

Learning block structured graphs in Gaussian graphical models

Alessandro Colombi^a, Raffaele Argiento^b, Lucia Paci^c, and Alessia Pini^c

^a Department of Economics, Management and Statistics, Università degli studi di Milano-Bicocca

^b Department of Economics, Università di Bergamo

^c Department of Statistical Sciences, Università Cattolica del Sacro Cuore

Abstract

Within the framework of Gaussian graphical models, a prior distribution for the underlying graph is introduced to induce a block structure in the adjacency matrix of the graph and learning relationships between fixed groups of variables. A novel sampling strategy named Double Reversible Jumps Markov chain Monte Carlo is developed for block structural learning, under the conjugate G-Wishart prior. The algorithm proposes moves that add or remove not just a single link but an entire group of edges. The method is then applied to smooth functional data. The classical smoothing procedure is improved by placing a graphical model on the basis expansion coefficients, providing an estimate of their conditional independence structure. Since the elements of a B-Spline basis have compact support, the independence structure is reflected on well-defined portions of the domain. A known partition of the functional domain is exploited to investigate relationships among the substances within the compound.

Keywords: Bayesian statistics; Conditional independence; functional data analysis; G-Wishart prior; Reversible jump MCMC

1 Introduction

Probabilistic graphical modeling is a powerful tool to study the dependence structure among a set of variables. The approach relies on the concept of conditional independence between variables, that is described through a map between a graph and a family of multivariate probability models. When such a family of probability distributions is chosen to be Gaussian, those models are known as Gaussian graphical models ([Lauritzen 1996](#)). Those models have been widely applied in many research fields to infer various types of networks, including in genomics where it is useful to understand how risk factors are related to each other for planning patient-specific therapies ([Dobra et al. 2004](#); [Jones et al. 2005](#); [Xia et al. 2018](#)).

Let \mathbf{Y} be a p -random vector distributed as a multivariate normal distribution with zero mean and precision matrix \mathbf{K} , i.e., $N_p(\mathbf{0}, \mathbf{K}^{-1})$. Under this normality assumption, the conditional independence relationship between variables can be represented in terms of the null elements of the precision matrix \mathbf{K} . Specifically, each node is associated to one of the variables of interest and the edges describe the structure of the non-zero elements of the precision matrix. The absence of a link between two vertices means that the two corresponding variables are conditionally independent, given all the others.

Usually, the structure of the underlying graph is unknown and needs to be estimated on the basis of the available data: this is referred to as (graph) structural learning. In a Bayesian framework, the specification of a prior on the graph space and, conditionally on the graph, a prior on the precision matrix is required. A common practice is to choose a discrete uniform distribution over the graph space \mathcal{G} , i.e., the space of all possible undirected graphs with p nodes. This is appealing for its simplicity but it is not a convenient choice to encourage sparsity, as it assigns most of the probability mass to graphs with a "medium" number of edges ([Jones et al. 2005](#)). As an alternative, a prior on the set of edges of the graph can be placed, to induce a prior over \mathcal{G} . In this setting, the most natural choice is to assume independent Bernoulli(θ_e) priors for each edge. The Bernoulli parameters θ_e may differ from edge to edge, but usually, a common value θ is assigned. For example, [Jones et al. \(2005\)](#) suggested to set $\theta = 2/(p - 1)$ to encourage sparsity in the graph; [Scott and](#)

[Carvalho \(2008\)](#) placed instead a Beta hyperprior on θ , a solution known as the multiplicity correction prior. Similarly, [Scutari \(2013\)](#) described a multivariate Bernoulli distribution where edges are not necessarily independent.

A common feature of the graph priors mentioned above is that they are non-informative, since the only type of prior believes they elicit in the model is the expected sparsity rate of the graph. In this work, instead, we develop a prior on the graph space that aims to be informative, according to prior information available for the application at hand. Since the graph describes the conditional dependence structure of variables involved in complex and, usually, high-dimensional phenomena, it is unrealistic to assume that prior knowledge is available for any one-to-one relationships between the observed quantities. Rather, it is more reasonable to envision that variables are grouped in smaller subsets. This is common in biological applications where the groups may be families of bacteria ([Osborne et al. 2021](#)), or genomics where groups of genes are known to be part of a common process ([Yook et al. 2004](#)). Also in market basket analysis products and customers can be easily grouped ([Giudici and Castelo 2003](#)).

Additionally, in some applications, the variables of interest may have a natural ordering, yielding to groups of nodes that are contiguous with respect to such ordering. In this case, if prior information is available about a possible partition of the nodes, the modeling assumptions must reflect that nodes cannot be re-labeled. For example, in spectrometric data analysis, the goal is usually to investigate relationships among the substances within the compound. To this end, a nonparametric regression coupled with a Gaussian graphical model on basis expansion coefficients can be employed, providing an estimate of their conditional independence structure ([Codazzi et al. 2022](#)). Since the elements of a B-Spline basis have compact support, the independence structure of the smoothing coefficients is reflected on portions of the spectrum, that is known a prior to be grouped in intervals of chemical interest. Here, the nodes representing the spline coefficients are naturally ordered and the fixed groups of nodes turn out to be contiguous in the functional domain. Thus, a prior on the graph should elicit such relevant features.

In this paper, we propose a class of prior distributions on the graph space that leverages

information on the groups of nodes and encodes a block structure in the adjacency matrix associated with the graph. Our approach consists of mapping the graph G into a block structured multigraph representation G_B , where blocks of links are represented by a single edge between two groups of nodes. Then, independent Bernoulli priors are assumed for the edges of the multigraph G_B . In other words, we allow nodes in different groups to be only fully connected or not connected at all. As a result, posterior learning aims at discovering the underlying pattern between groups of nodes, on the basis of the available data. We call this novel class of priors as *block graph priors*.

Bayesian posterior inference of the graph is usually performed through Markov chain Monte Carlo (MCMC) under the conjugate G-Wishart prior distribution on the precision matrix ([Roverato 2002](#); [Atay-Kayis and Massam 2005](#)). However, posterior computation is expensive for general non-decomposable graphs for two main reasons. Firstly, note that the cardinality of the graph space is $|\mathcal{G}| = 2^{\binom{p}{2}}$, i.e., it is large even if a moderate number of variables p is included. In practice, it can not be explicitly numerated but one needs to rely on search algorithms to explore it and learn what links should be included or not. However, even when the number of nodes is limited, it may be difficult to identify high posterior probability regions of the graph space.

Secondly, the challenge of developing efficient methods for structural learning is due to the presence of the G-Wishart prior distribution, which is defined conditionally on a graph G , and it is known only up to an intractable normalizing constant. Explicit formulas do exists for special cases such as complete or decomposable graphs ([Dawid and Lauritzen 1993](#)), which however are hard to justify from an applied side and increasingly restrictive as the number of nodes increases. [Uhler et al. \(2018\)](#) provided a recursive expression for the normalizing constant, but the procedure is computationally efficient only for some specific types of graphs. In practice, the normalizing constant is usually evaluated by means of numerical approximations such as an importance sampler ([Roverato 2002](#); [Dellaportas et al. 2003](#)), a Monte Carlo approximation ([Atay-Kayis and Massam 2005](#)), and a Laplace approximation ([Moghaddam et al. 2009](#); [Lenkoski and Dobra 2011](#)). Unfortunately, these methods become unstable with an increasing number of nodes (see [Jones et al. 2005](#) and

[Wang and Li 2012](#) for further details). Recent solutions have been proposed in the literature; for example [Wang et al. \(2015\)](#) leverages on the partial analytical structure of the G-Wishart distribution while [Mohammadi and Wit \(2015\)](#) rely on an approximation of the ratio of two normalizing constants arising when two models are compared. Also, [van den Boom et al. \(2022\)](#) used a delayed acceptance MCMC ([Christen and Fox 2005](#)) coupled with an informed proposal distribution ([Zanella 2020](#)) on the graph space to enable embarrassingly parallel computation. However, all these approaches are suited for comparing models whose graphs differ by a single edge and so they are inappropriate to address block structural learning. Rather, a MCMC method that modifies more than one link at a time is needed in our setting.

In this work, we introduce a Reversible Jump MCMC sampler ([Giudici and Green 1999](#); [Dobra et al. 2011](#)), defined over the joint space of the graph and the precision matrix, that leverages the structure induced by the block graph prior. In particular, we generalize the procedure of [Lenkoski \(2013\)](#) so that the algorithm modifies, at each step of the chain, an entire block of edges to guarantee a block structure of the adjacency matrix associated with the graph. Moreover, the Reversible Jump algorithm is coupled with the Exchange algorithm ([Murray et al. 2006](#)) to avoid the calculation of the G-Wishart normalizing constant, consisting of a second reversible move. We refer to the novel sampling method as the *Block Double Reversible Jump* (BDRJ) algorithm. As a result, the algorithm builds a Markov chain that visits only the subspace of block structured graphs, which is in general much smaller relative to the original graph space, and allows to infer the relationships among the fixed groups of nodes.

The remainder of the paper is organized as follows. Section 2 introduces the class of block structured graph priors and Section 3 provides the novel sampling strategy. In Section 4 we present a simulation study along with a comparison against an existing approach. Section 5 illustrates our method for functional data analysis. Finally, we conclude with a brief discussion in Section 6.

2 Block Structured Graph prior

2.1 From a graph to a block multigraph

Let \mathbf{Y} be a p -random vector distributed as a multivariate normal distribution with zero mean and precision matrix \mathbf{K} , i.e., $N_p(\mathbf{0}, \mathbf{K}^{-1})$; without loss of generality, we assume here \mathbf{Y} to be centered around zero. Let $G = (V, E)$ be an undirected graph, where $V = \{1, \dots, p\}$ is the set of p nodes and $E = \{(i, j) \mid i < j, i, j \in V\}$ is the set of undirected edges between the nodes. \mathbf{Y} is said to be Markov with respect to G if, for any edge (i, j) that does not belong to E , the i -th and j -th variables of \mathbf{Y} are conditionally independent given all the others, i.e., $Y_i \perp\!\!\!\perp Y_j \mid \mathbf{Y}_{-(ij)}$, where $\mathbf{Y}_{-(ij)}$ is the random vector containing all elements in \mathbf{Y} except the i -th and the j -th. Under the normality assumption, the conditional independence relationship between variables can be represented in terms of the null elements of the precision matrix \mathbf{K} . Specifically, each node is associated to one of the variables of interest, and its edges describe the structure of the non-zero elements of the precision matrix. The absence of a link between two nodes means that the two corresponding variables are conditionally independent, given all the others. Hence, the following equivalence provides an interpretation of the graph

$$Y_i \perp\!\!\!\perp Y_j \mid \mathbf{Y}_{-(ij)} \iff (i, j) \notin E \iff k_{ij} = 0,$$

where k_{ij} is the entry (i, j) of the precision matrix \mathbf{K} . The graph G is usually unknown and must be learned from the data. In a Bayesian framework it is considered to be a random variable having values in the space \mathcal{G} of all possible undirected graphs having p nodes. The cardinality of such a space is $|\mathcal{G}|=2^{\binom{p}{2}}$ since self-loops are not allowed.

