Bayesian Learning of Graph Substructures[*]

Willem van den Boom[†], Maria De Iorio[‡,§,¶] and Alexandros Beskos[‖,**]

Abstract. Graphical models provide a powerful methodology for learning the conditional independence structure in multivariate data. Inference is often focused on estimating individual edges in the latent graph. Nonetheless, there is increasing interest in inferring more complex structures, such as communities, for multiple reasons, including more effective information retrieval and better interpretability. Stochastic blockmodels offer a powerful tool to detect such structure in a network. We thus propose to exploit advances in random graph theory and embed them within the graphical models framework. A consequence of this approach is the propagation of the uncertainty in graph estimation to large-scale structure learning. We consider Bayesian nonparametric stochastic blockmodels as priors on the graph. We extend such models to consider clique-based blocks and to multiple graph settings introducing a novel prior process based on a Dependent Dirichlet process. Moreover, we devise a tailored computation strategy of Bayes factors for block structure based on the Savage-Dickey ratio to test for presence of larger structure in a graph. We demonstrate our approach in simulations as well as on real data applications in finance and transcriptomics.

Keywords: Bayesian nonparametrics, degree-corrected stochastic blockmodels, Dependent Dirichlet process, Gaussian graphical models, multiple graphical models, multivariate data analysis.

1 Introduction

Graphical models provide a flexible tool to describe the conditional independence structure in multivariate data: the nodes of the graph represent variables and the edges amongst them define conditional dependence (Lauritzen, 1996). Most inferential approaches focus on estimation of individual edges rather than on identification of informative structure in a graph on a larger scale. This is despite the fact that such large-scale structure is often present and of interest in multivariate data (Ravasz et al., 2002; Yook et al., 2004). Moreover, estimation of a single edge is (often extremely) sensitive to the number of observations as well as to the presence of specific nodes in the graph. We therefore propose graphical models that also enable learning of large-scale structure. These models build on extensive work in random graphs and networks (Newman, 2011;

[†]Yong Loo Lin School of Medicine, National University of Singapore, vandenboom@nus.edu.sg
[‡]Yong Loo Lin School of Medicine, National University of Singapore, mdi@nus.edu.sg
[§]Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research
[¶]Department of Statistical Science, University College London
[‖]Department of Statistical Science, University College London, a.beskos@ucl.ac.uk
[**]Alan Turing Institute, UK

Fortunato and Hric, 2016; Lee and Wilkinson, 2019) such as the stochastic blockmodel (Holland et al., 1983).

It is important to stress the distinction between random graph theory and graphical models as two parallel, large research areas with only limited interplay such as the work by Bornn and Caron (2011) and Peixoto (2019). Within the first field, models for random graphs have evolved substantially from initial approaches such as the Erdős–Rényi model (Erdős and Rényi, 1959) to methods for large-scale structure. They usually involve the description of network formation/evolution. See Fienberg (2012) and Barabási (2016) for an overview. Such developments contrast with the literature on graphical models (Friedman et al., 2007; Armstrong et al., 2009; Zhou et al., 2011; Maathuis et al., 2019; Ni et al., 2022) that aims to infer a graph from multivariate data. In this context, focus of inference is usually to determine the presence of an edge between two nodes whereas modelling of large-scale graph structures is often neglected (Bornn and Caron, 2011).

The rationale underpinning our work derives from the following consideration. In the random graph literature, there is major interest on large-scale structures as they often arise in applications. A common example is provided by scale-free networks which imply a hub structure (Yook et al., 2004). More recently, Newman (2011) advocates for more complex structures such as modularity. This consideration motivates the need to investigate such components also when inferring graphs from multivariate data. On the other hand, the Bayesian graphical model literature commonly focuses on single edges, and specification of a prior on graph space is achieved assuming the same probability of inclusion for each edge with all edges being independent. This prior corresponds to the Erdős–Rényi model.

Focusing on single edges can be restrictive in many applications, often preventing detection of important data features. For instance, assume we are interested in estimating a graph from gene expression data. It could be of biological interest (e.g. disease aetiology) to group genes in co-expression modules (i.e. block, larger structure) (Yook et al., 2004). Similarly in metabolomics, it is of interest to identify metabolites that are involved in the same biochemical reaction/pathway (Ravasz et al., 2002). Social networks provide another area where such graph substructures are relevant. For instance, we might estimate a graph from voting records of members of parliament with the goal of identifying political factions.

The increasing interest in estimating large structures in multivariate data is reflected in recent work. For example, Zhang (2018) first estimates a graph from the data and, then, identifies large-scale structure using random graph methods. Such an approach is suboptimal, for instance because it does not propagate the uncertainty from network estimation to the estimation of large-scale structure. In the machine learning literature, methods for identification of graph substructure can be found in Marlin et al. (2009).

Our work is positioned in this new line of research. Exploiting advances from random graph theory, we propose graphical models able to accommodate single-edge as well as block structure. The benefits of joint graph and structure recovery compared to a two-step approach are multiple: (i) if present, large-scale structure can guide graph estimation; (ii) ad hoc specification of a graph estimator (e.g. through an edge inclusion

probability threshold) is not required; (iii) data-driven detection of structure or lack thereof; (iv) uncertainty in graph estimation propagates to large-scale structure learning; (v) extension to complex set-ups (e.g. different biological conditions) is in principle straightforward which leads to; (vi) effective use of information as the developed framework allows combining data from multiple sources in a principled way. We consider both single and multiple graph scenarios as well as different blockmodels, namely the usual stochastic blockmodel and also one where blocks are cliques. Here we focus on blockmodels because of their popularity, but the setting is general and other priors could be employed.

One of our contributions is an algorithm (derived as by-product of the MCMC) to compute Bayes factors to test for the presence of block structure, which is equivalent to the presence of clusters in a nonparametric partition model. Our approach, based on the Savage-Dickey ratio (Dickey, 1971), offers computational advantages over existing methods (Basu and Chib, 2003; Legramanti et al., 2022a). To define blocks in multiple graphs, we introduce a novel Bayesian nonparametric prior. Specifically, we propose a Dependent Dirichlet process that does not enforce exchangeability within groups as in previous approaches (e.g. MacEachern, 1999; De Iorio et al., 2004; Müller et al., 2004; Camerlenghi et al., 2019; Quintana et al., 2022).

The paper is structured as follows. Sections 1.1 and 1.2 review related work on blockmodels and graphical models, respectively. Section 2 introduces Gaussian graphical models (GGMs, Dempster, 1972). In Section 3, we propose various priors on graphs that allow recovery of large-scale structure. Section 4 introduces Bayes factors for testing for block structure. We demonstrate the proposed approach in simulation studies in Section 5 and on real data applications in Section 6. We conclude the paper in Section 7. The paper is accompanied by Supplementary Material which contains further details on the methods and results.

1.1 Stochastic Blockmodels

Arguably, the most widely used model for large-scale structure in graphs is the blockmodel (Fienberg, 2012) which is therefore our starting point. A stochastic blockmodel (Holland et al., 1983) consists of a partition of the set of nodes into blocks or communities, where we use both terms interchangeably. Then, nodes in the same block are more likely to be connected than nodes from different blocks. Thus, the structure of interest is the clustering of nodes and the connectivity within and between these clusters. Introducing block structure in graph estimation allows highlighting macro-organisation (instead of focusing on single edges) and important hubs/connectivity clusters, which ultimately will aid interpretation of the results and hypothesis generation.

To this end, the key modelling strategy that we adopt is to employ tools from the Bayesian nonparametric literature for estimation of block structure in a graph, as such a strategy provides uncertainty propagation across the full graphical model and data-driven determination of blocks' number and membership. We limit this literature review of blockmodels to Bayesian nonparametric approaches though many others exist (e.g. Fortunato and Hric, 2016; Abbe, 2018; Funke and Becker, 2019; Lee and Wilkinson,

2019; Gao and Ma, 2021). See Schmidt and Mørup (2013) for an introduction to Bayesian nonparametric modelling of graphs including blockmodels. Kemp et al. (2006) introduce a blockmodel where the prior on the partition of nodes is a Chinese restaurant process (CRP, Pitman, 2006) which closely relates to the Dirichlet process (DP, Ferguson, 1973). Geng et al. (2018), Gao et al. (2020) and Jiang and Tokdar (2021) place a mixture prior with random number of components (Miller and Harrison, 2017; Argiento and De Iorio, 2022) on the partition: Geng et al. (2018) obtain posterior consistency results for the number of blocks, and Jiang and Tokdar (2021) do so for the partition and the edge probabilities. Gao et al. (2020) provide posterior concentration rates for the edge probabilities and show that their posterior mean achieves the minimax rate. Legramanti et al. (2022b) employ Gibbs-type partition priors which generalise both the CRP and the mixture with random number of components.

In general, these approaches require also specification of prior edge inclusion probabilities jointly with the block structure prior. For instance, Kemp et al. (2006), Geng et al. (2018) and Legramanti et al. (2022b) place Beta distributions on the edge probabilities, and Reyes and Rodríguez (2016) and Jiang and Tokdar (2021) add structure by using different priors for within- and between-block edge probabilities, while Tan and De Iorio (2019) use a DP to build a joint prior on the partition of nodes and edge probabilities. Additionally, they extend the model to a degree-corrected blockmodel, i.e. they introduce a popularity parameter for each node. Passino and Heard (2020) and Loyal and Chen (2022) consider Bayesian blockmodels where edge probabilities derive from a latent space embedding. Caron and Fox (2017) model edge probabilities by associating edges with realisations from a Poisson process with rate described by a random measure. Herlau et al. (2016) and Todeschini et al. (2020) extend this approach, respectively, to (i) blockmodels, with blocks corresponding to subsets of the support of a random measure; and (ii) to overlapping community detection, with each community corresponding to an element of a compound random measure. The Bayesian nonparametric blockmodel by Peixoto (2017) defines a generative process based on a random partition prior on the block configuration where structural constraints are imposed on the number of edges across blocks. This approach avoids explicit modelling of edge inclusion probabilities.

1.2   Learning Block Structure in Graphical Models

Proposals for the estimation of large-scale structures in graphical models can be categorised in two main strategies: (i) regularisation methods; (ii) imposing structure on the precision matrix. Within the first framework, Ambroise et al. (2009) and Marlin et al. (2009) do not model graphs explicitly, but learn a block structure as part of a shrinkage estimator for the precision matrix as in the graphical lasso (Friedman et al., 2007), where every block is characterised by its own regularisation parameter.

Within the second framework, Sun et al. (2014) consider a GGM with a CRP as prior on the partition of nodes. Then, the partition informs the sparsity pattern of the scale matrix of the Wishart prior on the precision matrix rather than of the precision matrix itself as is commonly done in GGMs. See Section S3 of Supplementary Material (van den Boom et al., 2022b) for details. Marlin and Murphy (2009) impose sparsity in the precision matrix of a GGM by first approximating the joint distribution of the

nodes via the specification of the conditional distribution of each node given the others. Then, they impose a continuous spike-and-slab prior on "edge weights" that capture the association of a node with the others. Finally, the prior on edge weights incorporates a block structure. Sun et al. (2015) fix the number of blocks, place a Dirichlet prior on the partition of the nodes in an exponential graphical model and compute a point estimate of the partition using an expectation-maximisation algorithm.

