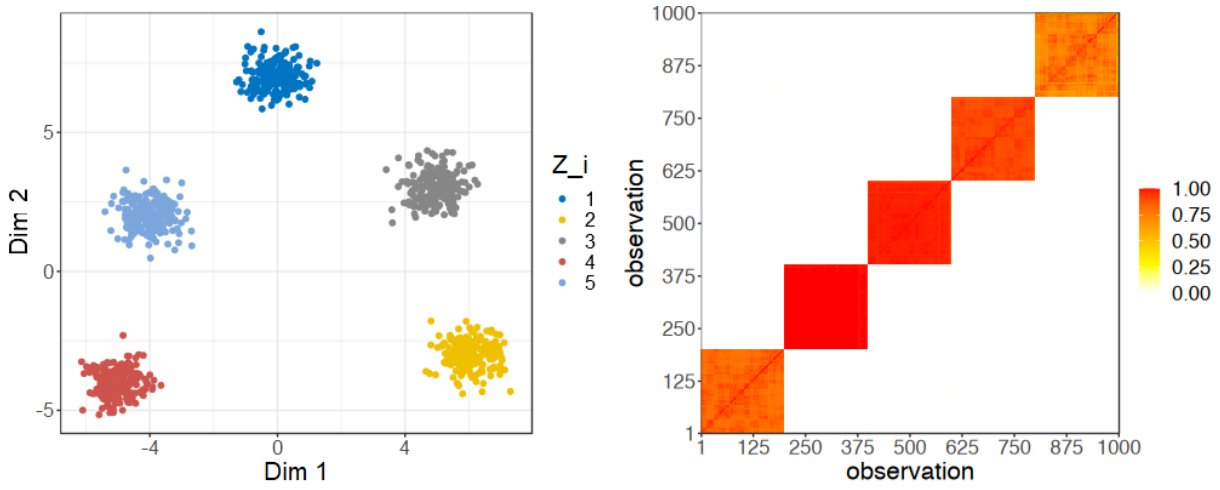
Figure S.4: Mis-specified simulation scenario. Left Panel: Scatter plot of the simulated data. Observations are colored according to estimated cluster membership. The line segments denote estimated edges, with clusters at the end of the line segments being vertex clusters, and clusters along the line segments being edge clusters. The gray shade of the line segments indicates the probability of assigning observations to the respective edge. Right panel: Posterior co-clustering probabilities for observations assigned to vertices.

S.5.3 Non-Connected Graph Scenario

Here we investigate how the model works in a scenario with no meaningful notion of the connected graph in the data. More precisely, we simulate from a mixture of five vertex clusters, exactly as in Section S.5.1, but without any edge components.

The simulated $N = 1000$ observations and inference under the GARP are shown in Figure S.5a.

Figure S.5 shows that the GARP was able to recover well the simulated truth in the point estimate also under this not-connected graph simulation truth.



(a) (b)

Figure S.5: Non-connected graph simulated data. Left Panel: Scatter plot of the simulated data. Observations are colored according to estimated cluster membership. Correctly recovering the simulation truth there are no estimated edge clusters. Right panel: Posterior co-clustering probabilities.

S.6 Proof of the Main Results

For easy reference, we provide in Table S.1 a brief statement of the various probability models used in the discussion and the results, and in the following list a brief summary of the main results. Here, Ex. 1 – 4 refer to the four examples for the EPPF from Section 3.

$G^{(N)}$	GARP model (4)
$\widetilde{G^{(N)}}$	relaxed model (7), $\widetilde{G^{(N)}}$ is proportional to $G^{(N)}$, but w/o $\mathbb{1}(E_N)$
$\widetilde{G^{(N)}_{\text{VZ}}}$	law under $\widetilde{G^{(N)}}$ on $(T_i,\, i = 1, \ldots, N)$, $T_i = \begin{cases} V_i & \text{if } V_i = 0 \\ (V_i, Z_i) & \text{if } V_i = 1 \end{cases}$
$\widetilde{G_{\text{VZ}}}$	Kolmogorov-consistent extension of $\widetilde{G^{(N)}_{\text{VZ}}}$ to $N \in \mathbb{N}$
$G^{(\infty)}$	inf exch. law that eventually matches the predictives under Ex 1 or 2 (or any MFM)
$G^{(\infty)}_N$	marginal law under $G^{(\infty)}$

Table S.1: Probability models used in the discussion and main results

For notational simplicity, we refer with $\widetilde{G_{\text{VZ}}}$ also to the marginal laws of the stochastic process $(T_i)_{i\in\mathbb{N}}$ as well as the law of M_v and $(\pi_m)_{m=1}^{M_v}$ in (S.1) since they do not depend on the dimension N. Finally, we refer with $G^{(N)}(\cdot)$ to the probability density and mass functions of random variables under the GARP model (1)–(4). More generally, given a probability measure P we denote by $P\{E\}$ the probability measure evaluated in a measurable set E

and by $P(a)$ the corresponding (when it exists) with respect to Lebesgue (i.e., pdf) or counting measure (i.e., pmf) evaluated in a point a.

Propositions 1 and S.5: Characterizations of the GARP $G^{(N)}$ as truncation of $\widetilde{G^{(N)}}$, which in turn is characterized as (i) a gCRP or, (ii) a composition of random discrete prob measures, respectively.

Proposition 2: Analytical statement of $\widetilde{G^{(N)}}\{E_N\}$ for a general Gibbs-type prior.