The starting point for our proposed model is that we assume that the p observed variables are grouped in M mutually exclusive groups that are known a priori. Each group has cardinality n_i and $\sum_{i=1}^M n_i = p$. We admit the possibility of having some $n_i = 1$, as long as $M < p$. Groups whose cardinality is equal to one are called *singletons*.

In our setting, the groups are given and the underlying graph has to satisfy a precise block structure. To accomplish that, we define a new space of undirected graphs whose nodes represent the groups of variables and edges represent the relationships between them.

Namely, let $V_B = \{B_1, \dots, B_M\}$ be a partition of V in M groups that is available a priori. Then, we define $G_B = (V_B, E_B)$ to be an undirected graph whose nodes are the sets $B_k, k = 1, \dots, M$ and that allows for self-loops if $n_k > 1$. Namely, the set of edges E_B is given by

$$E_B = \left\{ (l, m) \mid l, m \in V_B, \wedge l < m, \quad (l, l) \mid l \in V_B, \wedge n_l > 1 \right\}.$$

In graph theory, graphs that have self-loops are called *multigraphs*. We denote by \mathcal{G}_B the set of all possible multigraphs G_B having V_B as set of nodes.

To clarify the relationship between \mathcal{G}_B and \mathcal{G} , consider $G_B \in \mathcal{G}_B$ and $G \in \mathcal{G}$. By definition, the set of nodes of the multigraph G_B is obtained by grouping together the nodes of the graph G . What about the set of edges? Is there any relation between the two sets? The following map defines a relationship between them. Let $\rho : \mathcal{G}_B \rightarrow \mathcal{G}$, such that $G_B = (V_B, E_B) \mapsto G = (V, E)$ by the following transformations

$$\begin{aligned} V &= \{B_{l,h}, l = 1, \dots, n_h, h = 1, \dots, M\} = \{1, \dots, p\} \\ \text{if } (l, m) \in E_B &\Rightarrow (i, j) \in E \quad \forall i \in B_l, \forall j \in B_m \\ \text{if } (l, m) \notin E_B &\Rightarrow (i, j) \notin E \quad \forall i \in B_l, \forall j \in B_m \end{aligned} \tag{1}$$

A visual representation of this mapping is given in Figure 1. Once ρ is set we are able to associate each G_B in \mathcal{G}_B to one and only one G in \mathcal{G} , since ρ is clearly injective. We refer to G_B as the multigraph form of G .

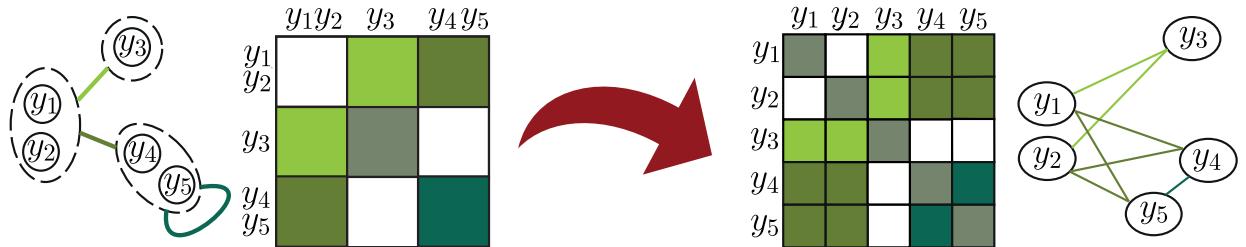


Figure 1: The map from multigraph $G_B \in \mathcal{G}_B$ (left panel) to its block structured form $G \in \mathcal{B}$ (right panel).

Nevertheless, ρ is not surjective which implies that there are graphs G that do not have a representative in \mathcal{G}_B . Indeed, only those graphs whose adjacency matrix have a particular block structure can be represented in a multigraph form. A non surjective map is the key

ingredient to define a subset of \mathcal{G} of block structured graphs that satisfy the modelling assumptions. Let \mathcal{B} be the image of ρ , i.e., the subset of \mathcal{G} containing all the graphs having p nodes and a block structure consistent with V_B . Moreover, $\rho : \mathcal{G}_B \rightarrow \mathcal{B}$ is a bijection, which means that every graph $G \in \mathcal{B}$ is associated to its representative $G_B \in \mathcal{G}_B$ via ρ^{-1} . We say that $G \in \mathcal{B}$ is the block graph representation of the multigraph $G_B \in \mathcal{G}_B$. This representation of block graphs allows us to work in a space where we can use standard tools of graphical analysis. In a different setting, [Cremaschi et al. \(2021\)](#) adopts a similar approach to model the conditional dependence across Markov processes.

2.2 Prior on the graph space

The map described in Section 2.1 allows us to introduce a class of priors \mathcal{G} that encodes the knowledge about the partition of the nodes. A customary choice in the literature is to assume independent Bernoulli prior to links in \mathcal{E} , the set of all possible edges. Such assumption seems reasonable only if one may assume a priori independence between the edges. Nevertheless, to induce a block structure in the adjacency matrix of the graph, independent Bernoulli priors for the edges would not be appropriate.

The class of priors we propose is built on two assumptions: (i) the nodes have a natural ordering and their groups are known a priori; (ii) a zero mass probability must be placed to those graphs in \mathcal{G} where nodes in the same groups are not fully connected or not connected at all, i.e., to all graphs that belongs to $\mathcal{G} \setminus \mathcal{B}$. Based on these assumptions, we need to specify the probability of graphs that are in \mathcal{B} , which can be represented through a multigraph $G_B \in \mathcal{G}_B$ using Equation (1). Each graph in \mathcal{G}_B can be thought as a graph having M nodes that may have self-loops, therefore it is reasonable to assume that its links are independent and that a prior on \mathcal{G}_B can be defined as the one induced by assigning independent Bernoulli prior to its links. Finally, the prior probability of each graph in \mathcal{B} is set to be equal to the probability of its representative in \mathcal{G}_B , which can be obtained by applying the ρ^{-1} map.

Namely,

$$\pi(G) \propto \begin{cases} \pi_B(\rho^{-1}(G)), & \text{if } G \in \mathcal{B}, \\ 0, & \text{if } G \in \mathcal{G} \setminus \mathcal{B}, \end{cases} \quad (2)$$

where $\pi_B(\rho^{-1}(G)) = \pi_B(G_B) = \theta^{|E_B|}(1-\theta)^{\binom{M}{2}-|E_B|}$ is the Bernoulli prior over the set \mathcal{G}_B , where each link has prior probability of inclusion θ , which is fixed a priori. We refer to (2) as *block-Bernoulli prior*. In particular, (2) is the composition of two maps: $\pi_B(\cdot)$, that is a prior on the space \mathcal{G}_B and ρ^{-1} that is deterministic. This implies that the same reasoning is valid even if $\pi_B(\cdot)$ is replaced by any other prior distribution available in the literature of graphical models. The only constrain is that it must be a probability distribution over \mathcal{G}_B , not over \mathcal{G} , as it is usually done. We define the resulting class of priors to as the *block graph priors*.

3 Block Double Reversible Jump algorithm

A popular choice as prior for the precision matrix \mathbf{K} , conditional on the graph G , is the G-Wishart distribution, introduced by [Roverato \(2002\)](#) to deal with non-decomposable graphs. Following [Mohammadi and Wit \(2015\)](#), we work with a Shape-Inverse Scale parametrization of the G-Wishart distribution, that is, we say that $\mathbf{K} | G \sim \text{G-Wishart}(b, D)$ if its density is given by

$$P(\mathbf{K} | G) = I_G(b, D)^{-1} |\mathbf{K}|^{\frac{b-2}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{K}D)\right\} \mathbb{1}_{\mathbb{P}_G},$$

where $b > 2$ is the shape parameter, the inverse scale matrix D is symmetric and positive definite and \mathbb{P}_G is the space of all $p \times p$ symmetric and positive definite matrices whose null elements are associated to links absent in graph G , i.e., that are Markov with respect to G . Finally,

$$I_G(b, D) = \int_{\mathbb{P}_G} |\mathbf{K}|^{\frac{b-2}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{K}D)\right\} d\mathbf{K},$$

is the normalizing constant. In this work, b and D are fixed hyperparameters. Let $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ be a iid sample of size n from a $\text{N}_p(\mathbf{0}, \mathbf{K}^{-1})$, where the precision matrix \mathbf{K} is Markov with respect to the graph G . Thanks to conjugacy, the full conditional distribution of \mathbf{K} is $\mathbf{K} | G, \mathbf{y} \sim \text{G-Wishart}(b+n, D+U)$, where $U = \mathbf{y}^\top \mathbf{y}$.

The goal of the Bayesian structural learning is to compute the posterior distribution

$$P(G|\mathbf{y}) \propto \int_{\mathbb{P}_G} p(\mathbf{y}|\mathbf{K})p(\mathbf{K}|G)\pi(G)d\mathbf{K} \propto \pi(G) \frac{I_G(d+n, D+U)}{I_G(d, D)},$$

which is proportional to the ratio of the posterior and the prior G-Wishart normalizing constant. As anticipated in the Introduction, posterior inference with the G-Wishart distribution is challenging since the joint posterior distribution of the graph and the precision matrix is doubly-intractable ([Murray et al. 2006](#)). Indeed, the normalizing constant $I_G(b, D)$ does not have a simple analytical form for general non-decomposable graphs, making the computation of a Metropolis-Hastings acceptance probability not feasible. To address this issue, Monte Carlo ([Atay-Kayis and Massam 2005](#)) and Laplace approximations ([Moghaddam et al. 2009; Lenkoski and Dobra 2011](#)) have been introduced. Alternatively, [Mohammadi et al. \(2021\)](#) proposed an approximation of the ratio $I_G(b, D)/I_{G'}(b, D)$, which is, in practice, the quantity required in the computation of the acceptance-rejection probability of a proposed graph G' . However, all the existing methods are not able to exploit prior information about the grouped structure of the graph since they are suited to modify only one edge at each step of the MCMC algorithm. Rather, to ensure a block structure of the graph compatible with our prior believes, links can not be modified at will, at least, not in the space \mathcal{B} . To our knowledge, there are no solutions available in the literature to compute the G-Wishart normalizing constant ratio in our setting.