Bornn and Caron (2011) consider decomposable graphs, which allow modelling of cliques and separators separately, and use a product partition model as prior on the graph. Their prior can induce large cliques and, as such, allows the identification of larger structures than edges. Peixoto (2019) uses a stochastic blockmodel as prior for network reconstruction in two discrete-valued graphical models, i.e. the Ising model and an epidemic model of infection status across time. He shows empirically that joint estimation of the graph and block structure increases accuracy as compared to two-step approaches. Colombi et al. (2022) and Cremaschi et al. (2022) consider inference in GGMs under a known block structure with either all or no edges present between a pair of blocks.

2 Gaussian Graphical Models

Let the graph $G = (V, E)$ be defined by a set of edges $E \subset \{(i, j) \mid 1 \le i < j \le p\}$ that represent links among the nodes in $V = \{1, \ldots, p\}$. The data are represented by an $n \times p$ matrix Y with independent and identically distributed rows corresponding to p-dimensional random vectors whose elements are represented by nodes on the graph. A graphical model (Lauritzen, 1996) is a family of distributions on the rows which is Markov over G. That is, the distribution $p(Y \mid G)$ is such that the i-th and j-th columns of Y are independent conditionally on the other columns if and only if $(i, j) \notin E$.

While our development for learning large-scale structure applies to graphical models in general, here we focus on GGMs (Dempster, 1972), which consider a Gaussian law for $p(Y \mid G)$. Then, each row of Y is distributed according to a Multivariate Gaussian distribution $\mathcal{N}(0_{p \times 1}, \Omega^{-1})$ with precision matrix Ω. The conditional independence structure implied by G implies that $\Omega_{ij} = 0$ if and only if nodes i and j are not connected. For the complete matrix Y, $\Omega_{ij} = 0$ implies that the i-th and j-th columns of Y are independent conditionally on the others. In this context, a blockmodel on G enables learning of sparse block-structured precision matrices where the block structure is unknown.

A popular choice as prior $p(\Omega \mid G)$ for the precision matrix Ω conditional on the graph G is the G-Wishart distribution $\mathcal{W}_G(\delta, D)$ as it induces conjugacy and allows working with non-decomposable graphs (Giudici, 1996; Roverato, 2002). It is parameterised by degrees of freedom $\delta > 2$ and a positive-definite rate matrix D. Then (e.g. Atay-Kayis and Massam, 2005),

$$p(Y \mid G) \propto \int p(\Omega \mid G)\, p(Y \mid \Omega)\, d\Omega = \frac{I_G(\delta^\star, D^\star)}{(2\pi)^{np/2} I_G(\delta, D)}, \tag{1}$$

where $\delta^\star = \delta + n$, $D^\star = D + Y^\top Y$ and $I_G(\delta, D)$ is the normalising constant of the density of $\mathcal{W}_G(\delta, D)$. The constant $I_G(\delta, D)$ is not analytically available for general, non-decomposable G. Thus, we make use of the Markov chain Monte Carlo (MCMC) methodology from van den Boom et al. (2022a) and of a Laplace approximation of $I_G(\delta, D)$ from Moghaddam et al. (2009) to perform posterior inference on G.

3   Graph Priors for Large-Scale Structure Recovery

Key to learning large structure in graphs is specification of a prior $p(G)$ on graphs. To this end, we borrow ideas from random graph theory, adapting them effectively in our context.

Moreover, our approach is based on the Dirichlet process (Ferguson, 1973), a probability model for random probability distributions. Readers familiar with the DP can skip to the next subsection. If a random measure $H \sim \mathrm{DP}(\nu, H_0)$, then H is almost surely discrete. H_0 is the base measure, a distribution around which the DP is centred, while $\nu > 0$ denotes the precision parameter. Due to its discreteness, H admits the well-known "stick-breaking" construction (Sethuraman, 1994) and can be represented as a countable mixture of point masses: $H = \sum_{k=1}^\infty w_k \delta_{\beta'_k}$. Here $\delta_{\beta'_k}$ is a point mass at β'_k, the weights w_k are generated by rescaled Beta distributions, $w_k = \xi_k \prod_{l=1}^{k-1}(1 - \xi_l)$ with $\xi_k \overset{\text{i.i.d.}}{\sim} \mathrm{Beta}(1, \nu)$, and the locations $\{\beta'_k\}_{k=1}^\infty$ are i.i.d. samples from the base measure H_0. Finally, the sequences $\{\beta'_k\}_{k=1}^\infty$ and $\{\xi_k\}_{k=1}^\infty$ are independent.

3.1   Degree-Corrected Stochastic Blockmodel

The fundamental idea behind our strategy is the following. Each node i in the graph forms a connection with another node j according to (i) its own propensity (or popularity) captured by the parameter θ_i; (ii) its block membership captured by the "interaction" parameter β_{ij}, with nodes in the same block having higher probability to share an edge. The popularity parameter can be thought of as the node-specific propensity to form connections with other nodes. To guarantee parsimony, popularity parameters are modelled assuming a DP prior. On the other hand, prior specification on block-specific parameters is more complex, as it depends on the prior on the partition of nodes, which defines the block structure. We exploit the discreteness of the DP to define the blocks, where each component in the DP discrete mixture (see the "stick-breaking construction") corresponds to a block and nodes allocated to the same component share the same block-specific interaction parameter. Figure 1 summarises the modelling strategy.

Our starting point is the Bayesian nonparametric degree-corrected stochastic blockmodel by Tan and De Iorio (2019) who propose a probit model for the edge inclusion probabilities. More specifically, $\Pr\{(i, j) \in E\} = \Phi(\mu_{ij})$, independently over distinct pairs (i, j) for $1 \leq i < j \leq p$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard Gaussian distribution $\mathcal{N}(0, 1)$. Then,

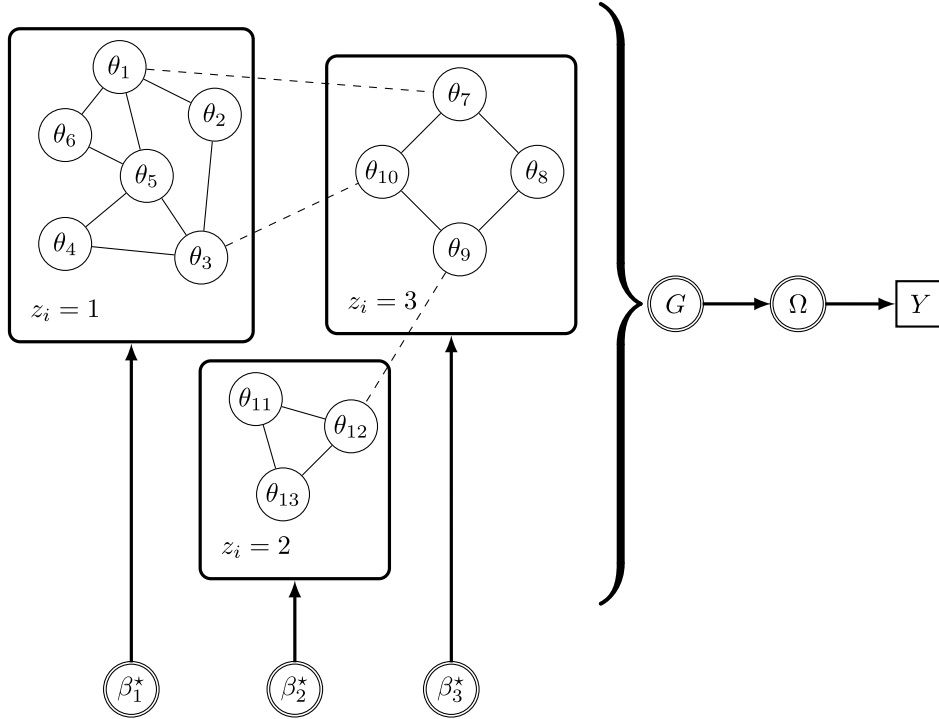$$\mu_{ij} = \theta_i + \theta_j + \beta_{ij}\, \mathbb{1}[z_i = z_j] \tag{2}$$

Figure 1: Visualisation of the degree-corrected stochastic blockmodel. Each node is characterised by its own popularity parameter θ_i. Block membership is captured by the allocation parameter z_i. Nodes in the same block share the same value for z_i as well as for the interaction parameter $\beta_i = \beta_{z_i}^\star$. There are more edges within (solid lines) than between (dashed lines) blocks represented by square boxes. The resulting graph G constrains the precision matrix Ω of the data Y. The direction of the arrows reflects the hierarchical specification of the model.

The allocation variable z_i denotes the community node i belongs to. The parameter β_{ij} measures the strength of interaction among members of the same community with nodes in the same block expected to share more edges among themselves than with nodes outside. The popularity parameter θ_i allows for degree correction in the blockmodel. That is, nodes have varying popularity as captured by the number of their neighbours, i.e. the number of nodes they are connected to via an edge.

To specify a prior on β_{ij}, we introduce auxiliary variable β_i for each node i and assume $\beta_i \mid H \overset{\text{i.i.d.}}{\sim} H$ for $i \in V$ where $H \sim \mathrm{DP}(\nu, H_0)$ with $H_0 = \mathcal{N}(0, s_\beta^2)$, $s_\beta^2 > 0$. The discreteness of the DP implies a positive probability of ties in a sample from $H = \sum_{k=1}^\infty w_k \delta_{\beta_k'}$ and this, in turn, induces a clustering structure so that the nodes will be grouped together in K clusters, where the number K of clusters is unknown and learnt from the data through the posterior distribution. In our context, each cluster

corresponds to a block, the parameters w_k denote the prior probabilities of belonging to each mixture component, and β'_k denotes the block-specific interaction parameter. Nodes are clustered based on their edge inclusion probability, so that nodes in the same block k share a common value β'_k such that in (2) $\beta_{ij} = \beta_i = \beta_j = \beta'_k$. Thus, the set of node-level parameters $\{\beta_i\}_{i=1}^p$ reduces a posteriori to the set of unique values $\beta_1^\star, \ldots, \beta_K^\star$ assigned to within-block edges. We denote block membership with the variable $z_i \in \{1, \ldots K\}$, $i \in V$.

Each node is characterised by its own popularity parameter θ_i, which allows for higher flexibility, but can also challenge identifiability of large-scale structures. This is due to the fact that some connection patterns might be explained nearly as well (in terms of graph likelihood) by either granular specification of the vector θ or by blocks. To avoid such identifiability issues and associated inflated posterior uncertainty, we reduce the number of unique elements of θ through clustering via a DP. Specifically, $\theta_i \mid F \overset{\text{i.i.d.}}{\sim} F$ for $i \in V$, where $F \sim \text{DP}(\alpha, F_0)$, $\alpha > 0$, with $F_0 = \mathcal{N}(0, s_\theta^2)$, $s_\theta^2 > 0$. Lastly, $\nu \sim \Gamma(a_\nu, b_\nu)$ and $\alpha \sim \Gamma(a_\alpha, b_\alpha)$, $a_\nu, b_\nu, a_\alpha, b_\alpha > 0$.