Theorem 1: Let $g^\infty = \lim_{N\to\infty} \widetilde{G^{(N)}}\{K_v = 1\}$ and $g_v^\infty = \lim_{n_v\to\infty} \widetilde{G^{(N)}}\{K_v = 1 \mid N_v = n_v\}$. Then

$$g^\infty = g_v^\infty = \begin{cases} 0 & \text{Ex 1 , 3, 4} \\ \gamma \in (0,1) & \text{Ex 2} \end{cases}$$

Theorem 2: $\widetilde{G_{\text{VZ}}}\{E_N \text{ eventually}\} = \begin{cases} 1 & \text{Ex 1 , 3, 4} \\ 1 - \widetilde{G_{\text{VZ}}}\{M_v = 1\} & \text{Ex 2} \end{cases}$.

Proposition 3: $\text{fEPPF}_K^{(N)}$ under the GARP model in (4).

Proposition 4: The data $\boldsymbol{y}$, the graph-aligned random partition induced by (V_i, Z_i) and the random partition Ψ_N are finitely exchangeable, but not a projection of an infinitely exchangeable process under our proposal (1)–(4).

Theorem 3 and Corollary 1: Under Ex 1, the predictive probabilities for V_i, Z_i under the GARP are eventually equal to the same under a Kolmogorov-consistent sequence $\left(G_N^{(\infty)}\right)$; statement of a Pólya urn and directing measure for $\left(G_N^{(\infty)}\right)$.

The same remains true for any MFM.

S.6.1 Proof of Proposition 1

Proof. We assume the GARP definition via the *relaxed model* in (7), (8), (9), and (10) and show that is equivalent to the definition in (4).

First, we note that in (4) the constraint $\mathbb{1}(E_N)$ can be rewritten as $\mathbb{1}(\{N_v = N\} \cup \{K_v > 1\})$. Note also that under $N_e = 0$ the second line in (4) does not arise. For notational simplicity, we naturally extend the definition of K_v and M_e by defining $K_v = M_e = 0$ if $N_v = 0$ and defining $\text{DM}_\cdot^{(0)}(\cdot) = \text{DM}_0^{(\cdot)}(\cdot) \equiv 1$.

Note also that (9) is equivalent to sample $\boldsymbol{Z}_v = (Z_i : i \in [N],\, V_i = 1)$ from

$$\widetilde{G^{(N)}}(\boldsymbol{Z_v} \mid \boldsymbol{V}) = \text{EPPF}_{K_v}^{(N_v)}(n_1, \ldots, n_{K_v} \mid \alpha, \sigma)/K_v! \tag{S.5}$$

The clustering indicators Z_i are a 1-to-1 mapping of the induced exchangeable random partition up to possible relabelings. By (conditional) exchangeability of the partition, any possible relabeling of $\boldsymbol{Z_v}$ has the same probability that is equal to the EPPF divided by the number of relabelings, i.e., $K_v!$.

Similarly, sampling from (10) is equivalent to sample $\boldsymbol{Z_e} = (Z_i : i \in [N], V_i = 0)$ from

$$\widetilde{G^{(N)}}(\boldsymbol{Z_e} \mid \boldsymbol{Z_v}) = \mathrm{DM}_{M_e}^{(N_e)}((n_{k,k'})_{k<k'_v} \mid \beta/M_e),$$

where $\mathrm{DM}_{M_e}^{(N_e)}((n_{k,k'})_{k<k'_v})$ denotes the marginal likelihood of the DM distribution for the categorical random variables, which is a function of the sufficient statistics $(n_{k,k'})_{k<k'_v}$, i.e., the ordered cardinalities of the different edges. In contrast to the EPPF and fEPPF, here some $n_{k,k'}$ can be 0, implying that there is no edge connecting the vertices k and k'.

Finally, we obtain (4) via the multiplication rule of probability, i.e.,

$$\widetilde{G^{(N)}}(\boldsymbol{V}, \boldsymbol{Z}) = \widetilde{G^{(N)}}(\boldsymbol{V})\, \widetilde{G^{(N)}}(\boldsymbol{Z_v} \mid \boldsymbol{V}) \cdot \widetilde{G^{(N)}}(\boldsymbol{Z_e} \mid \boldsymbol{V}, \boldsymbol{Z_v})$$

where $\widetilde{G^{(N)}}(\boldsymbol{V}) = p_v^{N_v}(1-p_v)^{N_e}$ by (8). □

S.6.2 Proof of Proposition 2

Proof. First, recall $E_N = \{N_v = N\} \cup \{K_v > 1\}$. That is, E_N occurs if and only if there are at least two vertex-clusters (i.e., $K_v > 1$) unless no observations are allocated to edge-clusters (i.e., $N_v = N$). Thus, by additivity of probability,

$$\begin{aligned}\widetilde{G^{(N)}}\{E_N\} &= \widetilde{G^{(N)}}\{\{N_v = N\} \cup \{K_v > 1\}\} \\ &= \widetilde{G^{(N)}}\{N_v = N\} + \widetilde{G^{(N)}}\{\{N_v \neq N\} \cap \{K_v > 1\}\},\end{aligned}$$

where $\widetilde{G^{(N)}}\{N_v = N\} = p_v^N$. In words, we decompose E_N into the union of the (disjoint) events "all clusters are vertices" and "not all observations are in vertices and there are at

least 2 vertex-clusters". The second term is further expanded by conditioning on N_v as:

$$\begin{aligned}
&\widetilde{G^{(N)}}\{\{N_v \neq N\} \cap \{K_v > 1\}\} = \\
&\widetilde{G^{(N)}}\{\{N_v \notin \{0,1,N\}\} \cap \{K_v > 1\}\} = \\
&\sum_{n_v=2}^{N-1} \widetilde{G^{(N)}}\{N_v = n_v\}\widetilde{G^{(N)}}\{K_v \neq 1 \mid N_v = n_v\} = \\
&\sum_{n_v=2}^{N-1} \binom{N}{n_v} p_v^{n_v}(1-p_v)^{n_v-1}\big[1 - \text{EPPF}_1^{(n_v)}(n_v)\big] = \\
&\sum_{n_v=2}^{N-1} \binom{N}{n_v} p_v^{n_v}(1-p_v)^{n_v-1}\big[1 - (1-\sigma)_{n_v-1} W_{n_v,1}\big],
\end{aligned}$$

where the last equality follows from the definition of the Gibbs-type priors. □

S.6.3 Proof of Theorem 1

Proof. First, note that the finite sample behavior of

$$g_{n_v} = \widetilde{G^{(N)}}\{K_{v,N} = 1 \mid N_{v,N} = n_v\} = \widetilde{G_{\text{VZ}}}\{K_{v,N} = 1 \mid N_{v,N} = n_v\} = \text{EPPF}_1^{(n_v)}(n_v)$$

is derived as a special case of the EPPF in the different examples in Section 3 of the main manuscript. From it, we can derive the large sample behavior g_{n_v} and the limit g_v^∞ reported in Table 2. Let $(x)_n = \Gamma(x+n)/\Gamma(x) = x(x+1)\cdots(x+n-1)$. To compute the rate of g_{n_v} we note that by the Stirling approximation

$$\frac{(x)_n}{n!} = \frac{\Gamma(x+n)}{\Gamma(x)n!} \asymp \frac{n^{x-1}}{\Gamma(x)} \quad \text{as } n \to \infty.$$

Note also that $(N_{v,N})_{N\in\mathbb{N}}$ is a ($\widetilde{G_{\text{VZ}}}$-almost surely) Markovian non-decreasing sequence of random integers such that

$$\frac{N_{v,N}}{N} \to p \quad \text{as } N \to \infty$$

$\widetilde{G_{\text{VZ}}}$-a.s. by the strong law of large numbers. Therefore, $N_{v,N}$ diverges $\widetilde{G_{\text{VZ}}}$-almost surely and $g_v^\infty \equiv \lim_{n_v\to\infty} \widetilde{G_{\text{VZ}}}\{K_v = 1 \mid N_{v,N} = n_v\} = \lim_{N\to\infty} \widetilde{G_{\text{VZ}}}\{K_v = 1\} = g^\infty$.

We note, as a remark, that to have g_v^∞ well defined we consider a sequence $(N = f(n_v))_{n_v\in\mathbb{N}}$ such that $f : \mathbb{N} \to \mathbb{N}$ and $f(n) \geq n$ for any $n \in \mathbb{N}$. Moreover, the hierarchical definitions of $\mathbf{V}$ and $\boldsymbol{Z_v}$ imply that $K_v = K_v(\boldsymbol{Z}_v)$ $\widetilde{G_{\text{VZ}}}$-almost surely, where $K_v = K_v(\boldsymbol{Z}_v)$ indicates a function of the N units (V_i, Z_i) that depends on $\boldsymbol{Z}$ only indirectly through the $N_{v,N}$ units allocated to vertices, i.e., $\boldsymbol{Z}_v$.

Finally, as derived in Section 4 of the main manuscript, $g_v^\infty = g^\infty = \lim_{N\to\infty} \widetilde{G_{\text{VZ}}}\{E_N^c\}$.

□

S.6.4 Proof of Theorem 2

Recall the definition of eventually. Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of events in the measurable space $(\Omega, \mathcal{F})$,

$$\{E_N \text{ eventually}\} = \liminf_N E_N = \cup_{\bar{N}=0}^{\infty} \cap_{N=\bar{N}}^{\infty} E_N.$$

In words, it is the set of $\omega \in \Omega$ such that there exists an integer $\bar{N}(\omega)$ such that for any integer $N \geq \bar{N}(\omega)$, $\omega \in E_N$.

Proof. **Case with a M_v-dimensional symmetric Dirichlet (where $M_v > 1$) or with a DP or with a PYP in** (4)**.**

First, since $K_{v,N}, N_{v,N}$ are functions of $T_{1:N}$ (that is of $(\boldsymbol{V}_{1:N}, \boldsymbol{Z}_{v,N})$) only, $\widetilde{G^{(N)}}(K_{v,N}, N_{v,N}) = \widetilde{G_{\text{VZ}}}(K_{v,N}, N_{v,N})$ for any $N \in \mathbb{N}$.

Note that under $\widetilde{G_{\text{VZ}}}$, $(K_{v,N})_N$ is an a.s. non-decreasing Markovian sequence of positive integers such that for any natural $N > 1$, $\widetilde{G_{\text{VZ}}}\{K_{v,N} > 1\} > 0$ and it can be computed from (8)-(9).