For this reason, we move a step forward and develop a Block Reversible Jump Monte Carlo Markov Chain sampler ([Giudici and Green 1999; Dobra et al. 2011](#)) defined over the joint space of graph and precision matrix. It generalizes the procedure by [Lenkoski \(2013\)](#) in such a way that it modifies, at each step of the chain, an entire block of edges to guarantee that the visited graphs always belong to the space of block structured graphs \mathcal{B} . By doing so, the search is limited to the subset of graphs whose structure is consistent with V_B . In particular, the algorithm is built upon a trans-dimensional version of the Exchange algorithm ([Murray et al. 2006](#)) that allows to avoid any calculation of the G-Wishart normalizing constant.

We denote the current state of the chain by $(\mathbf{K}^{[s]}, G^{[s]})$, where $\mathbf{K}^{[s]} \in \mathbb{P}_{G^{[s]}}$. Since the graph is constraining the support of the precision matrix, the Reversible Jump technique

is needed to handle trans-dimensional moves due to a different number of unknown entries of the precision matrix in subsequent iterations. Each step of the algorithm is made of two iterative steps. In the first step, the state (\mathbf{K}', G') is proposed and the graph G' is accepted or rejected. In particular, the first step consists of three parts: (i) a new graph G' is proposed in the neighbourhood of the current graph $G^{[s]}$; (ii) a matrix \mathbf{K}' compatible with the constraints imposed by G' is constructed. Note that, one must guarantee that the proposed matrix \mathbf{K}' is a precision matrix. Differently from [Giudici and Green \(1999\)](#), who proposed a Reversible Jump sampler that limits itself to visit decomposable graphs and requires to check the positive definiteness of the precision matrix, we follow [Dobra et al. \(2011\)](#) and [Lenkoski \(2013\)](#) that exploit a re-parametrization of the precision matrix based on its Cholesky decomposition, so that matrices are positive definite by construction. Finally, (iii) the acceptance-rejection rate is computed by exploiting a modified Exchange Algorithm that avoids the calculation of the G-Wishart normalizing constant. The algorithm requires a double reversible move, leading to a Double Reversible Jump procedure. In the following, each of these three parts is described.

The second MCMC step is to sample the precision matrix $\mathbf{K}^{[s+1]}$ from its full conditional, i.e., $\mathbf{K}^{[s+1]} \mid G^{[s]}, \mathbf{y} \sim \text{G-Wishart}(b + n, D + U)$. The resulting algorithm is given in Algorithm 1, with the details on the derivation of the acceptance-rejection probability deferred to Appendix A. The R package `BGSL`, implementing Algorithm 1, is available at the link github.com/alessandrocolombi/BGSL.

(i) Propose a new graph G'

A common factor in most of the existing MCMC methods for graphical models is to set up chains such that the proposed graph $G' = (V, E')$ belongs to the one-edge-away neighbourhood of G , which is defined as

$$nbd_p(G) := nbd_p^+(G) \cup nbd_p^-(G), \quad (3)$$

where $nbd_p^+(G)$ and $nbd_p^-(G)$ are the sets of undirected graphs having p nodes that can be obtained by adding, or removing, an edge to $G \in \mathcal{G}$, respectively. A step in the Markov chain that selects $G' \in nbd_p(G^{[s]})$ is said to be a local move. On the other hand, the

proposed BDRJ approach is innovative because it propose moves that modifies an entire block of links, not just a single one. In other words, our moves are local in \mathcal{G}_B but not in \mathcal{G} .

Our procedure leverages on the definition of the ρ function and the generalization of standard graphical models tools to the space \mathcal{B} . Hence, suppose $G^{[s]} \in \mathcal{B}$ and we aim to construct a new graph $G' \in \mathcal{B}$. Firstly, we map the current graph into its multigraph representative $G_B^{[s]} \in \mathcal{G}_B$. Then, with probability α_G , we add a new link or, with probability $(1 - \alpha_G)$, we remove one of its existing links. Namely, the new multigraph representation $G'_B \in \mathcal{G}_B$ is drawn from

$$q\left(G'_B | G^{[s]}\right) = \alpha_G \text{Unif}\left(nbd_M^{\mathcal{B},+}\left(\rho^{-1}\left(G^{[s]}\right)\right)\right) + (1 - \alpha_G) \text{Unif}\left(nbd_M^{\mathcal{B},-}\left(\rho^{-1}\left(G^{[s]}\right)\right)\right), \quad (4)$$

where, similarly to (3), $nbd_M^{\mathcal{B}}\left(G_B^{[s]}\right)$ is the one-edge-away neighbourhood of $G_B^{[s]} = \rho^{-1}\left(G^{[s]}\right)$ with respect to the space of multigraphs \mathcal{G}_B . Finally ρ is applied once again to map the resulting multigraph back in \mathcal{B} to obtain G' , i.e., we set $G' = \rho\left(G'_B\right)$.

If $\alpha_G = 0.5$, (4) gives equal probability to addition and delition moves. To lighten the notation, we always refer to this case. The proposal distribution (4) is preferred with respect to simply chose uniformly in the whole neighbourhood as in [Madigan and York \(1995\)](#). Indeed, [Dobra et al. \(2011\)](#) noticed that in a simple uniform edge sampling, the probability of proposing a move that adds (or deletes) an edge is too small if the current graph has a very large (or small) number of edges. Therefore, (4) guarantees a better mixing in the resulting Markov chain. A closer look at (4) reveals how our multigraph representation allows us to use standard tools of structural learning in the space \mathcal{G}_B to get non-standard proposal in the original space \mathcal{G} .

(ii) Construct the precision matrix $\mathbf{K}' | G'$

Once that the graph is selected, we need to specify a method to construct a proposed precision matrix \mathbf{K}' that satisfies the constraints imposed by the new graph. In principle, the method by [Wang and Li \(2012\)](#) based on the partial analytical structure of the G-Wishart appears to be an efficient choice. However, this solution strongly relies on the possibility of writing down an explicit formula for the full conditional of the elements of \mathbf{K} . Such

results, presented in [Roverato \(2002\)](#), can be handled in practice only if at each step of the algorithm only one link of the graph is modified. Instead, the proposal distribution presented in Section 3 modifies an arbitrary number of links. Differently, we build on a generalization of the Reversible Jump mechanism by [Lenkoski \(2013\)](#). The idea is that it is possible to guarantee the positive definiteness of \mathbf{K}' and the zero constraints imposed by \mathbf{G}' by working on the Cholesky decomposition of $\mathbf{K}^{[s]}$.

Indeed, $\mathbf{K}^{[s]} \in \mathbb{P}_{G^{[s]}}$ implies that it is possible to compute its Cholesky decomposition, $\mathbf{K}^{[s]} = (\boldsymbol{\Phi}^{[s]})^T \boldsymbol{\Phi}^{[s]}$, where $\boldsymbol{\Phi}^{[s]}$ is an upper triangular matrix. This is appealing because the zero constraints imposed by $\mathbf{G}^{[s]}$ on the off-diagonal elements of $\mathbf{K}^{[s]}$ induce a precise structure and properties on $\boldsymbol{\Phi}^{[s]}$. In particular, let $v(G^{[s]}) = \{(i, j) \mid i, j \in V, i = j \vee (i, j) \in E^{[s]}\}$ be the set of the diagonal elements and the links belonging to $\mathbf{G}^{[s]}$. Hence, $\boldsymbol{\Phi}^{v(G^{[s]})} = \{\boldsymbol{\Phi}_{ij} \mid i, j \in v(G^{[s]})\}$ is said to be the set of *free elements* of $\boldsymbol{\Phi}^{[s]}$. The remaining entries are uniquely determined through the completion operation ([Atay-Kayis and Massam 2005](#), Proposition 2) as a function of the free elements. We refer to these as *non-free elements*. See [Roverato \(2002\)](#) and [Atay-Kayis and Massam \(2005\)](#) for an exhaustive overview.

Suppose the proposed graph \mathbf{G}' is obtained from (4) by adding edge (l, m) to the multigraph representation of $\mathbf{G}^{[s]}$. The set of links that are changing in \mathcal{G} is then $L = \{(i, j) \mid i, j \in V, i < j, (i, j) \in E', (i, j) \notin E^{[s]}\}$. The cardinality $|L|$ is arbitrary and, in general, different from one. We call $V(L) = B_l \cup B_m$ the set of the vertices involved in the change.

Note that $v(\mathbf{G}') = v(G^{[s]}) \cup L$. Our solution to define the new free elements is to maintain the same value for all the ones that are not involved in the change and to set the new ones by perturbing the current, non free elements, independently and all with the same variance σ_g^2 . Namely, draw $\eta_h \stackrel{\text{ind}}{\sim} N(\boldsymbol{\Phi}_h^{[s]}, \sigma_g^2)$ and set $\boldsymbol{\Phi}'_h = \eta_h$ for each $h \in L$. Then, all non free elements of $\boldsymbol{\Phi}'$ are derived through completion operation ([Atay-Kayis and Massam 2005](#)) and the proposed precision matrix is $\mathbf{K}' = (\boldsymbol{\Phi}')^T \boldsymbol{\Phi}'$ is then obtained. Note that, we are generating a random variable $\boldsymbol{\eta}$ of length $|L|$ that matches the dimension gap between $\mathbf{K}^{[s]}$ and \mathbf{K}' . In case of dimension reduction, the move is deterministic since can be seen as an increasing move with a specific choice for the proposal vector. The acceptance rate, than,

is the reciprocal of the corresponding increasing move.

(iii) Compute the acceptance-rejection rate

Finally, we frame the previous mechanism in the Exchange algorithm (Murray et al. 2006) paradigm to eliminate the presence of the G-Wishart normalizing constants. Specifically, we employ the Double Reversible Jump procedure (Lenkoski 2013), that is the trans-dimensional equivalent of the Exchange algorithm.