The proposed model has wide applicability as blocks can represent communities. Algorithm S1 in Supplementary Material (van den Boom et al., 2022b) details an MCMC algorithm for posterior inference. We remark that the block-specific parameter $\beta_k^\star$ does not appear in the likelihood for a block k of size one, i.e. when $|\{i \mid z_i = k\}| = 1$, differently from conventional applications of the DP. Therefore, the Metropolis-Hastings proposal to add a block of size one does not require a proposal for the parameter specific to the new block. This property causes Algorithms 2 and 8 in Neal (2000), which we use to update the DP parameters, to coincide.

3.2   Southern Italian Community Structure

In the stochastic blockmodel, nodes in the same block are not necessarily connected. This level of flexibility is particularly desirable when the network is observed directly, and focus is on understanding network formation and evolution. On the other hand, in graphical models, the graph is a latent variable and, in applications, it might be appropriate to impose a more restrictive definition of block/community. More restrictions can also provide the benefit of improving computations as the space to explore gets smaller.

Here we assume that nodes in a block form a clique. In a clique, all pairs of nodes are connected by an edge. Indeed, the earliest approaches for community structure in graphs consider cliques (Luce and Perry, 1949; Festinger, 1949), the rationale being that a community is strongest if each pair of its members is connected. Cliques have for instance biological significance in protein-protein interaction networks (Yu et al., 2006). We thus introduce a Bayesian nonparametric graph prior where each block is a clique.

In this context, main object of inference are the cluster allocation variables z_i, $i \in V$, as in Section 3.1, as then the block structure is given: $\Pr\{(i, j) \in E\} = 1$ if $z_i = z_j$. We name such construction the Southern Italian community structure (SICS) with reference to traditional Southern Italian communities where everybody knows each other. Still, there can be connections between nodes from different blocks and, for these edges, we

assume a prior inclusion probability ρ: $\Pr\{(i,j) \in E\} = \rho$ for $1 \le i < j \le p$ and $z_i \ne z_j$. This construction defines a prior $p(G \mid z)$.

A prior on z completes the graph prior $p(G) = \int p(G \mid z)\, p(z) dz$. We assume that z follows the Chinese restaurant process (Pitman, 2006) with concentration parameter ν a priori for concreteness and consistency with the DP in Section 3.1. We note that our approach is flexible and other priors on the partition of nodes $z = (z_1, \dots, z_p)$ can be straightforwardly adopted borrowing from the rich Bayesian nonparametric literature. The CRP assumption implies:

$$\Pr(z_i = k \mid z_{-i}) = \begin{cases} \frac{n^{\beta}_{-i,k}}{p-1+\nu}, & k = 1, \dots, K^{-i} \\ \frac{\nu}{p-1+\nu}, & k = K^{-i}+1 \end{cases} \tag{3}$$

where $z_{-i} = \{z_j \mid j \ne i\}$, $n^{\beta}_{-i,k} = |\{j \in V \mid z_j = k,\, j \ne i\}|$ and K^{-i} is the number of unique elements in z_{-i}. Finally, $\rho \sim \mathcal{U}(0,1)$ and $\nu \sim \Gamma(a_\nu, b_\nu)$, $a_\nu, b_\nu > 0$, complete prior specification.

We now highlight some of the implications of the SICS prior on the overall graph structure and on the corresponding MCMC algorithm. In the standard GGM framework, moves on the posterior space of graphs usually involve a single edge and consequently, when updating the graph, we only need to integrate out one element of the Cholesky decomposition of the precision matrix (assumed to have a G-Wishart prior) leading to efficient computation (van den Boom et al., 2022a). In such a context, updating more than an edge at a time is extremely challenging. On the other hand, under the SICS framework, change of block membership for a single node can affect a number of edges in the graph:

- The node joins a new block, and forms connections with every node already present in that block, thus a number of edges are introduced in the graph.

- The node is removed from the current block (which is represented by a clique). Thus, as many edges as the number of nodes left in that block are removed. Some of these edges might be readded as there is a positive probability ρ of connection between nodes in different communities.

From a computational point of view, the structure in SICS poses a challenge to posterior inference using MCMC, due to the multiple edge change. Updates of z conditional on G are not possible. Instead, we devise MCMC steps to update z and G jointly, in addition to updating G conditional on z. Section S2.2 of Supplementary Material (van den Boom et al., 2022b) details the MCMC.

The SICS prior is a limiting case of the stochastic blockmodel, obtained taking the limit for $\beta_k^\star \to \infty$ in the model described in Section 3.1. In the limit, blocks become cliques with probability one. Palla et al. (2012) consider a latent factor model that is, conditionally on the factor loadings, equivalent to a GGM with the SICS prior where $\rho = 0$. That is, there is an edge between two nodes if and only if they belong to the same block.

3.3   Multiple Graphs

In applications, it is common that data (i.e. the rows of the observation matrix Y) are naturally clustered due to experimental conditions. For instance, Y could represent gene expression measurements with different groups of rows corresponding to different cancer types. One way to deal with such heterogeneity is a multiple graphical model (e.g. Peterson et al., 2015; Ma and Michailidis, 2016; Mitra et al., 2016; Tan et al., 2017), where each graph corresponds to an experimental condition (e.g. case/control status).

Consider multiple graphs $G_x = (V, E_x)$ and associated data (n_x, Y_x) for $x = 1, \ldots, q$. Here, x indexes the groups such that we have n_x observations in Y_x, with each group characterised by its own graph and data-generating process $p(Y_x \mid G_x)$. We model the graphs $G_1, \ldots, G_q$ jointly through the specification of a prior $p(G_1, \ldots, G_q)$. The goal is to identify common patterns, as well idiosyncratic edge/block structures. We now introduce the multiple graph extension of the degree-corrected stochastic blockmodel, described in Section 3.1. For each graph G_x, we have $\Pr\{(i, j) \in E_x\} = \Phi(\mu_{xij})$ and (2) becomes

$$\mu_{xij} = \theta_i + \theta_j + \beta_{xij}\, \mathbb{1}[z_{xi} = z_{xj} = k]$$

where z_{xi} is the allocation variable for group x, $x = 1, \ldots, q$, and node $i \in V$. Similarly to Section 3.1, we introduce auxiliary variable $\beta_{xi} \mid H \overset{\text{i.i.d.}}{\sim} H$ for $i \in V$, marginally for each x. Then, $\beta_{xij} = \beta_{xi} = \beta_{xj}$ when i, j belong to the same community under condition x. The other parameters of the blockmodel are shared across graphs and have priors as specified in Section 3.1. Thus, marginally for each x, we recover the blockmodel of Section 3.1.

We treat one group (and so graph) as baseline and the other graphs as offset from the baseline group. For ease of notation, we set $x = 1$ as baseline group and, for clarity, corresponding parameters by a subscript 'b'.

There is a vast literature on Dependent Dirichlet processes (DDPs, see, e.g., MacEachern, 1999; De Iorio et al., 2004; Müller et al., 2004; Camerlenghi et al., 2019; Quintana et al., 2022), where the goal is to cluster subjects based also on group information. These tools are not directly applicable to our context as we are actually clustering variables (i.e. nodes on the graph) observed on n_x subjects under each of q experimental conditions (groups). Since we are assuming that $\beta_{xi} \mid H \overset{\text{i.i.d.}}{\sim} H$ marginally for each group x, the same node i under different groups can either belong to a different cluster (block) or to the same. In the multiple graph context, it is then desirable to have $\Pr(z_{bi} = z_{xi}) > \Pr(z_{bi} = z_{xj})$ for $i \neq j$ and $x \geq 2$ to reflect that node i in G_b and in G_x correspond to the same variable and to encourage sharing of large-scale structures across graphs. In a multiple networks context, Reyes and Rodríguez (2016) require the same property. See also the discussion on identifiability for multiple blockmodels in Section 2.2 of Matias and Miele (2016) and arguments for the related concept of separate exchangeability in Lin et al. (2021). On the other hand, DDP models typically assume exchangeability within each group x across subjects which implies that $\Pr(z_{xi} = k) = \Pr(z_{xj} = k)$ a priori. We thus consider the following set-up, where the block structures $\{z_{xi}\}_{i=1}^p$, $x \geq 2$, are estimated as offsets from the baseline $\{z_{bi}\}_{i=1}^p$.

Let $\beta_{bi} \mid H \overset{\text{i.i.d.}}{\sim} H$ for $i \in V$ and $H \sim \text{DP}(\nu, H_0)$ where $H_0 = \mathcal{N}(0, s_\beta^2)$. Then, set $\beta_{xi} = \beta_{bi}$ with probability $\gamma \in (0,1)$ and $\beta_{xi} \mid H \sim H$ with probability $1 - \gamma$, independently for $i \in V$ and $x = 2, \ldots, q$. We note that our construction is invariant to any relabelling of the blocks. This is due to the fact that the distribution of β_{xi} is a mixture of a point mass at β_{bi} and H:

$$\beta_{xi} \mid \beta_{bi}, H \sim \gamma \delta_{\beta_{bi}} + (1 - \gamma)H$$

where H defines the overall block structure (across multiple graphs). Hence, our prior specification allows keeping the same block labels across multiple graphs, while biasing the probability that a node belongs to the same block across conditions. Moreover, the implied dependence across block structures closely resembles the type of dependence across partitions described in Page et al. (2022). In a dynamic clustering framework, they avoid the use of cluster labels by specifying a prior directly on random partitions, inflating the probability of belonging to the same cluster across time.

Posterior computations are greatly simplified by the introduction of binary "genealogical indicators" $g_{xi} \in \{0,1\}$ such that $\beta_{xi} = \beta_{bi}$ if $g_{xi} = 1$ and $\beta_{xi} \mid H \sim H$ if $g_{xi} = 0$. Note that even in the case $g_{xi} = 0$, there is a positive probability that $\beta_{xi} = \beta_{bi}$ due to the discrete nature of H. This implies that the probability of $\beta_{xi} = \beta_{bi}$, $x \geq 2$, is greater than $\Pr(g_{xi} = 1) = \gamma$ a priori conditionally on H. Section S2.3 of Supplementary Material (van den Boom et al., 2022b) details an MCMC algorithm for inference which involves a joint update for (g_{xi}, β_{xi}). The prior dependence among the $z_{1i}, \ldots, z_{qi}$ enables learning of block structure both within and across graphs. The indicators g_{xi} capture the extent to which structure in G_x is shared with G_b. At the same time, the cluster indicators $\{z_{xi}\}_{i=1}^p$ capture the within-graph block structure. Thus, the proposed prior construction allows for borrowing of large-scale information across graphs, as well as the detection of graph-specific blocks.