Moreover, by Kingman's representation theorem (see Kingman, 1978 and Theorem 14.7 in Ghosal and van der Vaart, 2017) the random partition can be characterized as arising from the ties obtained by sampling from a unique discrete probability measure $P_v = \sum_{m=1}^{M_v} \pi_m \delta_{\tilde{\boldsymbol{\theta}}}$ (we know that is M_v-symmetric Dirichlet or a DP or a PYP distributed) and the frequency of the kth largest partition block converges almost surely to kth largest random weight in $(\pi_m)_{m=1}^{M_v}$ for any $k \in 1, \ldots, M_v$. Therefore, together with the assumption $M_v \geq 2$, it implies that

$$\widetilde{G_{\text{VZ}}}\{\{K_{v,N} > 1\} \text{ eventually } w.r.t.\ N\} = 1.$$

To conclude the proof of (14), note that

$$E_N = \{N_{v,N} = N\} \cup \{K_{v,N} > 1\} \supset \{K_{v,N} > 1\}.$$

Thus, we have shown that $\widetilde{G_{\text{VZ}}}\{E_N \text{ eventually}\} = 1$.

To prove (15), first recall that for any $N \in \mathbb{N}$, $G^{(N)}$ and $\widetilde{G^{(N)}}$ denote the probability mass function of $(V_i, Z_i)_{i=1}^N$ under the GARP and the relaxed model, respectively. Next, for any $N, k \in \mathbb{N}$ and any set of possible points $a_k = (\mathbf{v}_{1:N+k}, \mathbf{z}_{1:N+k})$, by definition of conditional probability we have

$$\widetilde{G^{(N+k)}}(a_k \mid \mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}) = \frac{\widetilde{G^{(N+k)}}(a_k)}{\widetilde{G^{(N+k)}}\{\mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}\}}, \tag{S.6}$$

where, by additivity of probability,

$$\widetilde{G^{(N+k)}}\{\mathbf{V}_{1:N}=\mathbf{v}_{1:N},\mathbf{Z}_{v,N}=\mathbf{z}_{v,N}\}=\sum_{\left\{(v_i',z_i')_{i=1}^{N+k}:\mathbf{v}'_{1:N}=\mathbf{v}_{1:N},\,\mathbf{z}'_{v,N}=\mathbf{z}_{v,N}\right\}}\widetilde{G^{(N+k)}}((v_i',z_i')_{i=1}^{N+k}).$$

Moreover, for any $k,N\in\mathbb{N}$ and any possible points $a_k=\mathbf{v}_{1:N+k},\mathbf{z}_{1:N+k}$ such that $\{\boldsymbol{V}_{1:N}=\mathbf{v}_{1:N+k},\boldsymbol{Z}_{1:N}=\mathbf{z}_{1:N}\}$ entails that $\{K_{v,N}>1\}$ holds (and thus $\mathbb{1}(E_N)=1$) we have

$$G^{(N+k)}\big(a_k\mid\mathbf{V}_{1:N}=\mathbf{v}_{1:N},\mathbf{Z}_{v,N}=\mathbf{z}_{v,N}\big)=\frac{\widetilde{G^{(N+k)}}\big(a_k\big)}{\widetilde{G^{(N+k)}}\{\mathbf{V}_{1:N}=\mathbf{v}_{1:N},\mathbf{Z}_{v,N}=\mathbf{z}_{v,N}\}},$$

by (7) and definition of conditional probability.

To conclude the proof of (15) we note that, by (14), there exists a set $\mathcal{T}$ of sequences $(t_i)_{i=1}^{\infty}$ that are possible realizations of $(T_i)_{i=1}^{\infty}$ such that $\widetilde{G_{\text{VZ}}}\{\mathcal{T}\}=1$ and such that for any sequence $t=(t_i)_{i=1}^{\infty}\in\mathcal{T}$ there exists a $\bar{N}(t)\in\mathbb{N}$ such that $\{K_{v,N}(t)>1\}$ holds for any $N\geq\bar{N}(t)$. Therefore, for any $N\geq\bar{N}(t)$

$$G^{(N+k)}\big(a_k\mid\mathbf{V}_{1:N}=\mathbf{v}_{1:N},\mathbf{Z}_{v,N}=\mathbf{z}_{v,N}\big)=\widetilde{G^{(N+k)}}(a_k\mid\mathbf{V}_{1:N}=\mathbf{v}_{1:N},\mathbf{Z}_{v,N}=\mathbf{z}_{v,N}),$$

where $t_i=v_i$ if $v_i=0$ and $t_i=(v_i,z_i)$ if $v_i=1$. Thus we proved (15).

Case with a Gnedin process in (4).
Similarly to the previous case, note that under $\widetilde{G_{\text{VZ}}}$, $(K_{v,N})_N$ is an a.s. non-decreasing Markovian sequence of positive integers. Moreover, by Kingman representation theorem and the fact that $\widetilde{G_{\text{VZ}}}\{M_v<\infty\}=1$ we have that

$$\widetilde{G_{\text{VZ}}}\{\{K_{v,N}=M_v\}\text{ eventually }w.r.t.\ N\}=1.$$

Indeed, the random partition can be thought of as arising from the ties obtained by sampling from a unique discrete probability measure $P_v=\sum_{m=1}^{M_v}\pi_m\delta_{\tilde{\boldsymbol{\theta}}}$ (here distributed as a Gnedin process) and the frequency of the kth largest partition block converges almost surely to kth largest random weight in $(\pi_m)_{m=1}^{M_v}$ for any $k\in 1,\dots,M_v$.