Let $\widetilde{\mathbf{W}}$ a latent $p \times p$ symmetric and positive definite matrix that is Markov with respect to G' , i.e., $\widetilde{\mathbf{W}} \in \mathbb{P}_{G'}$. The augmented target is the joint distribution

$$p(\mathbf{K}, G, \widetilde{\mathbf{W}}, G' | \mathbf{y}) \propto p(\mathbf{K}, G | \mathbf{y}) p(\widetilde{\mathbf{W}} | G') q(G' | G),$$

so that the marginal density $p(\mathbf{K}, G | \mathbf{y})$ remains unchanged. Hence, the BDRJ considers switching between $(\mathbf{K}, G, \widetilde{\mathbf{W}}, G')$ to the alternative $(\mathbf{K}', G', \mathbf{W}^0, G)$, with $\mathbf{W}^0 \in \mathbb{P}_{G^{[s]}}$, by performing two Reversible Jump moves; a dimension increasing step from (\mathbf{K}, G) to (\mathbf{K}', G') according to the posterior parameters $b + n$ and $D + U$ of the G-Wishart distribution and a dimension decreasing jump from $(\widetilde{\mathbf{W}}, G')$ to (\mathbf{W}^0, G) according to the prior parameters b and D of the G-Wishart distribution. Thus, the normalizing constant ratio in the acceptance-rejection rate cancels out and the proposed graph G' is accepted with probability $\min(1, R^+)$, where

$$R^+ \propto \frac{p(\mathbf{K}' | G')}{p(\mathbf{K}^{[s]} | G)} \frac{p(\mathbf{W}^0 | G^{[s]})}{p(\widetilde{\mathbf{W}} | G')}.$$

The details are reported in Appendix A.

Our proposed method BDRJ borrows the skeleton of the Double Reversible Jump but modifies the proposal distribution to guarantee that $G' \in \mathcal{B}$ and, as a consequence, that multiple elements of the precision matrix are updated accordingly. Since only proposal distributions and priors have been changed, we still end up having a valid MCMC scheme which is now able to infer relationships in block structured graphs.

Algorithm 1: Block Double Reversible Jump

Suppose the chain to be in state $(\mathbf{K}^{[s]}, G^{[s]})$, with $\mathbf{K}^{[s]} \in P_{G^{[s]}}$ and $G^{[s]} \in \mathcal{B}$. For each iteration:

Step 1. Updating the graph G

- 1.1. Sample G'_B from $q(G'_B | G^{[s]})$ given by (4). Set $G' = \rho^{-1}(G^{[s]})$. Suppose an addition move is selected. Call L the set of new edges.
- 1.2. Draw $\tilde{\mathbf{W}} | G' \sim \text{G-Wishart}(b, D)$ from an exact sampler (Lenkoski 2013).
- 1.3. For each $h \in L$, draw $\eta_h \sim N(\Phi_h^{[s]}, \sigma_g^2)$
- 1.4. Set $(\Phi')^{\nu(G^{[s]})} = (\Phi^{[s]})^{\nu(G^{[s]})}$ and $\Phi'_h = \eta_h \quad \forall h \in L$.

Derive the remaining elements by completion operation and define

$$\mathbf{K}' = (\Phi')^T \Phi'.$$

- 1.5. Set $(\Phi^0)^{\nu(G^{[s]})} = (\tilde{\Phi})^{\nu(G^{[s]})}$. Derive the remaining elements by completion operation and define $\mathbf{W}^0 = (\Phi^0)^T \Phi^0$.
- 1.6. Compute $\gamma((\mathbf{K}^{[s]}, G^{[s]}) \rightarrow (\mathbf{K}', G')) = \min\{1, R^+\}$ where

$$R^+ = \frac{\exp\left\{-\frac{1}{2}\langle \mathbf{K}' - \mathbf{K}^{[s]}, D + U \rangle\right\}}{\exp\left\{-\frac{1}{2}\langle \tilde{\mathbf{W}} - \mathbf{W}^0, D \rangle\right\}} \frac{\pi(G')}{\pi(G^{[s]})} \prod_{i \in V(L)} \left(\frac{\Phi_{ii}^{[s]}}{\Phi_{ii}^0} \right)^{\nu_i^{G'} - \nu_i^{G^{[s]}}} \exp\left\{ \frac{1}{2\sigma_g^2} \sum_{h \in L} \left[(\Phi'_h - \Phi_h^{[s]})^2 - (\Phi_h^0 - \tilde{\Phi}_h)^2 \right] \right\},$$

where $\langle A, B \rangle$ denotes the trace of the product between A and B .

- 1.7. Draw $c \sim \text{Unif}[0, 1]$. if $c < \gamma$ then set $G^{[s+1]} = G'$.

Step 2. Updating the precision matrix \mathbf{K}

Draw $\mathbf{K}^{[s+1]} | G^{[s+1]}, \mathbf{y} \sim \text{G-Wishart}(b + n, D + U)$.

3.1 Posterior Inference

Posterior inference of the graph has to be performed with some care. In the ideal case, we would like to approximate its posterior distribution with the relative frequency of visits for each sampled graph. One way of providing a pointwise estimate of the graph structure is to select the maximum a posteriori strategy, which represents the mode of their posterior distribution. Unfortunately, as noticed in [Jones et al. \(2005\)](#), for problems with even a moderate number of nodes p , the space to be explored is so large that the graph frequency can not be viewed as a solid estimate of its posterior probability because the space of all graphs is so large that each particular graph may be encountered only a few times in the course of the MCMC sampling ([Peterson et al. 2015](#)).

A more practical and stable solution is instead to estimate the posterior edges inclusion marginally. Let S be the size of a MCMC output, then the posterior inclusion probabilities are estimated as

$$\hat{p}_{ij} = \frac{1}{S} \sum_{s=1}^S \mathbb{1}\left((i, j) \in E^{[s]}\right) \quad (5)$$

where $\mathbb{1}\left((i, j) \in E^{[s]}\right)$ is the indicator function representing the inclusion of the edge linking nodes i and j in the graph $G^{[s]} = (V, E^{[s]})$ visited during the s -th iteration. We call $\hat{\mathbf{P}}$ the upper triangular matrix having elements \hat{p}_{ij} , for $i = 1, \dots, p$ and $j = i, \dots, p$, which in practice are the proportion of the MCMC iteration after the burnin period in which the edge (i, j) was selected to be part of the graph. Since it contains the posterior probabilities of inclusion, this matrix represents our uncertainty about including or not a edge in the graph.

Pointwise graphical estimate \hat{G} is carried out by selecting all edges whose posterior inclusion probability in (5) exceeds a given threshold τ . A possible choice is $\tau = 0.5$, in analogy with the median probability model of [Barbieri and Berger \(2004\)](#), originally proposed in the linear regression setting. A second possibility is based on the Bayesian False Discovery Rate (BFDR; [Müller et al. 2007](#); [Peterson et al. 2015](#))

$$\text{BFDR} = \frac{\sum_{i < j} (1 - \hat{p}_{ij}) \mathbb{1}(\hat{p}_{ij} \geq s)}{\sum_{i < j} \mathbb{1}(\hat{p}_{ij} \geq s)}, \quad (6)$$

where τ is selected so that (6) is below 0.05.

For what concerns the precision matrix, we average over the MCMC samples ([Cremaschi et al. 2019](#); [Wang and Li 2012](#)) to obtain the posterior mean $\widehat{\mathbf{K}}$

$$\widehat{\mathbf{K}} = \frac{1}{S} \sum_{s=1}^S \mathbf{K}^{[s]}.$$

Note that, even if each $\mathbf{K}^{[s]}$ has a block structure induced by $G^{[s]}$, $\widehat{\mathbf{K}}$ does not share the same structure of the selected graph \widehat{G} .

4 Simulation Study

We carry out a simulation study to evaluate the ability of our methodology to recover the structure of the generating graph. We compare our performance to the Birth and Death approach (BDgraph for short) proposed by [Mohammadi and Wit \(2015\)](#) for Gaussian graphical models with a standard non-informative prior on the graph space. In the latter, authors derived an efficient MCMC method where moves are decided according to specific birth and death transition kernels for the edges. Every proposed move is accepted and so the chain reaches convergence very quickly. Moreover, the usage of an approximation of the G-Wishart normalizing constants ratio approximation ([Mohammadi et al. 2021](#)) allows to speed up calculations. Such a method is available in the R package BDgraph ([Mohammadi and Wit 2019](#)). Graph posterior estimates from both BDRJ and BDgraph approaches have been obtained by cutting the posterior probability of inclusion of each link with the threshold chosen via the BDPR in (6).

4.1 Performance evaluation

To assess the performance of recovering the graph structure, we compute the standardized Structural Hamming Distance (Std-SHD, [Tsamardinos et al. 2006](#)) from the underlying graph, which is, in the case of undirected graphs, equal to the number of wrongly estimated edges, standardized with respect to the number of all possible ones, i.e.,

$$\text{Std-SHD} = \frac{\text{FP} + \text{FN}}{\binom{p}{2}}, \quad (7)$$

where FP and FN are the number of false positives and false negatives, respectively. Inspired by [Mohammadi and Wit \(2015\)](#), [Kumar et al. \(2020\)](#) and [Osborne et al. \(2021\)](#), we also take in consideration the F_1 -score ([Baldi et al. 2000](#); [Powers 2011](#)), which is defined as

$$F_1\text{-score} = \frac{2TP}{2TP + FP + FN} \quad (8)$$

where TP, TN are the number of true positives and true negatives, respectively. Both indices lie between 0 and 1: for the Std-SHD lower values are preferred (0 value stands for perfect match), while for F_1 -score higher values correspond to better performances (1 value stands for perfect match). The main difference between the two indices is that (7) equally weights errors due to false positiveness or negativeness while (8) places higher importance on the number of correct discoveries made, that are the true positives. To visualize their difference, consider the following simple example: set the true graph to have a sparsity index equal to 0.1 and consider a trivial estimator, i.e., the empty graph. The resulting Std-SHD score would be equal to 0.1 that seems reasonably good even if the estimated graph is not capturing any significant information. On the other hand, the F_1 -score score is not deceived as it would be equal to 0.

4.2 Results

We consider two different scenarios to compare the ability of BDRJ and BDgraph to learn the structure of conditional independencies. In the first one, we generate data using an underlying graph whose adjacency matrix has a block structure, i.e., with fixed groups of nodes. In the second experiment we investigate the performance of the block graph prior model when the true graph has incomplete blocks and some isolated links.