We want to highlight that our construction differs from the hierarchical Dirichlet process (HDP, Teh et al., 2006), as assuming an HDP-type prior would imply that $\beta_{xi} \mid H_x \overset{\text{ind.}}{\sim} H_x$, $H_x \mid H \overset{\text{i.i.d.}}{\sim} \text{DP}(\cdot, H)$, $H \sim \text{DP}(\nu, H_0)$. This means that the β_{xi} have a positive probability to be equal across group x (and obviously node i), but the same node would not have higher probability to belong to the same block across groups.

There are proposals in the graphical model literature where a graph is considered baseline (e.g. Mukherjee and Speed, 2008; Telesca et al., 2012; Mitra et al., 2016; Tan et al., 2017), but their focus is on differences in individual edges instead of blocks. Moreover, Paul and Chen (2020) consider blockmodels with multiple graphs in a frequentist framework where the number of blocks is known and assume a hierarchical structure for the block memberships under each experimental condition, linking block membership to an unknown baseline membership. Reyes and Rodríguez (2016) and Stanley et al. (2016) induce dependence among graphs by assuming that they either share the same block structure or have unrelated block structures, which leads to a less flexible modelling tool than our approach. Amini et al. (2021) use an HDP to link block structure across graphs which presents the limitation described in the previous paragraph. Ma and Michailidis (2016) assume that multiple graphs share the same known block structure

and only edges between and across blocks might differ across experimental conditions. Edges are estimated using regularised nodewise regression (Zhou et al., 2011), instead of working directly on graph space. In a different context, previous work on hidden Markov models (Fox et al., 2008), including time-varying blockmodels (Ishiguro et al., 2010; Fan et al., 2015), involves a similar dependence for cluster indicators across time, but this dependence is induced through a more involved construction with a "spiked" base measure for the DP on the transition probability vector of the Markov chain.

In our construction, the distribution of $\{z_{xi}\}_{i=1}^p$, $x \geq 2$, is defined conditionally on $\{z_{bi}\}_{i=1}^p$ such that $\{g_{xi}\}_{i=1}^p$ captures differences from $\{z_{bi}\}_{i=1}^p$. Alternative dependence structures could be easily considered within our framework. For instance, instead of setting a group as baseline, we could specify a latent block structure $\{z_{0i}\}_{i=1}^p$ and then define $\{z_{xi}\}_{i=1}^p$, $x \geq 1$, as deviations from $\{z_{0i}\}_{i=1}^p$, for which a prior process needs to be specified (e.g. simply assume H as prior). Finally, the multiple graph set-up can be straightforwardly extended to the SICS prior from Section 3.2.

4 Testing for Large-Scale Structure

In this section, we describe a strategy to test if there is any block structure in an individual graph. Although the description below will only involve one graph for ease of explanation, the same techniques can be employed to test for the presence of structure in multiple graphs.

In model (2), block structure is represented through indicator vector z. Thus, testing for presence of block structure is equivalent to testing whether $K \geq 2$. In the Bayesian paradigm, we can use Bayes factors and here we describe a computational method for their evaluation based on the Savage-Dickey ratio (Dickey, 1971), as they are not available analytically. Consider a prespecified block structure $z^\star$. For instance, $z^\star$ can consist of a single ($K^\star = 1$) block, i.e. no large-scale structure, such that the test assesses the evidence for any block structure. Our method will often be computationally infeasible for other choices of $z^\star$ as discussed later but the idea applies to any $z^\star$ in principle.

The computation of Bayes factors for DP-based models has received attention in the literature though with some drawbacks: the method from Basu and Chib (2003) requires an extra MCMC run with z fixed to $z^\star$ and the use of sequential importance sampling, resulting in an involved strategy, not easily integrated into an existing MCMC implementation. Legramanti et al. (2022a) evaluate the marginal likelihood of each model using the harmonic mean approach (Newton and Raftery, 1994; Raftery et al., 2007), which can be unstable or slow to converge. Application of their method in our (and other's) context would benefit from the direct evaluation of $p(Y \mid z)$ which is not available in closed form. Without an analytical form for $p(Y \mid z)$, the harmonic mean approach requires an extra MCMC run with z fixed to $z^\star$ to approximate $p(Y \mid z^\star)$. Indeed, one of the main advantages of our method is that Bayes factors can be evaluated directly from the MCMC output for the model of interest.

More in details, the Bayes factor of the relative evidence of $z = z^\star$ (model $\mathcal{M}^\star$) over $z \sim p(z)$ (model $\mathcal{M}$) is

$$B = \frac{p(Y \mid \mathcal{M}^\star)}{p(Y \mid \mathcal{M})} = \frac{p(Y \mid z^\star)}{p(Y)} = \frac{p(Y, z^\star)}{p(Y)\, p(z^\star)} = \frac{p(z^\star \mid Y)}{p(z^\star)} \tag{4}$$

where the last ratio is the Savage-Dickey ratio. Note that the second equality uses the property that the prior on all remaining parameters in the model such as β_i are the same under $\mathcal{M}$ and $\mathcal{M}^\star$ in such a way that we recover the same model specification as $\mathcal{M}^\star$ when $z = z^\star$ in $\mathcal{M}$. Now, an estimate $\widehat{B}$ for B is obtained by plugging in the usual (in terms of sample frequency) estimate of $p(z^\star \mid Y)$ derived from the MCMC chain while $p(z^\star)$ is readily computed by numerical quadrature: $p(z) = \int p(z \mid \nu)\, p(\nu)\, d\nu$ where $p(\nu) = \Gamma(\nu \mid a_\nu, b_\nu)$ and (e.g. Legramanti et al., 2022b)

$$p(z \mid \nu) = \frac{\nu^K \prod_{k=1}^{K} |\{i \mid z_i = k\}|!}{\prod_{i=0}^{p-1}(\nu + i)}.$$

This scheme can be employed to compute $\widehat{B}$ from an MCMC chain as long as $p(z^\star)$ is not too small. In that case, reliably estimating $p(z^\star \mid Y)$ might be hard as the MCMC chain could visit $z^\star$ only rarely after convergence. Furthermore, $p(z^\star)$ will often be too small if $z^\star$ corresponds to multiple blocks due to the combinatorially many ways to assign p nodes to $K^\star \geq 2$ blocks, but it usually assumes reasonable values for $z^\star$ corresponding to absence of block structure ($K^\star = 1$), which refers to the conventional null hypothesis of no structure (i.e. "no effect"). For instance, we test for $K^\star = 1$ in the examples considered in this work. We remark that the methods from Basu and Chib (2003) and Legramanti et al. (2022a) do not have such limitation for small $p(z^\star)$, but in general require additional MCMC runs. Note that $p(z^\star \mid Y)$ being estimated as (close to) zero is not problematic, but leads to an accurate estimate of $B \approx 0$ (as long as $p(z^\star)$ is sufficiently far from zero). In Section S4 of Supplementary Material (van den Boom et al., 2022b), we show empirically that the proposed Bayes factor estimation converges faster than the harmonic mean approach.

5 Simulation Studies

We demonstrate the performance of our approach in two simulation scenarios. For all empirical results, we set the hyperparameters as $\gamma = 0.5$, $s_\beta^2 = s_\theta^2 = 1$, $a_\nu = b_\nu = a_\alpha = b_\alpha = 2$, $\delta = 3$ and $D = I_p$ unless otherwise stated. See Section S1 of Supplementary Material (van den Boom et al., 2022b) for an overview of the models.

5.1 Karate Club Network

We investigate the importance of uncertainty propagation when learning community structure in a graphical model. As true underlying graph G, we consider the karate club network (Zachary, 1977) which Tan and De Iorio (2019) analyse using the degree-corrected stochastic blockmodel of Section 3.1. The network's $p = 34$ nodes correspond

to members of a karate club while its 78 edges signify friendships between members. Conditionally on G, we sample a precision matrix $\Omega \mid G \sim \mathcal{W}_G(\delta, D)$. The n rows of Y are sampled according to the GGM in Section 2 independently from $\mathcal{N}(0_{p \times 1}, \Omega^{-1})$. Finally, we fit the model from Section 3.1 using 6000 MCMC iterations, discarding the first 1000 as burn-in. In this case, we set $a_\nu = b_\nu = a_\alpha = b_\alpha = 5$ as in Tan and De Iorio (2019) for a fair comparison.

We repeat the simulation for $n = 10^4, 10^3, 10^2, 10$ while keeping Ω the same and present the estimated community structure in Figure 2. For $n = 10^4$, the results are very close to those in Tan and De Iorio (2019) where the underlying network is known, with the two main blocks corresponding to the karate instructor Mr Hi and the club's president John A. The increased uncertainty in the estimation of G for smaller values of n obviously affects inference on the block structure, with too little information present in the data with only $n = 10$ observations to recover the two main blocks. This is also reflected in the estimate of the Bayes factor comparing the model with no block structure vs the model with $z \sim p(z)$: $\widehat{B} = 0$ for $n = 10^4, 10^3$, $\widehat{B} = 0.28$ for $n = 10^2$ and $\widehat{B} = 0.40$ for $n = 10$. These results show that uncertainty in graph estimation can have a major impact on community estimation and this uncertainty should not be ignored as is often done in applications where a two-step approach is adopted (first graph estimation and then blocks).

5.2 Block Structure Recovery

We now investigate how accurately the proposed methodology can recover block structure. We assign $p = 20$ nodes to K clusters by sampling z_i with replacement from $\{1, \ldots, K\}$ for $i \in V$. Then, we generate a graph G according to the SICS prior from Section 3.2 with between-block edge inclusion probability $\rho = 0.2$. Data Y corresponding to G are sampled as in Section 5.1. We consider the following scenarios: $K = 4$ for $n = 20, 100, 500, 1000$ and $n = 500$ for $K = 2, 3, 4, 5$. The performance of the algorithms is assessed over 50 replicates for each scenario.

We fit both models from Sections 3.1 and 3.2, as well as the model by Sun et al. (2014) (see Section S3 of Supplementary Material (van den Boom et al., 2022b) for a description) for comparison. We run the MCMC for 1000 iterations after a burn-in of 500 for the for the stochastic blockmodel and the model by Sun et al. (2014) while we record 5000 iterations after a burn-in of 1000 for the SICS model to account for the slower convergence and mixing of its MCMC.