Note that $\{K_{v,N}=M_v\}\subset\{K_{v,N}>1\}\cup\{M_v=1\}\subset E_N\cup\{M_v=1\}$, thus

$$\widetilde{G_{\text{VZ}}}\{\{K_{v,N}>1\}\cup\{M_v=1\}\text{ eventually }w.r.t.\ N\}=1$$

and

$$\widetilde{G_{\text{VZ}}}\big\{E_N\cup\{M_v=1\}\text{ eventually }w.r.t.\ N\big\}=1.\tag{S.7}$$

To conclude the proof we need to show that, for any $k\in\mathbb{N}$ and any possible set of points

$a_k = (\mathbf{v}_{1:N+k}, \mathbf{z}_{1:N+k})$

$$\widetilde{G}_{VZ}\left\{\left\{G^{(N+k)}(a_k \mid \mathbf{V}_{1:N}, \mathbf{Z}_{v,N}) = \widetilde{G^{(N+k)}}(a_k \mid \mathbf{V}_{1:N}, \mathbf{Z}_{v,N})\right\} \cup \{M_v = 1\} \text{ eventually}\right\} = 1. \tag{S.8}$$

To prove (S.8) we note that, by (S.7), there exists a set $\mathcal{T}$ of sequences $(t_i)_{i=1}^{\infty}$ that are possible realizations of $(T_i)_{i=1}^{\infty}$ such that $\widetilde{G_{\text{VZ}}}\{\mathcal{T}\} = 1$ and such that for any sequence $t = (t_i)_{i=1}^{\infty} \in \mathcal{T}$ there exists a $\bar{N}(t) \in \mathbb{N}$ such that $\{K_{v,N}(t) > 1\} \cup \{M_v(t) = 1\}$ (and thus E_N) holds for any $N \geq \bar{N}(t)$. Therefore, by (7) and definition of conditional probability, for any $N \geq \bar{N}(t)$

$$G^{(N+k)}\big(a_k \mid \mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}\big) = \widetilde{G^{(N+k)}}(a_k \mid \mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}),$$

where $t_i = v_i$ if $v_i = 0$ and $t_i = (v_i, z_i)$ if $v_i = 1$. □

S.6.5 Proof of Proposition 3

The fEPPF in Proposition 3 is computed via marginalization of the pmf of the GARP in (4) over all the quantities that are compatible with the cardinalities $\{c_1, \ldots, c_{K_v}\}$ of Ψ_N.

We state a more complete version of Proposition 3, now including a statement of the range of the three sums that appear in

$$\begin{aligned}\text{fEPPF}^{(N)}_{K_N}(|C_1|, \ldots, |C_{K_N}|) \propto \\ \sum_{N_v}\Bigg\{\binom{N}{N_v} p_v^{N_v}(1-p_v)^{N-N_v} \sum_{K_v}\Bigg[\binom{M_e}{K_N - K_v} \\ \sum_{(n_{k_1},\ldots,n_{K_v})} \text{EPPF}^{(N_v)}_{K_v}(n_{k_1}, \ldots, n_{K_v})\, \text{DM}^{(N-N_v)}_{M_e}((n_{k,k_{v'}})_{k<k'_v})\Bigg]\Bigg\}\end{aligned}$$

The first sum runs over $N_v \in [N]$ with the restriction that $N_v = N$ if $K_N \leq 2$. The second sum runs over $K_v \in [K_N]$ with the restrictions that

1. $K_v \geq 2$ if $K_N \geq 2$;
2. $K_v < K$ if $N_v \neq N$;
3. $K_v = K$ if $N_v = N$;
4. $K_N \leq N_v + \min\{M_e, N - N_v\}$, keeping in mind that $M_e \coloneqq \frac{K_v(K_v-1)}{2}$.

Finally, the last sum runs over $(n_1, \ldots, n_{K_v})$ where $\sum_{k=1}^{K_v} n_k = N_v$ and $n_1, \ldots, n_{K_v}$ are distinct elements of $\{c_1, \ldots c_{K_N}\}$ ordered, e.g., by cardinalities. And the non-zero edge-cluster sizes $n_{k,k'}$ are the remaining (ordered) elements of $(c_1, \ldots c_{K_N})$ that are not matched with vertex-cluster sizes n_k.

S.6.6 Proof of Proposition 4

Proof. **Finite exchangeability.**

First note that $\boldsymbol{Z_v} = (Z_{v,i})_{i=1}^{N_v} \coloneqq (Z_i : i \in [N],\, V_i = 1)$ identifies arbitrarily labeled vertex-clusters (e.g., in order of appearance). Hence, formally the vector $\boldsymbol{Z}_v$ and its relabeling are regarded as distinct objects, even though they identify the same vertex-partition.

Moreover, if the edge-clusters are relabelled according to the relabeling of the vertex-clusters this identifies the exact same graph-aligned random partition.

For instance, $(Z_1 = 1, Z_2 = 2, Z_3 = 5, Z_4 = (1, 2))$ entails the same graph-aligned partition as $(Z_1 = 3, Z_2 = 2, Z_3 = 1, Z_4 = (2, 3))$, but a different one than $(Z_1 = 3, Z_2 = 2, Z_3 = 3, Z_4 = (1, 2))$. A relabeling of Z_i which preserves the same graph-aligned random partition does not modify the likelihood distribution $G^{(N)}(\boldsymbol{y} \mid \boldsymbol{V}, \boldsymbol{Z})$ in (1), which is invariant under such a relabeling.