Experiment 1 - Complete Blocks

We set $n = 500$, $p = 40$ and $M = p/2$ groups of equal size, which leads to off-diagonal blocks of size 2×2 . The blocked underlying graphs have been randomly generated by sampling from $\pi(G \mid \theta)$ with different sparsity indices θ , uniformly distributed in the interval $[0.2, 0.6]$. Given the graph, the true precision matrix has been sampled from a

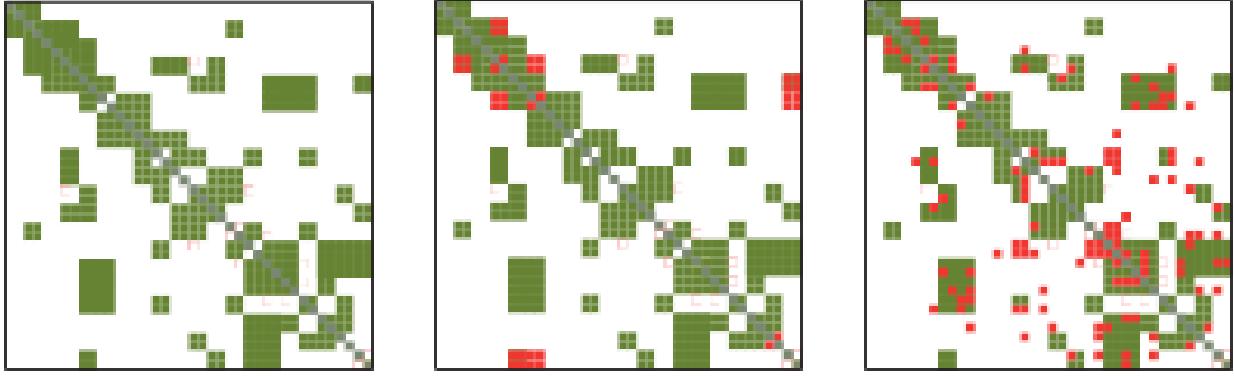


Figure 2: The true graph (left panel) and the estimated graph obtained using BDRJ (middle panel) and BDgraph (right panel) for Experiment 1. Green squares represent the included edges while red squares stand for edges that are wrongly estimated.

G-Wishart $(3, \mathbf{I}_p)$. For this study, σ_g^2 was set equal to 0.5, after a little tuning phase. MCMC sample is composed of 400,000 iterations plus 100,000 extra iteration as a burn-in period that were discarded. Figure 2 shows an example of the true graph and compares the final estimates of the forecited methods; green squares mean that there is an edge between the corresponding nodes. The middle and the right panels display the graph estimated by BDRJ and by BDgraph, respectively; red squares represent edges that are wrongly estimated. Clearly, the visual inspection suggests that BDRJ provides a better estimate of the underlying graph than BDgraph.

Figure 3 shows the boxplot of the Std-SHD and F_1 -scores over 15 replicates. We note that, overall, BDRJ outperforms BDgraph in terms of both indices. The number of misclassified edges is rather low for BDRJ, with median value of Std-SHD equal to 0.0455. Many true discoveries are achieved and indeed the median F_1 -score under BDRJ is 0.869. BDgraph performs worse with respect to both indices; the median Std-SHD is equal to 0.0622 while the F_1 -score is 0.807. The reason for these differences is that BDgraph does not recognize the block structure of the true graph, because of the lack of such prior information in the model. Rather, it tries to estimate every possible link independently from each other. This entails more errors in the final estimate as well as a less interpretable structure of the graph. It would be hard to explain why there are missing edges within some structures that are clearly grouped.

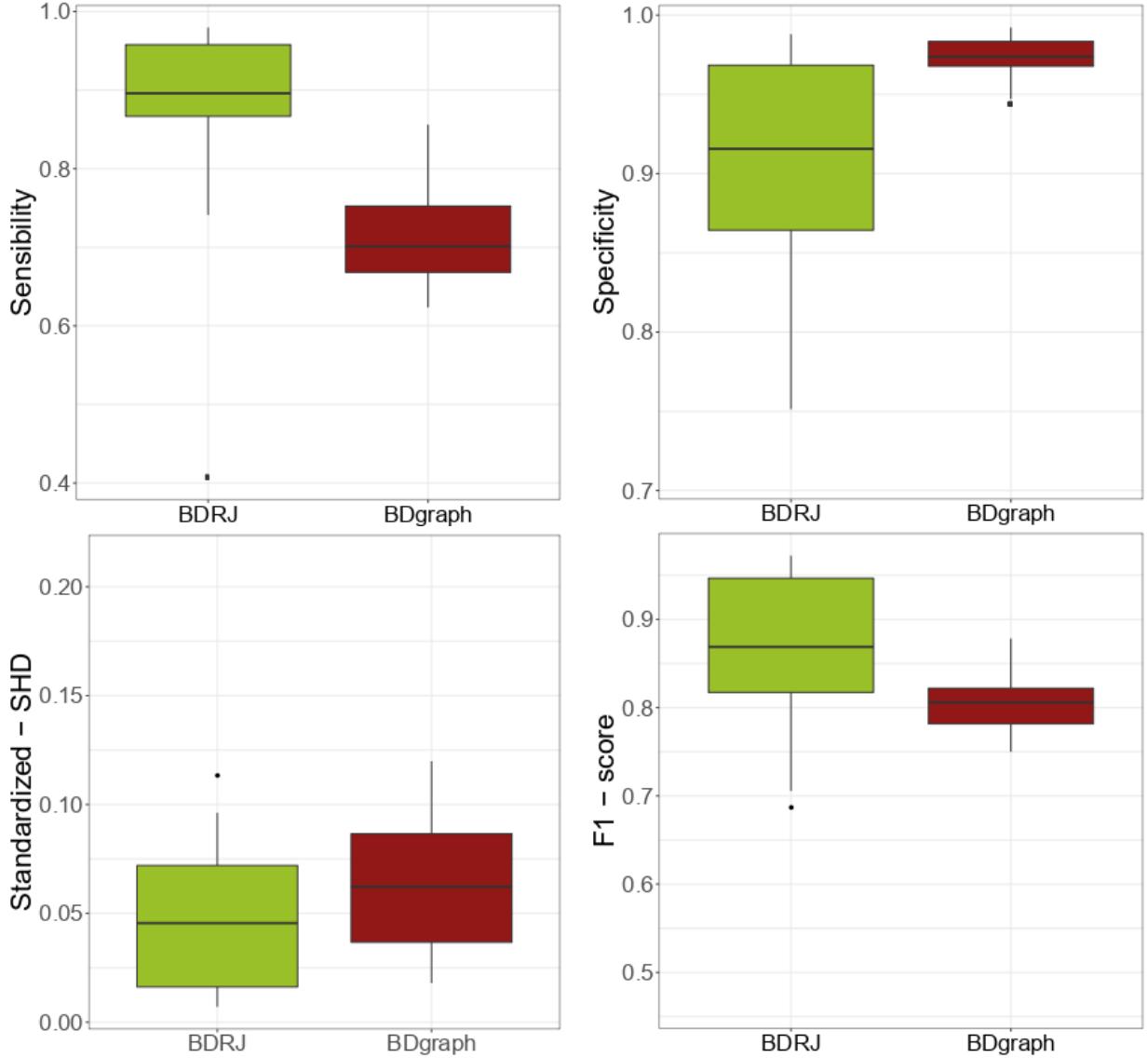


Figure 3: Boxplot of the Std-SHD (left panel), F₁-score (middle panel) and SL (right panel) over the replicates of Experiment 1.

Experiment 2 - Incomplete blocks

In this experiment we analyze the performance of our model when the underlying graph has incomplete blocks. To simulate the data under this scenario, we first draw a block structured graph from $\pi(G \mid \theta)$, where $\theta = 0.2$. Then, within each block, edges are removed with probability equal to 0.25. By doing so, blocks are incomplete but graphs maintains a meaningful structure. We set $n = 500$, $p = 30$, and $M = p/2$ groups of equal size; given the

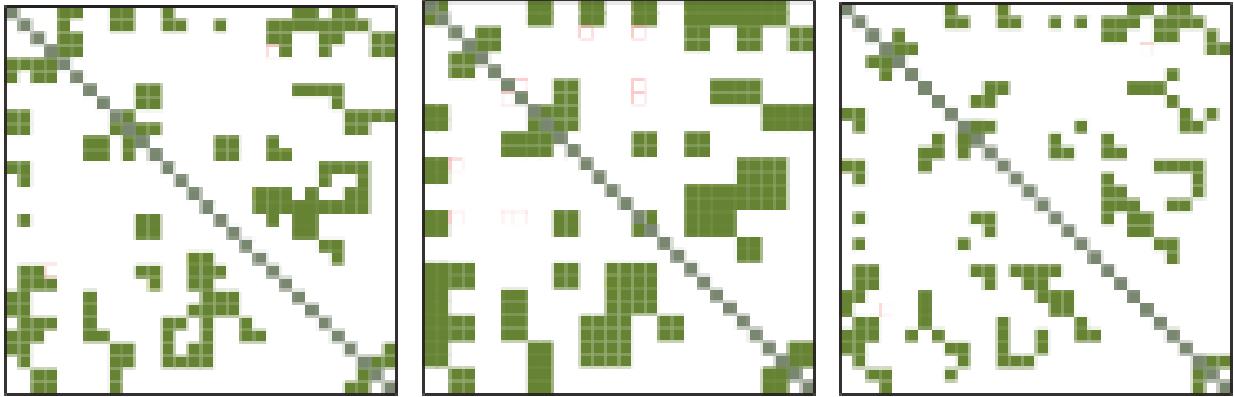


Figure 4: The true graph (left panel) and the estimated graph obtained using BDRJ (middle panel) and BDgraph (right panel) for Experiment 2. Green squares represent the included edges.

graph, the true precision matrix has been sampled from $\text{G-Wishart}(3, \mathbf{I}_p)$.

Figure 4 shows an example of the true graph (left panel) and the estimated one using BDRJ (middle panel) and BDgraph (right panel). A simple visual inspection of the figure suggests that our approach tends to include incomplete blocks rather than to discard them, leading to some false discoveries, as expected. If we had observed the opposite, that is the inclusion of a block only if it is complete, we would have ended up with an almost empty graph, i.e., no discoveries would have been done.