The cluster allocation vector z informs the block structure. As point estimate for z, we report the configuration that minimises the posterior expectation of Binder's (1978) loss function under equal misclassification costs, which is a common choice in the applied Bayesian nonparametrics literature (Lau and Green, 2007). See Appendix B of Argiento et al. (2014) for computational details. Briefly, this expectation of the loss measures the difference for all possible pairs of nodes between the posterior probability of co-clustering and the estimated cluster allocation. Following Sun et al. (2014), we use the Rand (1971) index to quantify the difference between the true allocation and
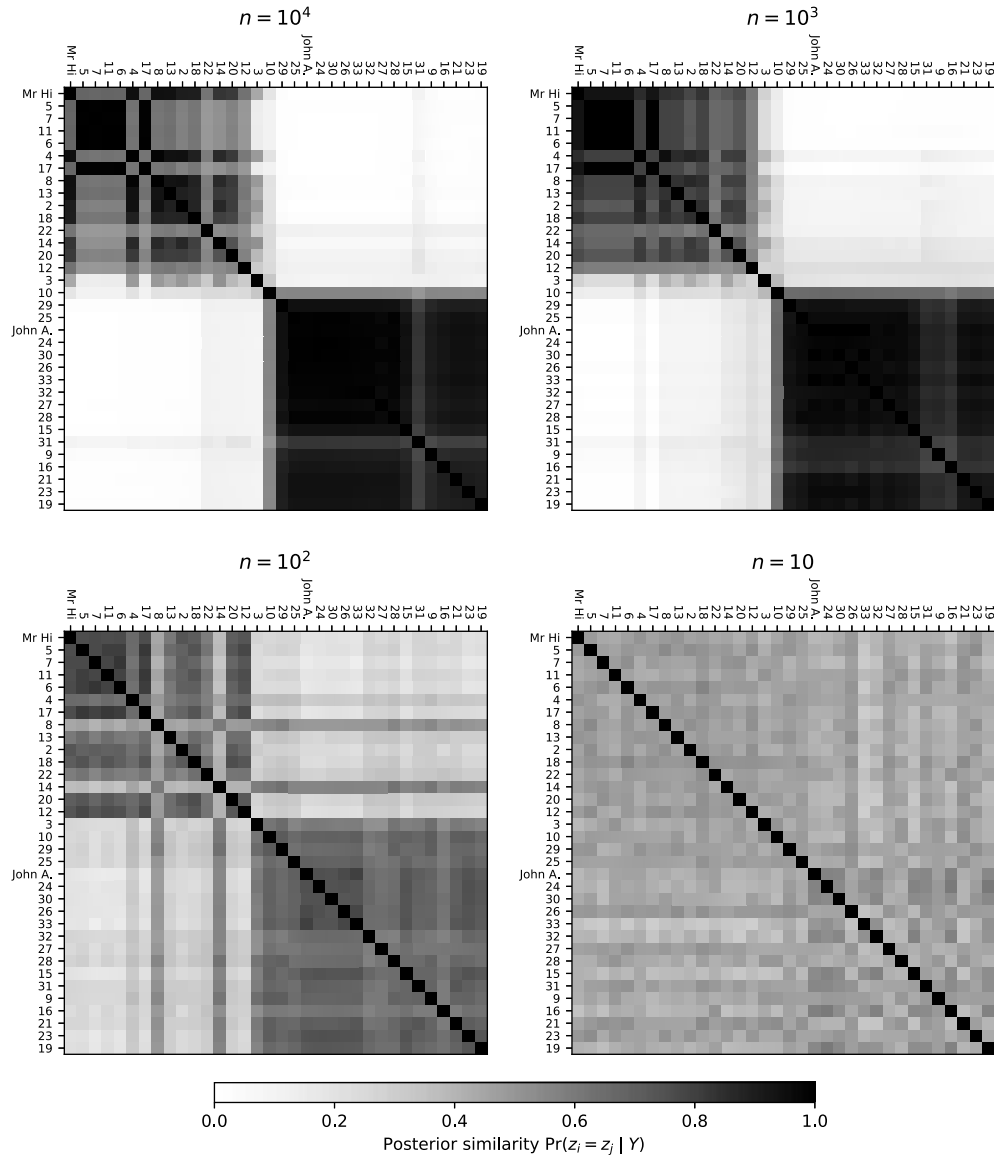
Figure 2: Karate club network: posterior similarity matrices for the simulation studies. The nodes are ordered as in Tan and De Iorio (2019).

its Binder point estimate. A Rand index of one corresponds to a perfect match while a lower value indicates worse block structure recovery.

Figure 3 shows that the proposed methodology recovers the block structure comparably to or substantially more accurately than the model by Sun et al. (2014). The
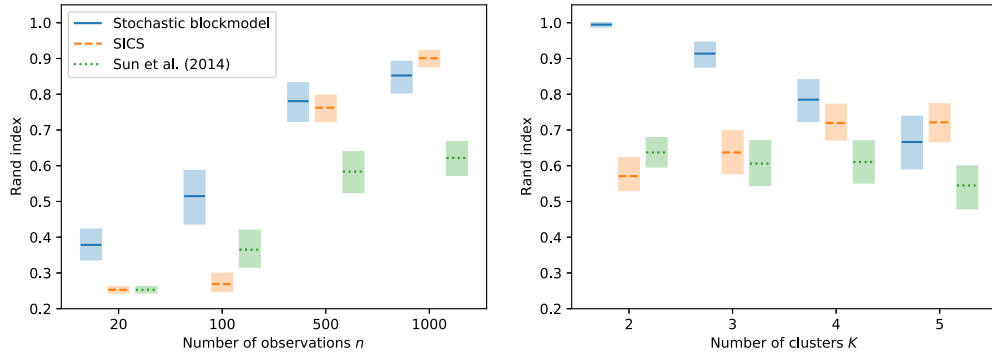
Figure 3: Block structure recovery: Rand index versus the number of observations (left) and the number of clusters (right). The lines represent means over the 50 replicates for the stochastic blockmodel, SICS and the model by Sun et al. (2014). The shaded areas are 95% bootstrapped confidence intervals.

superior performance of the stochastic blockmodel over SICS, when occurring, is most likely due to the fact that the SICS model imposes more stringent assumptions on the correlation structure of the data, which might not be captured with small sample sizes (left panel of Figure 3). This is in line with power considerations for detecting correlation in a frequentist framework, with large sample size usually required, especially for partial correlations (see, e.g., Castelo and Roverato, 2006; Knudson and Lindsey, 2014). Secondly, the SICS structure is more easily recovered when fewer nodes belong to a block (right panel of Figure 3) as this relaxes the assumption on the overall dependence structure among the random variables. Finally, posterior inference for the stochastic blockmodel is performed through an exact MCMC (van den Boom et al., 2022a) while, for the SICS, we employ a Laplace approximation for the graph likelihood to update z and G jointly. See Section S2 of Supplementary Material (van den Boom et al., 2022b).

6 Applications

We apply the proposed models to two real data sets. We discuss MCMC mixing and convergence in Section S5 of Supplementary Material (van den Boom et al., 2022b).

6.1 Mutual Fund Data

We consider data on monthly returns of $p = 59$ mutual funds described in Scott and Carvalho (2008). The funds are divided into four types by the sectors they invest in with 13 funds investing in U.S. bonds, 30 in U.S. stocks, 7 in both U.S. stocks and bonds, and 9 in international stocks. The data contain observations on $n = 86$ months. Here, we ignore the dependence of the returns across time and focus on the dependence between funds as in Scott and Carvalho (2008) and Marlin et al. (2009). Note that time dependence could be easily incorporated through a mean term. The returns are
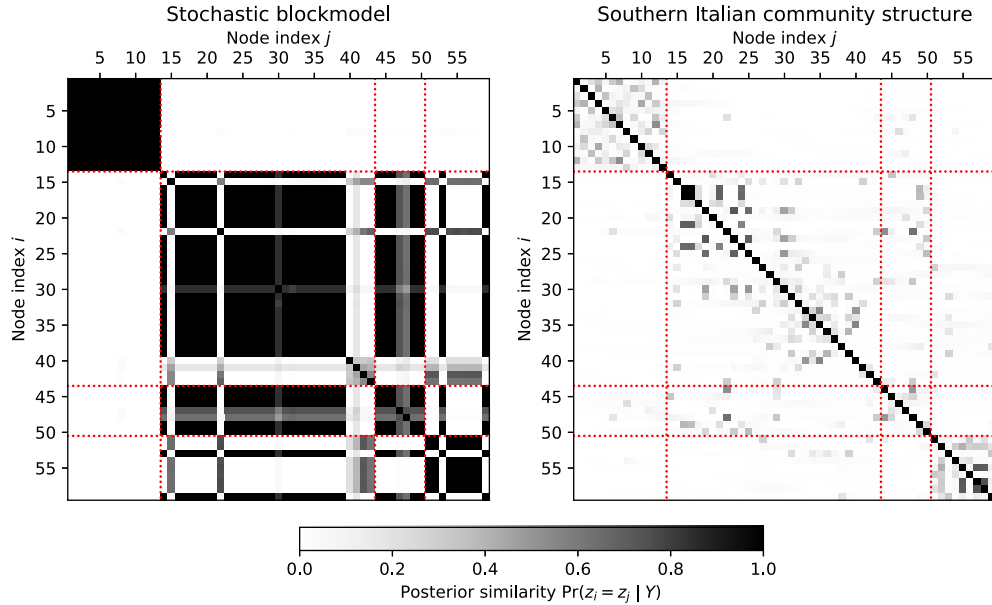
Figure 4: Mutual fund data: posterior similarity matrices. The red dotted lines demarcate the fund types.

quantile-normalised so that they marginally follow a standard Gaussian distribution. We fit both the degree-corrected stochastic blockmodel from Section 3.1 and SICS from Section 3.2. We run the MCMC chain for 15000 iterations discarding the first 5000 as burn-in for the stochastic blockmodel, and for 110000 iterations discarding the first 10000 for SICS.

The stochastic blockmodel identifies clear blocks of funds per Figures 4 (left panel) and 5. Specifically, the U.S. bonds, U.S. stocks and bonds, and international stocks funds are each blocked together without overlap between these fund types except for two international stocks funds that are grouped with the U.S. stocks and bonds. The other funds, which invest in U.S. stocks, are mostly blocked with the U.S. stocks and bonds, or with the international stocks but not with the U.S. bonds. These results are intuitive as funds with a mixture of U.S. stocks and bonds make investments which overlap with funds with only U.S. stocks, and correlation between the returns of U.S. and international stocks is likely. This identified block structure is notably more in line with the fund types than the blocking results presented in Marlin et al. (2009) obtained by shrinkage estimation and optimisation, where only a clear separation of the U.S. bonds funds from the others is detected.