By construction, the graph-aligned random partition (4) induced by (V_i, Z_i) is exchangeable, i.e., the joint law is invariant to permutation of the labels i. Note that we cannot state the same argument directly in terms of the pmf of (V_i, Z_i) since we have an arbitrary order of $\boldsymbol{Z}_v$, i.e., the order of arrival (irrelevant for the graph-aligned random partition) that gives probability zero to permutations of i's that entails a non-increasing sequence of $\boldsymbol{Z}_v$.

Since the likelihood of the sample (1) can be defined as a function of the graph-aligned random partition, we immediately obtain the exchangeability of the sample $(\boldsymbol{y})_{i=1}^{N}$.

Finally, since the random partition Ψ_N can be seen as the marginalization of the graph-aligned random partition, we also have finite exchangeability of Ψ_N as also shown via the fEPPF (3).

Lack of projectivity.

To prove that *infinity exchangeability* does not hold we show a simple counterexample where projectivity does not hold.

We first show the lack of projectivity for the graph-aligned random partition $G^{(N)}(\boldsymbol{V}, \boldsymbol{Z})$. It suffices to note that for a sample of size $N = 1$ the probability of assigning an observation to a vertex is 1, i.e., $G^{(1)}\{V_1 = 1\} = 1$, while it is strictly smaller than 1 for $N = 3$, since, by (4),

$$G^{(3)}\{V_1 = 0\} = G^{(3)}\{V_1 = 0, Z_1 = (1, 2), V_2 = 1, Z_2 = 1, V_3 = 1, Z_3 = 2\} > 0.$$

Next, we show the lack of projectivity for $\boldsymbol{y}$. The last argument also implies that in a sample of size $N = 1$ the marginal density of the observations $\boldsymbol{y}_1$ can be rewritten as

$$\int \mathrm{N}(\boldsymbol{y} \mid \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)\mathrm{dNIW}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^* \mid \boldsymbol{\mu}_0, \lambda_0, \kappa_0, \boldsymbol{\Sigma}_0) \tag{S.9}$$

while under $N = 3$ it is a mixture of (S.9) and an additional term corresponding to an allocation as an edge:

$$\int \mathrm{N}(\boldsymbol{y} \mid \boldsymbol{\mu}^*_{1,2}, \boldsymbol{\Sigma}^*_{1,2}) \mathrm{d}G^{(3)}(\boldsymbol{\mu}^*_{1,2}, \boldsymbol{\Sigma}^*_{1,2}),$$

with $G^{(3)}(\boldsymbol{\mu}^*_{1,2}, \boldsymbol{\Sigma}^*_{1,2})$ characterized by $\boldsymbol{\mu}^*_{1,2} = (\boldsymbol{\mu}^*_1 + \boldsymbol{\mu}^*_2)/2$ and $\boldsymbol{\Sigma}^*_{1,2} = f(\boldsymbol{\mu}^*_1, \boldsymbol{\mu}_2)$, where $\boldsymbol{\mu}^*_1$ and $\boldsymbol{\mu}^*_2$ are independent draws of a generalized Student-T distribution. This shows that $\boldsymbol{y}_i$ is not infinitely exchangeable.

Finally, we consider the random partition Ψ_N. Note that the probability of observations $i = 1, 2$ being clustered together in a sample of size 2 (i.e., of a partition with a single cluster), is equal to

$$\begin{aligned} G^{(2)}\{\Psi_2 = \{1,2\}\} &= \mathrm{fEPPF}^{(2)}_1(2) = \mathrm{EPPF}^{(2)}_1(2) > \\ &> G^{(3)}\{\Psi_3 : Z_1 = Z_2\} = \mathrm{EPPF}^{(2)}_1(2)\, G^{(3)}\{V_1 = V_2 = 1\}. \end{aligned}$$

Thus, in the last expression, the first factor is the probability of having the observations with labels $i = 1, 2$ in the same cluster given that they are in vertex-clusters, and the second factor is the probability of those two observations being assigned to vertex clusters. Note that, in the case of $N = 2$ the probability of the two observations to be assigned in vertex-cluster is 1. □

S.6.7 Proof of Theorem 3 and Corollary 1

Theorem 3 in the main manuscript shows that in some cases the predictive distributions of the GARP model eventually (i.e., for a large enough sample size N) can be characterized as a projection of the predictive distributions of a limiting infinitely exchangeable model, thus where projectivity holds.

Proof. **Proof of Theorem 3 (M_v-dimensional symmetric Dirichlet)**
(Case 1: $M_v = 1$)
For any $N \in \mathbb{N}$ our proposal degenerates to a single Gaussian model because $G^{(N)}$-a.s. all the observations are clustered together in a single vertex. In such a case it is immediate to check that we have projectivity and (18), (19) and (20) hold. However, this is clearly an uninteresting case from a modeling perspective.
(Case 2: $M_v > 1$)
First, recall that

$$\widetilde{G_{\mathrm{VZ}}}\left\{ \lim_{N\to\infty} \frac{N_{v,N}}{N} \to p_v \right\} = 1$$

by the strong law of large numbers.

Recall also that under $\widetilde{G_{\mathrm{VZ}}}$, $(K_{v,N})_N$ is an a.s. non-decreasing Markovian sequence of positive integers such that for any $N \in \mathbb{N}$, $K_{v,N} \leq M_v$ and $\widetilde{G_{\mathrm{VZ}}}\big\{K_{v,N} = \min(N, M_v)\big\} > 0$

and it can be computed from (8)-(9).