On the other hand, the BDgraph method does not make any assumption on the graph structure, so in principle, it should be able to recover the graph correctly. In practice, as soon as the dimension of the graph increases, this unlikely happens. The tendency here is instead to be more conservative, which leads to a lower number of false discoveries.

To clarify, we report in Figure 5 the boxplots of sensitivity index $\text{TP}/(\text{TP} + \text{FN})$ and specificity index $\text{TN}/(\text{TN} + \text{FP})$ computed over 15 simulated datasets. We see that our approach is always better than BDgraph in terms of sensitivity since it provides an high number of true discoveries and a low number of false negatives. This means that it is unlikely that a missing edge is instead present in the underlying graph. On the other hand, the BDgraph solution is better in terms of specificity, i.e., likely, an included edge represents an actual connection in the underlying graph. The specificity and sensitivity indices provide

a clear picture of the differences between the two approaches, which is no longer true if we look at the Std-SHD and the F_1 -score values in Figure 5. The latter score, in particular, is very similar for the two methods, meaning that the two approaches are equivalent in terms of misclassified edges. In other words, when dealing with a structured graph with incomplete blocks, the expected findings from a stochastic search in the complete graph space or in the block graph space are comparable, but with the latter depicting more interpretable results.

5 Analysis of fruit purees

We illustrate with a real world application how our model is able to exploit prior information to enrich the data analysis and to provide more interpretable results. The motivating problem is the analysis of spectrometric data of fruit purees (Zheng et al. 2019), which are publicly available at <https://data.mendeley.com/datasets/frrv2yd9rg>. See Holland et al. (1998) for an exhaustive description of the dataset. Our analysis focuses on 351 absorbance spectra of strawberry purees, that are displayed in Figure 6. Curves were measured on an equally spaced grid of 235 different wavelengths, whose range is $[899, 1802] \text{ cm}^{-1}$. The resulting spectra were then standardized with respect to the area under the curve, so that their final range is $[-0.1, 1.7]$. The shape of the spectra is very similar for all curves: they all exhibit a well-recognizable peak around the wavelength $1000 - 1200 \text{ cm}^{-1}$ and few secondary others, like the ones around 1600 cm^{-1} and 1700 cm^{-1} . The data were collected by means of infrared spectroscopy, that measures the interaction of infrared radiation with matter by absorption, emission, or reflection. Such technique is used to study and identify chemical substances or functional groups in solid, liquid, or gaseous forms. Indeed, from a chemical prospective, the spectrum can be seen as the identity card of the substances that are present within the compound. However the spectrum of heterogeneous compounds may be difficult to be analyzed due to overlapping emissions of different substances that interacts one another.

The problem has already been addressed in Codazzi et al. (2022), where data were firstly analyzed. From a mathematical point of view, a spectrum is a continuous function of

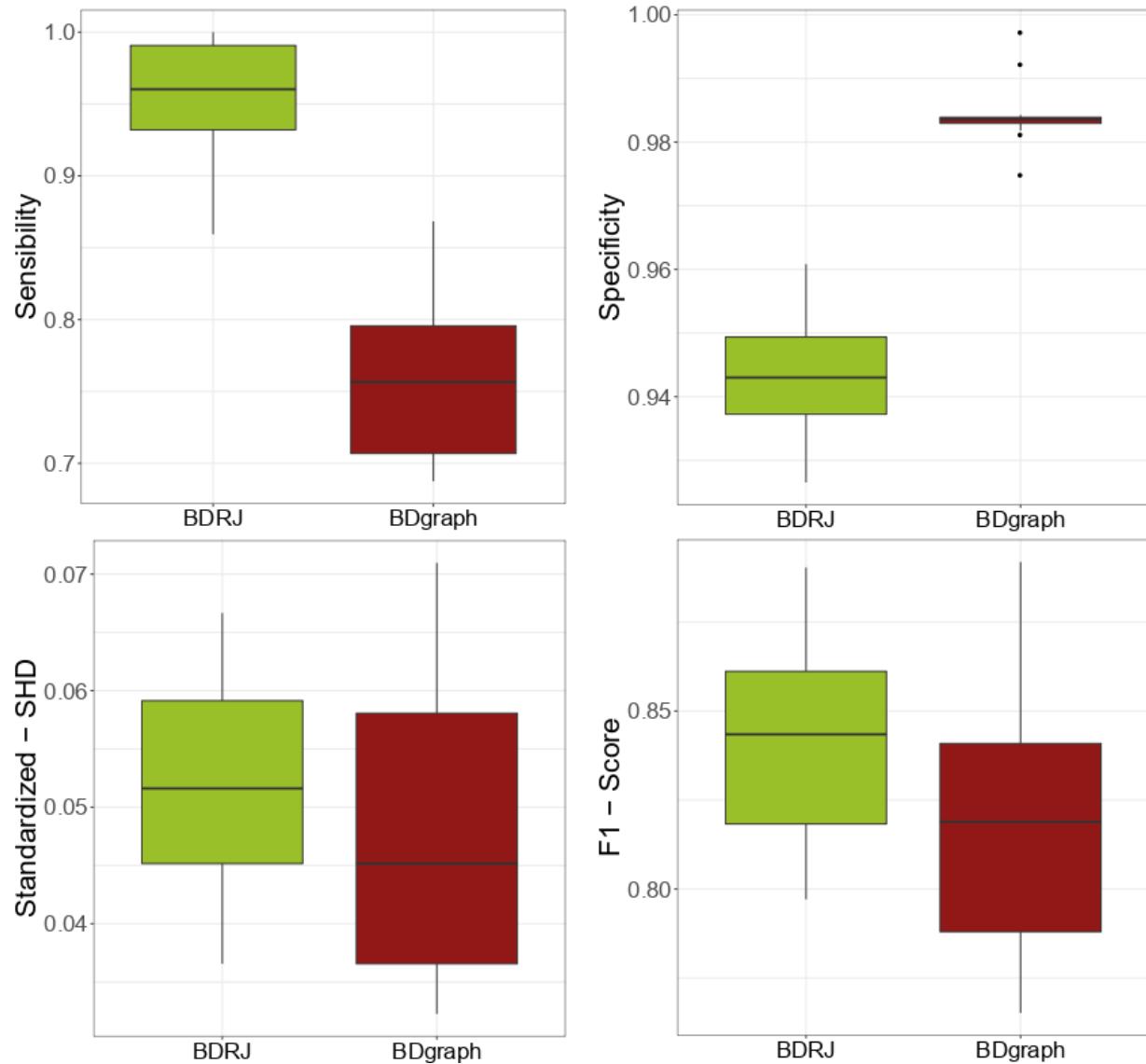


Figure 5: Boxplot of the Sensibility (top-left panel), Specificity (top-right panel), Std-SHD (bottom-left panel) and F_1 -score (bottom-right panel) over the replicates of Experiment 2.

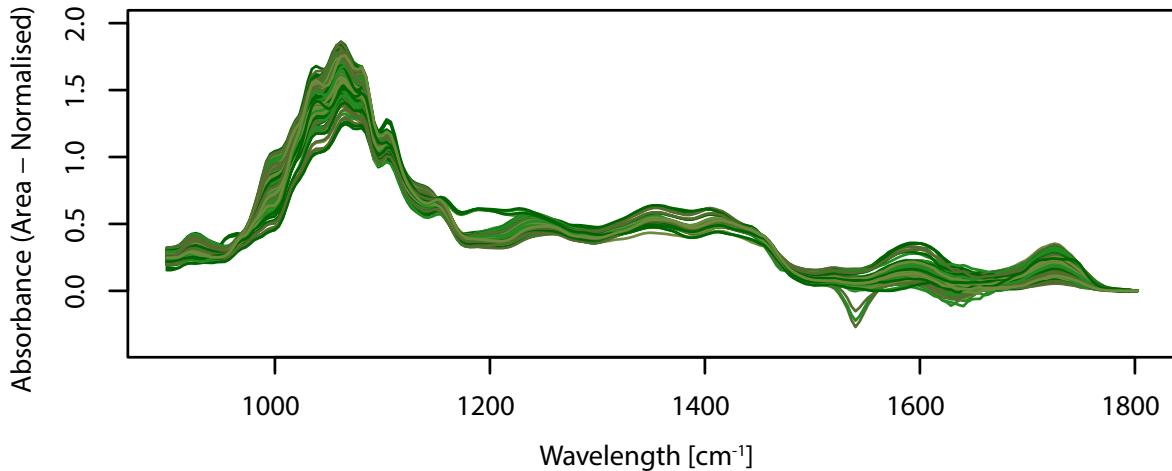


Figure 6: Plot of 351 spectra of absorbance of pure fruit purees, measured in 235 different wavelengths of the middle-infrared spectra.

the wavelength, and so the authors framed the problem within the functional data analysis setting. The classical smoothing strategy is enriched by placing a Gaussian graphical model on the basis expansion coefficients, providing an estimate of their conditional independence structure. Since the elements of a B-Spline basis have compact support, the independence structure is reflected on well defined portions of the domain. The Bayesian hierarchical formulation enables the borrowing of strength among different curves and the graphical model allows to share information along different subintervals of the functional datum. Finally, note that, in this application, the support of the spline basis functions coincide with bands of the spectrum. Therefore the problem of studying interactions between different substances simply translates in studying the dependencies between the basis expansion coefficients which can be read off from the graph.

[Codazzi et al. \(2022\)](#) used independent Bernoulli prior distributions on the graph edges and a BDgraph method to provide posterior inference on the graph and the precision matrix. As already discussed, this is a general non-informative setting that does not allow to include further prior information even when those are available. This is one of those situations: for infrared spectrography data, peaks of the signals are associated with the

vibrational modes of the different molecules present in the substance ([Atkins and De Paula 2013](#)). This is why the signals can be decomposed into different parts, corresponding to the peaks observed along the domain. As a matter of fact, domain-experts identified nine intervals of the spectrum of chemical interest, associated to the most significant peaks of the signal ([Defernez et al. 1995](#)).

Such a partition can be visualized in Figure 7, where different colors highlights the nine groups. Clearly, the nodes representing the basis expansion coefficients are ordered and so the groups are contiguous in the functional domain. The figure also shows the support of $p = 40$ spline functions; the number of basis was chosen following [Codazzi et al. \(2022\)](#), as a trade-off between good fitting of the smoothed curves and limited computational burden.