The SICS prior leads to a large number of blocks with a posterior mode at $K = 24$ blocks. A larger number of blocks with SICS than with the stochastic blockmodel is expected as SICS' definition of a block as a clique is more stringent such that larger blocks are less likely to appear. The large-scale pattern of the similarity matrix for
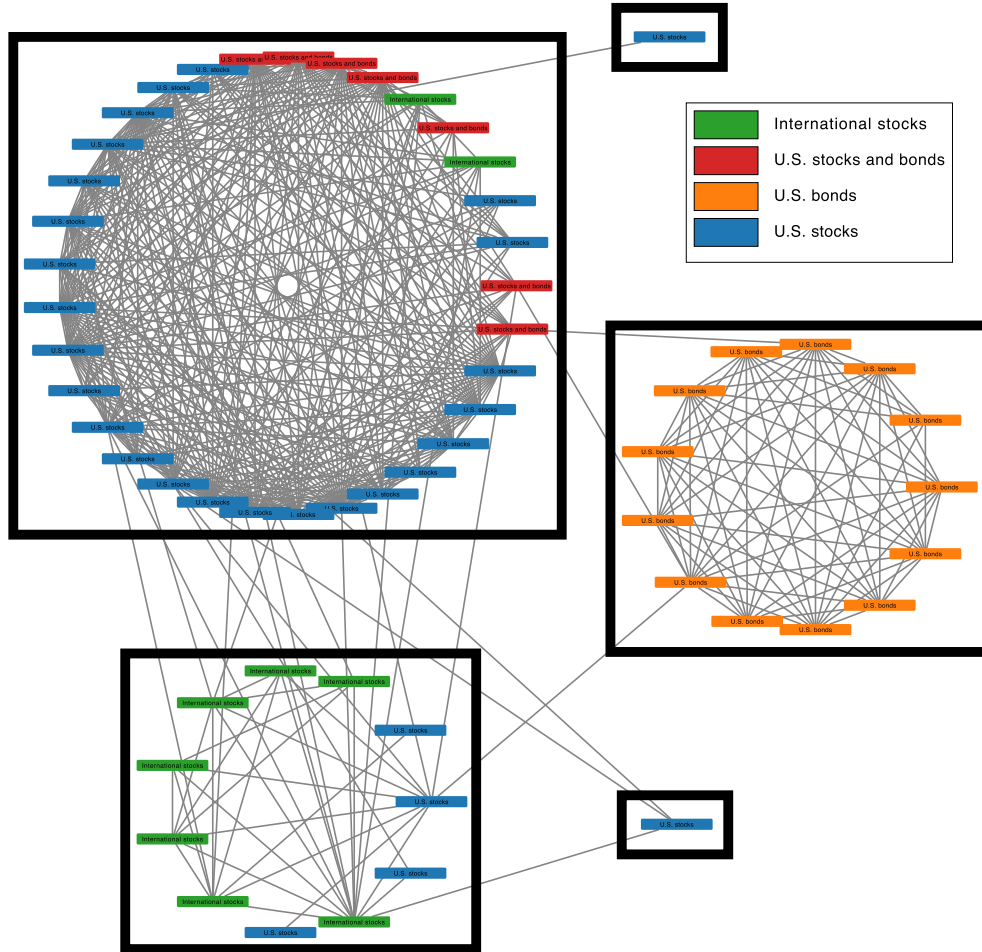
Figure 5: Mutual fund data: median probability graph (Barbieri and Berger, 2004), which consists of edges with posterior inclusion probability greater than 0.5, from the stochastic blockmodel. Colour of nodes refers to fund type. Boxes group nodes belonging to blocks estimated by minimising Binder's loss function.

SICS is still similar to that of the stochastic blockmodel in Figure 4, though with much lower values for $\Pr(z_i = z_j \mid Y)$. Still, the posterior fit indicates strong evidence for the presence of a block structure as the Bayes factor in favour of absence of block structure is $\widehat{B} = 0$.

Generally, the posterior inference contains information on whether the stochastic blockmodel or SICS is the most appropriate model for the data at hand. We assess this by computing the Bayes factor of the stochastic blockmodel versus SICS using the harmonic mean approach. This results in a log Bayes factor of 235 indicating strong evidence that the stochastic blockmodel fits the data better than SICS.
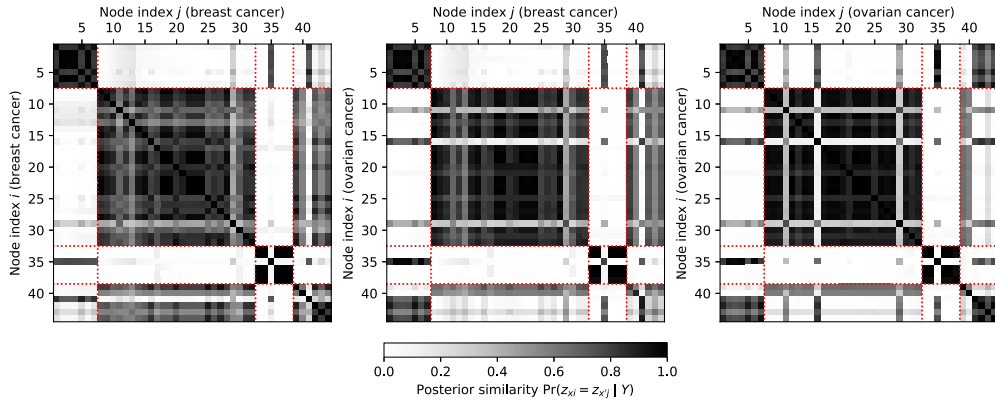
Figure 6: Gene expression data: posterior similarity matrices. The three matrices visualise the posterior probabilities of $z_{1i} = z_{1j}$, $z_{1i} = z_{2j}$ and $z_{2i} = z_{2j}$, respectively, where $x = 1$ corresponds to breast cancer and $x = 2$ to ovarian cancer. The red dotted lines demarcate the modules identified by Zhang (2018).

6.2 Gene Expression Data

As a second application, we consider gene expression levels, the interactions between which are often represented as networks. An important concept in the gene network literature is that of module, which is a densely connected subgraph. Thus, learning the block structure of a graph allows module detection. Typically, a two-step approach is adopted: first the graph is estimated from the gene expressions, and then the modules are derived from the graph estimate (see, e.g., Zhang, 2018), which underestimates uncertainty and often leads to false positives.

We analyse data on gene expressions from $n_1 = 590$ breast cancer tissue samples and $n_2 = 561$ ovarian cancer samples from The Cancer Genome Atlas. We focus on $p = 44$ genes identified by Zhang (2018) as spread across four estimated modules (Modules 6, 14, 36 and 39 in Table 2 of the cited paper) which are highly enriched in terms of Gene Ontology (GO, Ashburner et al., 2000) annotations. For each cancer, the gene expressions are quantile-normalised to marginally follow a standard Gaussian distribution. We apply the proposed multiple graph methodology from Section 3.3 with $q = 2$ separate groups, corresponding to the two different cancers. We run the MCMC algorithm for 55000 iterations, discarding the first 5000 as burn-in.

Posterior inference on block structure, shown in Figure 6, carries strong similarities with the modules identified by Zhang (2018). For ease of discussion, we refer to the modules from Zhang (2018) as Module 1 (comprising Nodes 1 through 7), 2 (Nodes 8 through 32), 3 (Nodes 33 through 38) and 4 (Nodes 39 through 44), and highlight them in Figure 6. The proposed methodology finds differences in block structure between breast and ovarian cancer (see middle panel of Figure 6) as well as differences from Zhang's modules, which are forced to be the same across both cancers by construction. Across both breast and ovarian cancer, we find that Nodes 39 and 40 (GSTM3 and

BCAR3 genes, respectively) are grouped with genes from Module 2, which has GO annotations relating to inflammatory response. For Breast cancer, we cluster Nodes 42 through 44 (GSTM1, GSTM2 and GSTM5) with Module 2 while we put them together with Module 1, which has GO annotations relating to pattern specification, for ovarian cancer. Note that these three genes are paralogs of each other which suggests that they have similar function. Finally, Nodes 35 and 41 (HOXB13 and GSTM4) are together with nodes in Module 1 for both cancers. These results show the flexibility of the proposed model to capture differences as well as commonalities in large-scale dependence structure across multiple biological conditions.

7   Discussion

In this work, we combine advances from random graph theory with graphical models to obtain joint estimation of the graph and its large-scale structure. The resulting graphical models are able to go beyond estimation of individual edges to provide inference on the community structure of the graph, while appropriately propagating uncertainty in the estimation. We introduce a novel DDP prior process tailored to the multiple graph setting and propose a convenient computation of Bayes factors in partition models. Advantages of the proposed approach include interpretability, flexibility (due to the nonparametric component) and wide applicability. We focus on two different block structures: stochastic blockmodels and SICS. We note that the SICS prior is more suitable in applications where strong partial correlation between a small number of nodes is expected.

Alternative priors on the block structure could be considered. For instance, Gibbs-type priors (Gnedin and Pitman, 2005) and microclustering priors (Betancourt et al., 2020) are drop-in replacements for the respective DP terms used in (3), and cover a wide range of partition priors such as mixture with random number of components (Miller and Harrison, 2017; Geng et al., 2018; Argiento and De Iorio, 2022). In general, standard Bayesian nonparametric priors assume that the location parameters are i.i.d. draws from a base measure. Note that, in our case, the location parameters, $\beta_k^\star$, correspond to the within-block interaction strength and they can be arbitrarily close or far given the prior assumptions. On the other hand, in the context of mixture with random number of components, it is easier to introduce constraints on the locations, if the application or interpretability require it. For instance, we might impose that the $\beta_k^\star$ vary significantly across blocks and, to that end, assume a repulsive mixture prior (Petralia et al., 2012).

We focus on GGMs for convenience and because of their popularity. Our methodological contribution is however not constrained to this specific set-up and can be extended to work with other graphical models, e.g. to graphs with discrete or mixed type nodes. Moreover, our computational strategy to estimate Bayes factors finds general applicability in the context of Bayesian nonparametric models to test the presence/absence of a partition structure.

We conclude the discussion with few remarks on asymptotic properties of the proposed methodology. Posterior contraction on the graph space does not imply contraction for the block structure and vice versa: posterior contraction of the graph requires the

number of observations $n \to \infty$. See Lee and Cao (2021) and Niu et al. (2021), and references therein. On the other hand, asymptotic results for Bayesian learning of block structure involve the number of nodes $p \to \infty$ (e.g. Geng et al., 2018; Gao et al., 2020; Jiang and Tokdar, 2021). For n large, the graph might be estimated with high precision. In that scenario, considerable posterior uncertainty about the block structure can remain as such uncertainty can be present even if the graph is observed. Furthermore, for large p, block structure can be recovered accurately even if the estimates of the overall graph connection pattern are characterised by high uncertainty.

Supplementary Material

Supplement (DOI: 10.1214/22-BA1338SUPP; .pdf). Overview of notation and models, details of the MCMC algorithms, description of the model by Sun et al. (2014), empirical results for the Bayes factor approach, and MCMC trace plots.
Code for the empirical results is available at `https://github.com/willemvandenboom/graph-substructures`.