Moreover, by Kingman's representation theorem (see Kingman, 1978 and Theorem 14.7 in Ghosal and van der Vaart, 2017) the random partition can be thought of as arising from the ties obtained by sampling from a unique discrete probability measure $P_v = \sum_{m=1}^{M_v} \pi_m \delta_{\tilde{\boldsymbol{\theta}}}$ (we know that is M_v-symmetric Dirichlet distributed) and the frequency of the kth largest partition block converges almost surely to kth largest random weight in $(\pi_m)_{m=1}^{M_v}$ for any $k \in \{1, \ldots, M_v\}$. Therefore, together with the assumption that M_v is finite $\widetilde{G_{\text{VZ}}}\{\lim_{N\to\infty} K_{v,N} = M_v\} = 1$. Thus, since $K_{v,N}$ are random integers,

$$\widetilde{G_{\text{VZ}}}\left\{\{K_{v,N} = M_v\} \text{ eventually w.r.t. } N\right\} = 1. \tag{S.10}$$

Note also that

$$\{K_{v,N} = M_v\} \subset E_N.$$

Thus, for any $N, k \in \mathbb{N}$ and $a_k = (v_i, z_i)_{i=1}^{N+k}$ such that $\{\mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}\}$ entails that $\{K_{v,N} = M_v\}$ holds (and so $\mathbb{1}(E_N) = 1$) we have

$$\begin{aligned}
&G^{(N+k)}\big((v_i, z_i)_{i=1}^{N+k} \mid \mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}\big) = \\
&\frac{\widetilde{G^{(N+k)}}\big((v_i, z_i)_{i=1}^{N+k}\big)}{\widetilde{G^{(N+k)}}\{\mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}\}} = \\
&\widetilde{G^{(N+k)}}\big((v_i, z_i)_{i=1}^{N+k} \mid \mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}\big),
\end{aligned}$$

by definition of conditional probability and (7).

Note that, for any $N \in \mathbb{N}$, $K_{v,N} = M_v$ entails that $K_{v,N+k} = M_v$ and $M_{e,N+k} = M_e^+ \coloneqq \frac{M_v(M_v-1)}{2}$ for any $k = 0, 1, \ldots$. Therefore, by definition of $\widetilde{G^{(N)}}$ and the fact that $\{\mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}\}$ entails that $\{K_{v,N} = M_v\}$ and $\{M_{e,N} = M_e^+\}$ hold, for any $k \in \mathbb{N}$ we have

$$\begin{aligned}
&\widetilde{G^{(N+k)}}\big((v_i, z_i)_{i=1}^{N+k} \mid \mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}\big) = \\
&\frac{G_{N+k}^{(\infty)}\big((v_i, z_i)_{i=1}^{N+k}\big)}{G_{N+k}^{(\infty)}\{\mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}\}} = \\
&G_{N+k}^{(\infty)}\big((v_i, z_i)_{i=1}^{N+k} \mid \mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}\big),
\end{aligned}$$

where $G_{N+k}^{(\infty)}$ refers to the pmf of $\mathbf{V}_{1:N+k}, \mathbf{Z}_{1:N+k}$ defined in (20). We now explicitly call such law $G_{N+k}^{(\infty)}$ (i.e., with the subscript) to stress the dimension to show that $(G_N^{(\infty)})_{N\in\mathbb{N}}$ are indeed Kolmogorov consistent and can be seen as the projection of the law of a stochastic process $G^{(\infty)}$.

To conclude the proof of (17) recall that by (S.10), there exists a set $\mathcal{T}$ of sequences $(t_i)_{i=1}^{\infty}$ that are possible realizations of $(T_i)_{i=1}^{\infty}$ such that $\widetilde{G_{\text{VZ}}}\{\mathcal{T}\} = 1$ and such that for any

sequence $t = (t_i)_{i=1}^{\infty} \in \mathcal{T}$ there exists a $\bar{N}(t) \in \mathbb{N}$ such that $\{K_{v,N}(t) = M_v\}$ (and thus also $\{M_{e,N}(t) = M_e^+\}$) holds for any $N \geq \bar{N}(t)$. Therefore, for any $N \geq \bar{N}(t)$

$$G^{(N+k)}\big(a_k \mid \mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}\big) = G^{(\infty)}_{N+k}(a_k \mid \mathbf{V}_{1:N} = \mathbf{v}_{1:N}, \mathbf{Z}_{v,N} = \mathbf{z}_{v,N}),$$

where $t_i = v_i$ if $v_i = 0$ and $t_i = (v_i, z_i)$ if $v_i = 1$.

To check the projectivity of $(G_N^{(\infty)})_N$ we note that for any $N \in \mathbb{N}$ and possible values $(v_i, z_i)_{i\in[N]}$

$$\begin{aligned} G_N^{(\infty)}((v_i, z_i)_{i\in[N]}) &= p_v^{N_v} \mathrm{EPPF}^{(N_v)}_{M_v}(n_1, \ldots, n_{K_v} \mid \alpha, \sigma)/K_v! \\ &\quad (1-p_v)^{N_e} \mathrm{DM}^{(N_e)}_{M_v(M_v-1)/2}((n_{k,k'})_{k<k'} \mid \beta/M_e) \\ &= \sum_{v_{N+1}, z_{N+1}} G^{(\infty)}_{N+1}((v_i, z_i)_{i\in[N]}) = G^{(\infty)}((V, Z)_{i\in[N]}). \end{aligned}$$

The second and third equalities hold by projectivity of the EPPF and DM (where the sum is over all possible values of v_{N+1}, z_{N+1}). We denote by $G^{(\infty)}$ the infinite-dimensional GARP defined via such Kolmogorov consistent finite-dimensional distributions.