Let $\mathbf{y}_t = (y_t(s_1), \dots, y_t(s_r))^\top$ be the absorbance spectrum at all observed wavelengths of curve t , with $t = 1, \dots, T = 351$. We employ the block structured Gaussian graphical model described in Section 2 to smooth the functional data and accommodate prior knowledge on the subintervals of the spectrum. In particular, we assume $\pi_B(G_B | \theta) = \theta^{|E_B|}(1-\theta)^{\binom{M}{2}-|E_B|}$, where $\theta = 2/(M - 1) = 0.25$, that is the choice suggested by [Jones et al. \(2005\)](#) but taking into account that the number of nodes in the multigraph space is $M = 9$, not $p = 40$. All the other prior distributions and hyperparameters are set as in [Codazzi et al. \(2022\)](#). Summing up, the Bayesian hierarchical model is defined as follows:

$$\begin{aligned} \mathbf{y}_t | \boldsymbol{\beta}_t, \tau^2 &\stackrel{\text{iid}}{\sim} N_r \left(\boldsymbol{\Omega} \boldsymbol{\beta}_t, \tau^2 \mathbf{I}_r \right), \\ \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_T | \boldsymbol{\mu}, \mathbf{K} &\stackrel{\text{iid}}{\sim} N_p \left(\boldsymbol{\mu}, \mathbf{K}^{-1} \right), \\ \mathbf{K} | G &\sim \text{G-Wishart}(d, \mathbf{D}), \\ G &\sim \pi(G), \\ \boldsymbol{\mu} &\sim N_p \left(\mathbf{0}, \sigma^2 \mathbf{I}_p \right) \text{ and} \\ \tau^2 &\sim \text{IG}(a, b), \end{aligned} \tag{9}$$

where the lj -th element of the matrix $\boldsymbol{\Omega}$ is the j -th basis function evaluated at the l -th grid point s_l , $l = 1, \dots, r = 235$, \mathbf{I}_p denotes the identity matrix of size p and $\text{IG}(a, b)$ denotes the Inverse Gamma distribution with shape parameter a and rate parameter b . Posterior inference of model (9) is obtained via the BDRJ algorithm run for 450,000 iterations after

a burn-in of 50,000 and a thinning value of 25. After some tuning, we set $\sigma_g^2 = 1$.

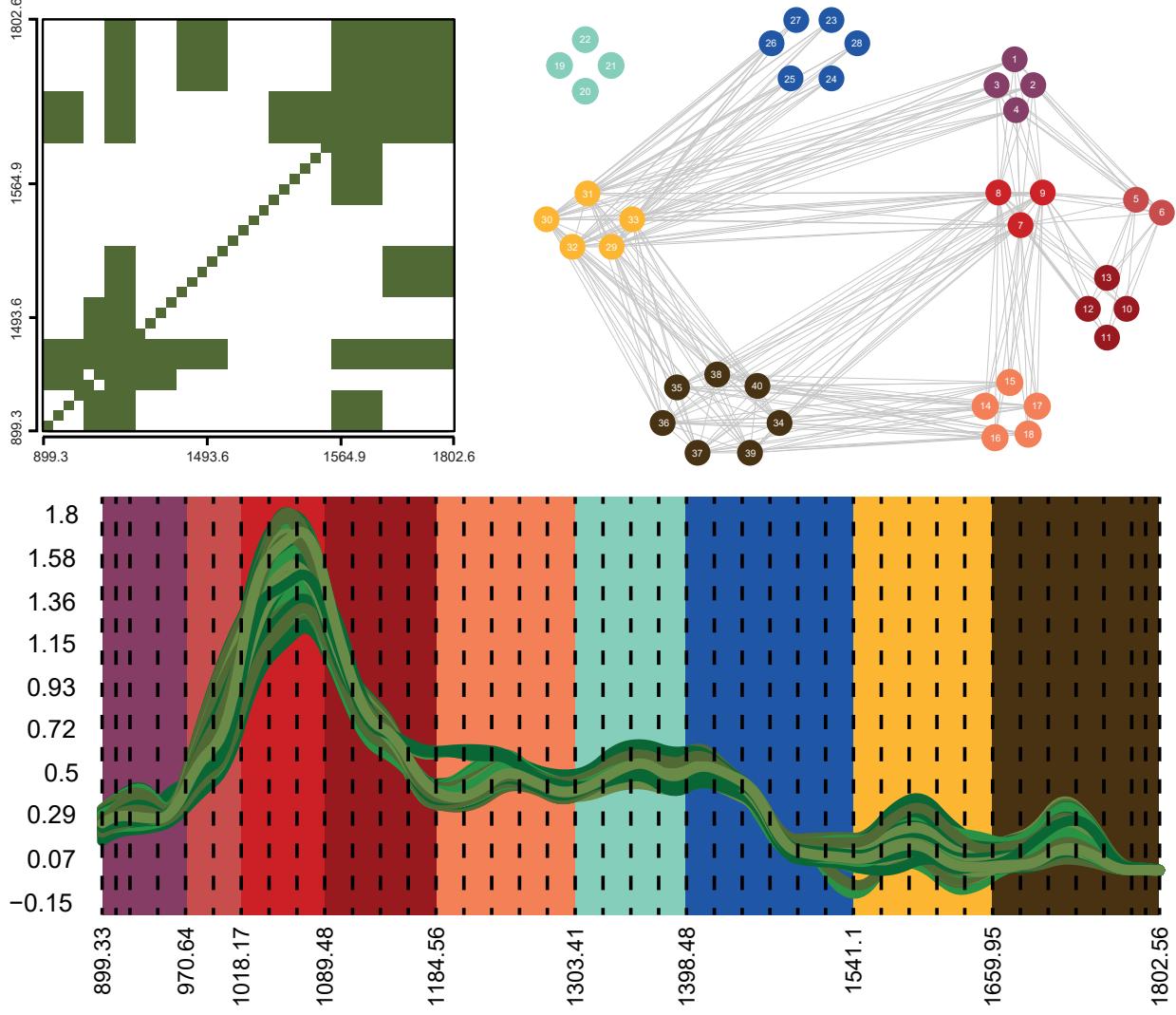


Figure 7: Top panels: posterior estimate of the graph obtained using the BFDR criterion. Nodes are colored according to different portions of the spectra highlighted in the bottom panel, where different colors represent the nine groups.

The left panel of Figure 7 shows the adjacency matrix estimated using the BFDR described in Section 3.1, where filled boxes represent the selected edges. The right panel displays the corresponding network; nodes are colored according to their group membership. The posterior adjacency matrix is characterized by two large diagonal blocks: one represents interactions within the main peak (the three red groups), which also reveals short-term interactions with the first group and the one immediately after. Similarly, the

other diagonal block represents short-term interactions within the two peaks in the tail of the spectrum as well as connections between them.

However, the peculiarity of a graphical model is the possibility of investigating long-term interactions and we indeed found out some off-diagonal blocks. The most connected group is the red one, $[1018.17 - 1089.48]cm^{-1}$, which is the central part of the main peak. It has four long-term interactions, in particular it is connected to both the final peaks. Another connection is the one between the first peak and the second-to-last peak (in yellow). It is the only long-term connection that does not involve the main peak nor the last one. Finally, note that four nodes in the interval $[1303.41 - 1398.48]cm^{-1}$ are totally disconnected, which is probably due to the fact that curves are really flat in that interval, meaning that no particular substance is absorbing in that region. Therefore, it makes sense that it is not correlated with the others. Such a disconnected component is also reported in [Codazzi et al. \(2022\)](#).

Our estimated structure of dependencies and the one reported in [Codazzi et al. \(2022\)](#) are similar, even though there are some differences due to the different modelling assumptions. The work of [Codazzi et al. \(2022\)](#) relied on a BDgraph method, which does not look for block structured graph. Therefore the estimated graph is more fragmented, i.e., there are incomplete extra diagonal blocks and some isolated links. Moreover, the graph is sparser than the one showed in Figure 7. This is coherent with the simulation experiments performed in Section 4, where we empirically showed that BDgraph is more parsimonious in including edges, which leads to a lower number of false positiveness but also to a less number of discoveries with respect to BDRJ. On the other hand our method is, by construction, prone to false positiveness. This difference is clearly visible when analysing the dependencies in the tail of the spectrum, $[1541.1 - 1802.56]cm^{-1}$. The groups are large, with five and seven nodes, respectively. According to BDgraph, some connections among the nodes in such groups are present, but only few are selected. BDRJ is also able to recognize such interactions, however it includes a large diagonal block of nodes due to the size of the involved groups. Moreover, we also notice that the inverse is possible. BDgraph found some isolated links that, according to BDRJ, are not strong enough to justify the

inclusion of a whole block and therefore are filtered away.

Overall, our approach improves the interpretability of the research findings. We recall that in an absorbance spectrum each chemical group absorbs light at a specific wavelength, therefore the main goal is to study the interactions between the peaks of the spectrum, as suggested by the experts. As a consequence, the BDgraph solution is overly detailed since incomplete blocks do not provide information about the relationships between the peaks but only about each portion of the peaks; isolated links are even more difficult to be explained. On the other hand, our model takes into account prior knowledge and balances the mathematical exactness and interpretability of the results.

6 Discussion

We introduced a new class of priors, called block graph priors, that allows to include, in a Gaussian graphical model, prior knowledge available about a partition of the nodes. A novel representation of block structured graphs is presented and then exploited to build the prior distribution on the space of block graph.

Posterior inference for general non-decomposable graphs is carried out, under the G-Wishart prior, through a novel sampling strategy. The BDRJ algorithm is a trans-dimensional Monte Carlo Markov chain sampler on the joint space of graphs and precision matrices that, at each iteration, modifies an arbitrary number of edges. This feature is new to the literature, since all existing methods we are aware of are able to change only one link at a time. With a double reversible move, the algorithm avoids the calculation of the normalizing constant $I_G(b, D)$, thus overcoming the main inferential challenge with G-Wishart priors. Moreover, thanks to dimensional reduction of the graph space, the stochastic search is usually able to get sharper solutions, in particular with respect to the F₁-score index.

We recall that our model is grounded on the hypothesis that variables in different groups are fully connected, or not connected at all. This modeling hypothesis is assumed to encode prior information about the groups of nodes. One side effect of such assumption is that we are no longer able to identify the precise structure within each block. However, as soon as the dimension of the graph increases, this modeling feature leads to easier interpretation

of the graph estimates and helps to understand the global behaviour of the phenomenon generating the data, rather than looking for all possible, even meaningless, one-to-one relationships among the variables.