References

Abbe, E. (2018). "Community detection and stochastic block models: Recent developments." *Journal of Machine Learning Research*, 18(177): 1–86. MR3827065. 3

Ambroise, C., Chiquet, J., and Matias, C. (2009). "Inferring sparse Gaussian graphical models with latent structure." *Electronic Journal of Statistics*, 3: 205–238. MR2495837. doi: `https://doi.org/10.1214/08-EJS314`. 4

Amini, A. A., Paez, M. S., and Lin, L. (2021). "Hierarchical stochastic block model for community detection in multiplex networks." arXiv:1904.05330v2. 11

Argiento, R., Cremaschi, A., and Guglielmi, A. (2014). "A "density-based" algorithm for cluster analysis using species sampling Gaussian mixture models." *Journal of Computational and Graphical Statistics*, 23(4): 1126–1142. MR3270714. doi: `https://doi.org/10.1080/10618600.2013.856796`. 14

Argiento, R. and De Iorio, M. (2022). "Is infinity that far? A Bayesian nonparametric perspective of finite mixture models." *Annals of Statistics*, forthcoming. arXiv:1904.09733v1. 4, 20

Armstrong, H., Carter, C. K., Wong, K. F. K., and Kohn, R. (2009). "Bayesian covariance matrix estimation using a mixture of decomposable graphical models." *Statistics and Computing*, 19(3): 303–316. MR2516221. doi: `https://doi.org/10.1007/s11222-008-9093-8`. 2

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). "Gene Ontology: Tool for the unification of biology." *Nature Genetics*, 25(1): 25–29. doi: `https://doi.org/10.1038/75556`. 19

Atay-Kayis, A. and Massam, H. (2005). "A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models." *Biometrika*, 92(2): 317–335. MR2201362. doi: https://doi.org/10.1093/biomet/92.2.317. 5

Barabási, A.-L. (2016). *Network Science*. Cambridge University Press, Cambridge, England. 2

Barbieri, M. M. and Berger, J. O. (2004). "Optimal predictive model selection." *Annals of Statistics*, 32(3): 870–897. MR2065192. doi: https://doi.org/10.1214/009053604000000238. 18

Basu, S. and Chib, S. (2003). "Marginal likelihood and Bayes factors for Dirichlet process mixture models." *Journal of the American Statistical Association*, 98(461): 224–235. MR1965688. doi: https://doi.org/10.1198/01621450338861947. 3, 12, 13

Betancourt, B., Zanella, G., and Steorts, R. C. (2020). "Random partition models for microclustering tasks." *Journal of the American Statistical Association*. Advance online publication. MR4480707. doi: https://doi.org/10.1080/01621459.2020.1841647. 20

Binder, D. A. (1978). "Bayesian cluster analysis." *Biometrika*, 65(1): 31–38. MR0501592. doi: https://doi.org/10.1093/biomet/65.1.31. 14, 15, 18

Bornn, L. and Caron, F. (2011). "Bayesian clustering in decomposable graphs." *Bayesian Analysis*, 6(4): 829–846. MR2869966. doi: https://doi.org/10.1214/11-BA630. 2, 5

Camerlenghi, F., Dunson, D. B., Lijoi, A., Prünster, I., and Rodríguez, A. (2019). "Latent nested nonparametric priors (with discussion)." *Bayesian Analysis*, 14(4): 1303–1356. MR4044854. doi: https://doi.org/10.1214/19-BA1169. 3, 10

Caron, F. and Fox, E. B. (2017). "Sparse graphs using exchangeable random measures." *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 79(5): 1295–1366. MR3731666. doi: https://doi.org/10.1111/rssb.12233. 4

Castelo, R. and Roverato, A. (2006). "A robust procedure for Gaussian graphical model search from microarray data with p larger than n." *Journal of Machine Learning Research*, 7(94): 2621–2650. MR2274453. 16

Colombi, A., Argiento, R., Paci, L., and Pini, A. (2022). "Learning block structured graphs in Gaussian graphical models." arXiv:2206.14274v2. 5

Cremaschi, A., Argiento, R., De Iorio, M., Cai, S., Chong, Y. S., Meaney, M., and Kee, M. (2022). "Seemingly unrelated multi-state processes: A Bayesian semiparametric approach." *Bayesian Analysis*. Advance online publication. doi: https://doi.org/10.1214/22-BA1326. 5

De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004). "An ANOVA model for dependent random measures." *Journal of the American Statistical Association*, 99(465): 205–215. MR2054299. doi: https://doi.org/10.1198/016214504000000205. 3, 10

Dempster, A. P. (1972). "Covariance selection." *Biometrics*, 28(1): 157–175. MR3931974. 3, 5

Dickey, J. M. (1971). "The weighted likelihood ratio, linear hypotheses on normal location parameters." *Annals of Mathematical Statistics*, 42(1): 204–223. MR0309225. doi: https://doi.org/10.1214/aoms/1177693507. 3, 12

Erdős, P. and Rényi, A. (1959). "On random graphs I." *Publicationes Mathematicae Debrecen*, 6(3-4): 290–297. MR0120167. 2

Fan, X., Cao, L., and Xu, R. Y. D. (2015). "Dynamic infinite mixed-membership stochastic blockmodel." *IEEE Transactions on Neural Networks and Learning Systems*, 26(9): 2072–2085. MR3453141. doi: https://doi.org/10.1109/TNNLS.2014.2369374. 12

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *Annals of Statistics*, 1(2): 209–230. MR0350949. doi: https://doi.org/10.1214/aos/1176342360. 4, 6

Festinger, L. (1949). "The analysis of sociograms using matrix algebra." *Human Relations*, 2(2): 153–158. doi: https://doi.org/10.1177/001872674900200205. 8

Fienberg, S. E. (2012). "A brief history of statistical models for network analysis and open challenges." *Journal of Computational and Graphical Statistics*, 21(4): 825–839. MR3005799. doi: https://doi.org/10.1080/10618600.2012.738106. 2, 3

Fortunato, S. and Hric, D. (2016). "Community detection in networks: A user guide." *Physics Reports*, 659: 1–44. MR3566093. doi: https://doi.org/10.1016/j.physrep.2016.09.002. 1, 3

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2008). "An HDP-HMM for systems with state persistence." In *Proc. of the 25th Int. Conf. on Machine Learning*, 312–319. ACM Press. 12

Friedman, J., Hastie, T., and Tibshirani, R. (2007). "Sparse inverse covariance estimation with the graphical lasso." *Biostatistics*, 9(3): 432–441. doi: https://doi.org/10.1093/biostatistics/kxm045. 2, 4

Funke, T. and Becker, T. (2019). "Stochastic block models: A comparison of variants and inference methods." *PLOS ONE*, 14(4): e0215296. doi: https://doi.org/10.1371/journal.pone.0215296. 3

Gao, C. and Ma, Z. (2021). "Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing." *Statistical Science*, 36(1): 16–33. MR4194201. doi: https://doi.org/10.1214/19-STS736. 3

Gao, C., van der Vaart, A. W., and Zhou, H. H. (2020). "A general framework for Bayes structured linear models." *Annals of Statistics*, 48(5): 2848–2878. MR4152123. doi: https://doi.org/10.1214/19-AOS1909. 4, 21

Geng, J., Bhattacharya, A., and Pati, D. (2018). "Probabilistic community detection with unknown number of communities." *Journal of the American Statistical Asso-*

ciation, 114(526): 893–905. MR3963189. doi: https://doi.org/10.1080/01621459. 2018.1458618.   4, 20, 21

Giudici, P. (1996). "Learning in graphical Gaussian models." In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics 5*, 621–628. Oxford University Press, Oxford. MR1425431.   5

Gnedin, A. V. and Pitman, J. (2005). "Exchangeable Gibbs partitions and Stirling triangles." *Zapiski Nauchnykh Seminarov*, 138(3): 5674–5685. MR2160320. doi: https:// doi.org/10.1007/s10958-006-0335-z.   20

Herlau, T., Schmidt, M. N., and Mørup, M. (2016). "Completely random measures for modelling block-structured sparse networks." In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.   4

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). "Stochastic blockmodels: First steps." *Social Networks*, 5(2): 109–137. MR0718088. doi: https://doi.org/10.1016/ 0378-8733(83)90021-7.   2, 3

Ishiguro, K., Iwata, T., Ueda, N., and Tenenbaum, J. (2010). "Dynamic infinite relational model for time-varying relational data analysis." In Lafferty, J., Williams, C., Shawe-Taylor, J., Zemel, R., and Culotta, A. (eds.), *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc.   12

Jiang, S. and Tokdar, S. (2021). "Consistent Bayesian community detection." arXiv:2101.06531v1.   4, 21

Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., and Ueda, N. (2006). "Learning systems of concepts with an infinite relational model." In *Proc. of the 21st National Conf. on Artificial Intelligence – Volume 1*, 381–388. AAAI Press.   4

Knudson, D. V. and Lindsey, C. (2014). "Type I and type II errors in correlations of various sample sizes." *Comprehensive Psychology*, 3. doi: https://doi.org/10. 2466/03.CP.3.1.   16

Lau, J. W. and Green, P. J. (2007). "Bayesian model-based clustering procedures." *Journal of Computational and Graphical Statistics*, 16(3): 526–558. MR2351079. doi: https://doi.org/10.1198/106186007X238855.   14

Lauritzen, S. L. (1996). *Graphical Models*. Oxford Statistical Science Series. The Clarendon Press, Oxford University Press, New York. MR1419991.   1, 5

Lee, C. and Wilkinson, D. J. (2019). "A review of stochastic block models and extensions for graph clustering." *Applied Network Science*, 4(122): 1–50. doi: https://doi.org/ 10.1007/s41109-019-0232-2.   1, 3

Lee, K. and Cao, X. (2021). "Bayesian inference for high-dimensional decomposable graphs." *Electronic Journal of Statistics*, 15(1): 1549–1582. MR4255308. doi: https://doi.org/10.1214/21-ejs1822.   21

Legramanti, S., Rigon, T., and Durante, D. (2022a). "Bayesian testing for exoge-

nous partition structures in stochastic block models." *Sankhya A*, 84(1): 108–126. MR4402748. doi: https://doi.org/10.1007/s13171-020-00231-2. 3, 12, 13

Legramanti, S., Rigon, T., Durante, D., and Dunson, D. B. (2022b). "Extended stochastic block models with application to criminal networks." *Annals of Applied Statistics*, 16(4): 2369–2395. MR4489215. doi: https://doi.org/10.1214/21-aoas1595. 4, 13

Lin, Q., Rebaudo, G., and Müller, P. (2021). "Separate exchangeability as modeling principle in Bayesian nonparametrics." arXiv:2112.07755v1. 10

Loyal, J. D. and Chen, Y. (2022). "A Bayesian nonparametric latent space approach to modeling evolving communities in dynamic networks." *Bayesian Analysis*. Advance online publication. doi: https://doi.org/10.1214/21-BA1300. 4

Luce, R. D. and Perry, A. D. (1949). "A method of matrix analysis of group structure." *Psychometrika*, 14(2): 95–116. MR0035974. doi: https://doi.org/10.1007/BF02289146. 8

Ma, J. and Michailidis, G. (2016). "Joint structural estimation of multiple graphical models." *Journal of Machine Learning Research*, 17(166): 1–48. MR3555057. 10, 11

Maathuis, M., Drton, M., Lauritzen, S., and Wainwright, M. (2019). *Handbook of Graphical Models*. Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL. MR3889064. 2

MacEachern, S. N. (1999). "Dependent nonparametric processes." In *ASA 1999 Proc. of the Section on Bayesian Statistical Science*. American Statistical Association. 3, 10

Marlin, B. M. and Murphy, K. P. (2009). "Sparse Gaussian graphical models with unknown block structure." In *Proc. of the 26th Annual Int. Conf. on Machine Learning*, 705–712. ACM Press. 4