From $G^{(\infty)}$ (and its Kolmogorov consistent finite-dimensional) we derive the urn schemes in (18) via the definition of conditional probability. The ratio boils down to (18) thanks to the product form of the EPPF and of the DM.

Finally, note that via the characterization of the EPPF and DM in terms of discrete random probabilities (see e.g., Section S.2), the induced law on $(\boldsymbol{\theta}_i)_{i=1}^N$ can thus be characterized by first sampling $V_i \overset{\text{iid}}{\sim} \mathrm{Bern}(p_v)$ and $\boldsymbol{\theta}_i \mid P_v, V_i = 1 \overset{\text{ind}}{\sim} P_v \coloneqq \sum_{m=1}^M \pi_m \delta_{\tilde{\boldsymbol{\theta}}_m}$ and $\boldsymbol{\theta}_i \mid P_e, V_i = 0 \overset{\text{ind}}{\sim} P_e \coloneqq \sum_{k<k'<M} \pi_{k,k'} \delta_{\tilde{\boldsymbol{\theta}}_{k,k'}}$. Thus we derive (19) marginalizing with respect to $\mathbf{V}$ and by the uniqueness of the directing measure.

Proof of corollary 1

First, we write explicitly the statement of Corollary 1.

Corollary S.2 (Corollary 1 of the main manuscript)**.** *Under the* GARP *with a Gnedin process (Example 2) in* (4) *there exists a finite random sample size* $\bar{N}$ *such that for any* $N > \bar{N}$ *the predictive distributions under the proposed GARP model given* M_v *are* $\widetilde{G_{VZ}}$*-a.s. equal to the predictive distributions given* M_v *under a Kolmogorov consistent* $G^{(\infty)}$*, i.e., for any possible sequence of sets of points* $(a_k)_{k\in\mathbb{N}}$*, with* $a_k = \mathbf{v}_{1:N+k}, \mathbf{z}_{1:N+k})$

$$\widetilde{G_{VZ}}\Big\{\big\{G^{(\infty)}_{N+k}(a_k \mid \mathbf{V}_{1:N}, \mathbf{Z}_{1:N}, M_v) = G^{(N+k)}(a_k \mid \mathbf{V}_{1:N}, \mathbf{Z}_{1:N}, M_v)\ \forall\, k\big\}\ \textit{eventually}\Big\} = 1.$$

Moreover, $G^{(\infty)}(\cdot \mid \mathbf{V}_{1:N}, \mathbf{Z}_{1:N}, M_v)$ *can be characterized by the urn scheme in* (18) *and* $G^{(\infty)}(\cdot \mid M_v)$ *by the pmf* (20) *and by an exchangeable sequence with directing measure being the law of* $P \mid M_v$ *as in* (19)*. Finally,* $G^{(\infty)}(M_v = m) = \widetilde{G_{VZ}}(M_v = m) = \frac{\gamma(1-\gamma)_{m-1}}{m!}$.

Note that $\widetilde{G_{\text{VZ}}}$-a.s. $M_v \in \mathbb{N}$ and that for any realization of $M_v = m \in \mathbb{N}$ we are back to the finite symmetric Dirichlet GARP and thus the result follows from Theorem 3. □

S.7 Software, Runtime, etc.

The results reported in this article are based on 10,000 MCMC iterations with the initial 5,000 iterations discarded as burn-in. The remaining samples were further thinned by an interval 2. We programmed everything in `R`. The analyses are performed with a Lenovo ThinkStation P330 with 16Gb RAM (Windows 10), using a `R` version 4.2.3. The MCMC algorithm takes 29.8 minutes.

References

Binder, D. A. (1978). Bayesian cluster analysis. *Biometrika*, **65**, 31–38.

Dahl, D. B., Johnson, D. J., and Müller, P. (2022a). Salso: search algorithms and loss functions for Bayesian clustering. *R package version 0.3.29.*

Dahl, D. B., Johnson, D. J., and Müller, P. (2022b). Search algorithms and loss functions for Bayesian clustering. *J. Comput. Graph. Stat.,* **31**, 1189–1201.

Franzolini, B. and Rebaudo, G. (2024). Entropy regularization in probabilistic clustering. *Stat. Methods Appt.,* **in press.**

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference.* Cambridge Univ. Press.

Kingman, J. F. C. (1978). The representation of partition structures. *J. London Math. Soc.,* **18**, 374-380.

Meilă, M. (2007). Comparing clusterings–an information based distance. *J. Multivar. Anal.,* **98**, 873–895.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Stat.,* **9**, 249–265.

R Core Team (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria.

Robert, C. P., and Roberts, G. (2021). Rao–Blackwellisation in the Markov Chain Monte Carlo Era. *Int. Stat. Rev.*, **89**, 237–249.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *J. Am. Stat. Assoc.,* **101**, 1566–1581.

Wade, S. and Ghahramani, Z. (2018). Bayesian cluster analysis: point estimation and credible balls. *Bayesian Anal.,* **13**, 559–626.