Marlin, B. M., Schmidt, M., and Murphy, K. P. (2009). "Group sparse priors for covariance estimation." In *Proc. of the Twenty-Fifth Conf. on Uncertainty in Artificial Intelligence*, 383–392. AUAI Press. 2, 4, 16, 17

Matias, C. and Miele, V. (2016). "Statistical clustering of temporal networks through a dynamic stochastic block model." *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 79(4): 1119–1141. MR3689311. doi: https://doi.org/10.1111/rssb.12200. 10

Miller, J. W. and Harrison, M. T. (2017). "Mixture models with a prior on the number of components." *Journal of the American Statistical Association*, 113(521): 340–356. MR3803469. doi: https://doi.org/10.1080/01621459.2016.1255636. 4, 20

Mitra, R., Müller, P., and Ji, Y. (2016). "Bayesian graphical models for differential pathways." *Bayesian Analysis*, 11(1): 99–124. MR3447093. doi: https://doi.org/10.1214/14-BA931. 10, 11

Moghaddam, B., Khan, E., Murphy, K. P., and Marlin, B. M. (2009). "Accelerating Bayesian structural inference for non-decomposable Gaussian graphical models." In

Advances in Neural Information Processing Systems, volume 22. Curran Associates, Inc.   6

Mukherjee, S. and Speed, T. P. (2008). "Network inference using informative priors." *Proceedings of the National Academy of Sciences of the United States of America*, 105(38): 14313–14318. doi: https://doi.org/10.1073/pnas.0802272105.   11

Müller, P., Quintana, F., and Rosner, G. (2004). "A method for combining inference across related nonparametric Bayesian models." *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 66(3): 735–749. MR2088779. doi: https://doi.org/10.1111/j.1467-9868.2004.05564.x.   3, 10

Neal, R. M. (2000). "Markov chain sampling methods for Dirichlet process mixture models." *Journal of Computational and Graphical Statistics*, 9(2): 249–265. MR1823804. doi: https://doi.org/10.2307/1390653.   8

Newman, M. E. J. (2011). "Communities, modules and large-scale structure in networks." *Nature Physics*, 8(1): 25–31. MR2676073. doi: https://doi.org/10.1093/acprof:oso/9780199206650.001.0001.   1, 2

Newton, M. A. and Raftery, A. E. (1994). "Approximate Bayesian inference with the weighted likelihood bootstrap." *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 56(1): 3–48. MR1257793. doi: https://doi.org/10.1111/j.2517-6161.1994.tb01956.x.   12

Ni, Y., Baladandayuthapani, V., Vannucci, M., and Stingo, F. C. (2022). "Bayesian graphical models for modern biological applications." *Statistical Methods & Applications*, 31(2): 197–225. MR4426829. doi: https://doi.org/10.1007/s10260-021-00572-8.   2

Niu, Y., Pati, D., and Mallick, B. K. (2021). "Bayesian graph selection consistency under model misspecification." *Bernoulli*, 27(1): 636–672. MR4177384. doi: https://doi.org/10.3150/20-BEJ1253.   21

Page, G. L., Quintana, F. A., and Dahl, D. B. (2022). "Dependent modeling of temporal sequences of random partitions." *Journal of Computational and Graphical Statistics*, 31(2): 614–627. MR4425090. doi: https://doi.org/10.1080/10618600.2021.1987255.   11

Palla, K., Ghahramani, Z., and Knowles, D. (2012). "A nonparametric variable clustering model." In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.   9

Passino, F. S. and Heard, N. A. (2020). "Bayesian estimation of the latent dimension and communities in stochastic blockmodels." *Statistics and Computing*, 30(5): 1291–1307. MR4137252. doi: https://doi.org/10.1007/s11222-020-09946-6.   4

Paul, S. and Chen, Y. (2020). "A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging." *Annals of Applied Statistics*, 14(2): 993–1029. MR4117838. doi: https://doi.org/10.1214/20-AOAS1339.   11

Peixoto, T. P. (2017). "Nonparametric Bayesian inference of the microcanonical stochastic block model." *Physical Review E*, 95(1): 012317. doi: https://doi.org/10.1103/PhysRevE.95.012317. 4

Peixoto, T. P. (2019). "Network reconstruction and community detection from dynamics." *Physical Review Letters*, 123(12): 128301. doi: https://doi.org/10.1103/PhysRevLett.123.128301. 2, 5

Peterson, C., Stingo, F. C., and Vannucci, M. (2015). "Bayesian inference of multiple gaussian graphical models." *Journal of the American Statistical Association*, 110(509): 159–174. MR3338494. doi: https://doi.org/10.1080/01621459.2014.896806. 10

Petralia, F., Rao, V., and Dunson, D. (2012). "Repulsive mixtures." In Pereira, F., Burges, C., Bottou, L., and Weinberger, K. (eds.), *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc. 20

Pitman, J. (2006). *Combinatorial Stochastic Processes: Ecole d'Eté de Probabilités de Saint-Flour XXXII – 2002*, volume 1875 of *Lecture Notes in Mathematics*. Springer, Berlin. Edited by Jean Picard. MR2245368. 4, 9

Quintana, F. A., Müller, P., Jara, A., and MacEachern, S. N. (2022). "The dependent Dirichlet process and related models." *Statistical Science*, 37(1): 24–41. MR4371095. doi: https://doi.org/10.1214/20-sts819. 3, 10

Raftery, A. E., Newton, M. A., Satagopan, J. M., and Krivitsky, P. N. (2007). "Estimating the integrated likelihood via posterior simulation using the harmonic mean identity." In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M. (eds.), *Bayesian Statistics 8*. Oxford University Press, London. MR2433201. 12

Rand, W. M. (1971). "Objective criteria for the evaluation of clustering methods." *Journal of the American Statistical Association*, 66(336): 846–850. 14, 15, 16

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabási, A.-L. (2002). "Hierarchical organization of modularity in metabolic networks." *Science*, 297(5586): 1551–1555. doi: https://doi.org/10.1126/science.1073374. 1, 2

Reyes, P. E. and Rodríguez, A. (2016). "Stochastic blockmodels for exchangeable collections of networks." arXiv:1606.05277v1. 4, 10, 11

Roverato, A. (2002). "Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models." *Scandinavian Journal of Statistics*, 29(3): 391–411. MR1925566. doi: https://doi.org/10.1111/1467-9469.00297. 5

Schmidt, M. N. and Mørup, M. (2013). "Nonparametric Bayesian modeling of complex networks: an introduction." *IEEE Signal Processing Magazine*, 30(3): 110–128. doi: https://doi.org/10.1109/MSP.2012.2235191. 4

Scott, J. G. and Carvalho, C. M. (2008). "Feature-inclusion stochastic search for Gaus-

sian graphical models." *Journal of Computational and Graphical Statistics*, 17(4): 790–808. MR2649067. doi: https://doi.org/10.1198/106186008X382683. 16

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors." *Statistica Sinica*, 4: 639–650. MR1309433. 6

Stanley, N., Shai, S., Taylor, D., and Mucha, P. J. (2016). "Clustering network layers with the strata multilayer stochastic block model." *IEEE Transactions on Network Science and Engineering*, 3(2): 95–105. MR3515211. doi: https://doi.org/10.1109/TNSE.2016.2537545. 11

Sun, S., Wang, H., and Xu, J. (2015). "Inferring block structure of graphical models in exponential families." In Lebanon, G. and Vishwanathan, S. V. N. (eds.), *Proc. of the 18th Int. Conf. on Artificial Intelligence and Statistics*, volume 38 of *Proc. of Machine Learning Research*, 939–947. PMLR. 5

Sun, S., Zhu, Y., and Xu, J. (2014). "Adaptive variable clustering in Gaussian graphical models." In Kaski, S. and Corander, J. (eds.), *Proc. of the 17th Int. Conf. on Artificial Intelligence and Statistics*, volume 33 of *Proc. of Machine Learning Research*, 931–939. PMLR. 4, 14, 15, 16, 21

Tan, L. S. and De Iorio, M. (2019). "Dynamic degree-corrected blockmodels for social networks: A nonparametric approach." *Statistical Modelling*, 19(4): 386–411. MR3984288. doi: https://doi.org/10.1177/1471082X18770760. 4, 6, 13, 14, 15

Tan, L. S. L., Jasra, A., De Iorio, M., and Ebbels, T. M. D. (2017). "Bayesian inference for multiple Gaussian graphical models with application to metabolic association networks." *Annals of Applied Statistics*, 11(4): 2222–2251. MR3743295. doi: https://doi.org/10.1214/17-AOAS1076. 10, 11

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). "Hierarchical Dirichlet processes." *Journal of the American Statistical Association*, 101(476): 1566–1581. MR2279480. doi: https://doi.org/10.1198/016214506000000302. 11

Telesca, D., Müller, P., Kornblau, S. M., Suchard, M. A., and Ji, Y. (2012). "Modeling protein expression and protein signaling pathways." *Journal of the American Statistical Association*, 107(500): 1372–1384. MR3036401. doi: https://doi.org/10.1080/01621459.2012.706121. 11

Todeschini, A., Miscouridou, X., and Caron, F. (2020). "Exchangeable random measures for sparse and modular graphs with overlapping communities." *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 82(2): 487–520. MR4084173. doi: https://doi.org/10.1111/rssb.12363. 4

van den Boom, W., Beskos, A., and De Iorio, M. (2022a). "The *G*-Wishart weighted proposal algorithm: Efficient posterior computation for Gaussian graphical models." *Journal of Computational and Graphical Statistics*. Advance online publication. doi: https://doi.org/10.1080/10618600.2022.2050250. 6, 9, 16

van den Boom, W., De Iorio, M., and Beskos, A. (2022b). "Supplementary Material for "Bayesian learning of graph substructures"." *Bayesian Analysis*. doi: https://doi.org/10.1214/22-BA1338SUPP. 4, 8, 9, 11, 13, 14, 16

Yook, S.-H., Oltvai, Z. N., and Barabási, A.-L. (2004). "Functional and topological characterization of protein interaction networks." *Proteomics*, 4(4): 928–942. doi: https://doi.org/10.1002/pmic.200300636. 1, 2

Yu, H., Paccanaro, A., Trifonov, V., and Gerstein, M. (2006). "Predicting interactions in protein networks by completing defective cliques." *Bioinformatics*, 22(7): 823–829. doi: https://doi.org/10.1093/bioinformatics/btl014. 8

Zachary, W. W. (1977). "An information flow model for conflict and fission in small groups." *Journal of Anthropological Research*, 33(4): 452–473. doi: https://doi.org/10.1086/jar.33.4.3629752. 13

Zhang, S. (2018). "Comparisons of gene coexpression network modules in breast cancer and ovarian cancer." *BMC Systems Biology*, 12(S1): 57–87. doi: https://doi.org/10.1186/s12918-018-0530-9. 2, 19

Zhou, S., Rütimann, P., Xu, M., and Bühlmann, P. (2011). "High-dimensional covariance estimation based on Gaussian graphical models." *Journal of Machine Learning Research*, 12: 2975–3026. MR2854354. 2, 12