A limitation of the current sampling strategy is the lower acceptance rate compared to stochastic search methods based on local moves. The issue can be mitigated by running the algorithm for many iterations, at the cost of additional computational burden. Also, a natural concern regards the need of tuning the σ_g^2 parameter, which defines the perturbation of the elements of the precision matrix that are modified in the construction of the proposed state. This parameter plays a key role in the definition of the chain, and its value has to be fixed a priori. A possible solution is to use an adaptive scheme or to generalize the recent sequential Monte Carlo method of [van den Boom et al. \(2022\)](#) to the case of global moves, i.e., when an arbitrary set of edges of the graph is changed.

A further extension can consider other types of graphical models, such as log-linear models or to non-Gaussian data by using a copula transition and extend them to the framework of block structured graphs.

Finally, in this work, we considered the case of groups of nodes that are contiguous and known a priori. In a more general framework, one could be interested in learning also the partition of the nodes. The work of [van den Boom et al. \(2022\)](#) represents a first step in such direction. Their method uses a stochastic blockmodel as a prior for the graph, where nodes in a block are assumed to form a clique. However, their sampling strategy relies on the Laplace approximation of the G-Wishart normalizing constant of [Moghaddam et al. \(2009\)](#) and it cannot be applied to graphs with a fixed labeling of the nodes. A modification of the BDRJ scheme may be, instead, a valid alternative for Bayesian learning of random block structures of graphs.

References

- Atay-Kayis, A. and H. Massam (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika* 92, 317–335.

Atkins, P. and J. De Paula (2013). *Elements of physical chemistry*. Oxford University Press, USA.

Baldi, P., S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* 16(5), 412–424.

Barbieri, M. M. and J. O. Berger (2004). Optimal predictive model selection. *Annals of Statistics* 32(3), 870–897.

Christen, J. A. and C. Fox (2005). Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical Statistics* 14(4), 795–810.

Codazzi, L., A. Colombi, M. Gianella, R. Argiento, L. Paci, and A. Pini (2022). Gaussian graphical modeling for spectrometric data analysis. *Computational Statististics & Data Analysis*.

Cremašchi, A., R. Argiento, M. De Iorio, C. Shirong, Y. S. Chong, M. J. Meaney, and M. Z. Kee (2021). Seemingly unrelated multi-state processes: a Bayesian semiparametric approach. *arXiv:2106.03072*.

Cremašchi, A., R. Argiento, K. Shoemaker, C. Peterson, and M. Vannucci (2019). Hierarchical normalized completely random measures for robust graphical modeling. *Bayesian Analysis* 14(4), 1271–1301.

Dawid, A. P. and S. L. Lauritzen (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics* 21(3), 1272 – 1317.

Defernez, M., E. K. Kemsley, and R. H. Wilson (1995). Use of infrared spectroscopy and chemometrics for the authentication of fruit purees. *Journal of Agricultural and Food Chemistry* 43(1), 109–113.

Dellaportas, P., P. Giudici, and G. Roberts (2003). Bayesian inference for nondecomposable graphical Gaussian models. *Sankhyā: The Indian Journal of Statistics* 65(1), 43–55.

- Dobra, A., C. Hans, B. Jones, J. R. Nevis, G. Yao, and M. West (2004). Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis* 90(1), 196–212.
- Dobra, A., A. Lenkoski, and A. Rodriguez (2011). Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. *Journal of the American Statistical Association* 106(496), 1418–1433.
- Giudici, P. and R. Castelo (2003). Improving Markov chain Monte Carlo model search for data mining. *Machine learning* 50(1-2), 127–158.
- Giudici, P. and P. Green (1999). Decomposable graphical Gaussian model determination. *Biometrika* 86(4), 785–801.
- Holland, J. K., E. K. Kemsley, and R. H. Wilson (1998). Use of Fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purées. *Journal of the Science of Food and Agriculture* 76(2), 263–269.
- Jones, B., C. Carvalho, A. Dobra, C. Hans, C. Carter, and M. West (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science* 20, 388–400.
- Kumar, S., J. Ying, J. V. de Miranda Cardoso, and D. P. Palomar (2020). A unified framework for structured graph learning via spectral constraints. *Journal of Machine Learning Research* 21(22), 1–60.
- Lauritzen, S. L. (1996). *Graphical models*. Oxford: Oxford University Press.
- Lenkoski, A. (2013). A direct sampler for G-Wishart variates. *Stat* 2(1), 119–128.
- Lenkoski, A. and A. Dobra (2011). Computational aspects related to inference in Gaussian graphical models with the G-Wishart prior. *Journal of Computational and Graphical Statistics* 20(1), 140–157.
- Madigan, D. and J. York (1995). Bayesian graphical models for discrete data. *International Statistical Review* 63, 215–232.

- Moghaddam, B., E. Khan, K. P. Murphy, and B. M. Marlin (2009). Accelerating Bayesian structural inference for non-decomposable Gaussian graphical models. *Advances in Neural Information Processing Systems* 22, 1285–1293.
- Mohammadi, A. and E. C. Wit (2015). Bayesian structure learning in sparse Gaussian graphical models. *Bayesian Analysis* 10(1), 109–138.
- Mohammadi, R., H. Massam, and G. Letac (2021). Accelerating Bayesian structure learning in sparse Gaussian graphical models. *Journal of the American Statistical Association* 0(0), 1–14.
- Mohammadi, R. and E. C. Wit (2019). BDgraph: An R package for Bayesian structure learning in graphical models. *Journal of Statistical Software* 89(3), 1–30.
- Müller, P., G. Parmigiani, and K. Rice (2007). FDR and Bayesian multiple comparisons rules. In J. M. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West (Eds.), *Bayesian Statistics 8*. Oxford: Oxford University Press.
- Murray, I., Z. Ghahramani, and D. MacKay (2006). Mcmc for doubly-intractable distributions. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pp. 359–366. AUAI Press.
- Osborne, N., C. B. Peterson, and M. Vannucci (2021). Latent network estimation and variable selection for compositional data via variational EM. *Journal of Computational and Graphical Statistics* 0(0), 1–22.
- Peterson, C., F. C. Stingo, and M. Vannucci (2015). Bayesian inference of multiple Gaussian graphical models. *Journal of the American Statistical Association* 110(509), 159–174.
- Powers, D. M. W. (2011). Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies* 2, 37–63.

- 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.
- Scott, J. and C. Carvalho (2008). Feature-inclusion stochastic search for Gaussian graphical models. *Journal of Computational and Graphical Statistics* 17(4), 790–808.
- Scutari, M. (2013). On the prior and posterior distributions used in graphical modelling. *Bayesian Analysis* 8(3), 505–532.
- Tsamardinos, I., L. E. Brown, C. F. Aliferis, and A. W. Moore (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning* 65(1), 31–78.
- Uhler, C., A. Lenkoski, and D. Richards (2018). Exact formulas for the normalizing constants of wishart distributions for graphical models. *The Annals of Statistics* 46(1), 90–118.
- van den Boom, W., A. Beskos, and M. D. Iorio (2022). The G-Wishart weighted proposal algorithm: Efficient posterior computation for Gaussian graphical models. *Journal of Computational and Graphical Statistics* 0(0), 1–10.
- van den Boom, W., M. De Iorio, and A. Beskos (2022). Bayesian learning of graph substructures. *arXiv: 2203.11664*.
- van den Boom, W., A. Jasra, M. De Iorio, A. Beskos, and J. G. Eriksson (2022). Unbiased approximation of posteriors via coupled particle Markov chain Monte Carlo. *Statistics and Computing* 32(3), 1–19.
- Wang, H. et al. (2015). Scaling it up: Stochastic search structure learning in graphical models. *Bayesian Analysis* 10(2), 351–377.
- Wang, H. and S. Z. Li (2012). Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electronic Journal of Statistics* 6, 168–198.

- Xia, Y., T. Cai, and T. T. Cai (2018). Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions. *Journal of the American Statistical Association* 113(521), 328–339.
- Yook, S.-H., Z. N. Oltvai, and A.-L. Barabási (2004). Functional and topological characterization of protein interaction networks. *PROTEOMICS* 4(4), 928–942.
- Zanella, G. (2020). Informed proposals for local MCMC in discrete spaces. *Journal of the American Statistical Association* 115(530), 852–865.
- Zheng, W., H. Shu, H. Tang, and H. Zhang (2019). Spectra data classification with kernel extreme learning machine. *Chemometrics and Intelligent Laboratory Systems* 192, 103815.

Appendix A

Acceptance-rejection rate

Consider the following hierarchical model

$$\begin{aligned} \mathbf{y}_1, \dots, \mathbf{y}_n \mid \mathbf{K} &\stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \mathbf{K}) \\ \mathbf{K} \mid G &\sim \text{G-Wishart}(b, D) \\ G &\sim \pi(G) \end{aligned} \tag{10}$$

where $\mathbf{y}_i \in \mathbb{R}^p$, $\forall i = 1 : p$ and \mathbf{K} is a precision matrix belonging to \mathbb{P}_G . Let $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$. The parameter of interest is the graph G , hence \mathbf{K} can be seen a nuisance parameter and marginalized out. Hence, the goal is to compute the posterior distribution ([Atay-Kayis and Massam 2005](#))

$$P(G|\mathbf{Y}) \propto \pi(G) \int_{\mathbb{P}_G} p(\mathbf{K}|G)p(\mathbf{Y}|\mathbf{K})d\mathbf{K} \propto \pi(G) \frac{I_G(d+n, D+U)}{I_G(d, D)}, \tag{11}$$

where $U = \mathbf{Y}^T \mathbf{Y} = \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^T$. The normalizing constant I_G is intractable and the existing numerical approximation methods show very high variability, as reported in [Jones et al. \(2005\)](#); [Mohammadi and Wit \(2015\)](#); [Wang and Li \(2012\)](#) (forse anche [Mohammadi et al. \(2021\)](#)). To the best of our knowledge, any method that uses such approximations should be avoided. As a consequence, one line of research exploit the approximation of the ratio of two normalizing constants given in [Mohammadi et al. \(2021\)](#), while another one any direct evaluation of I_G by leveraging on the exchange algorithm ([Murray et al. 2006](#)), see for example [Wang and Li \(2012\)](#); [van den Boom et al. \(2022\)](#) and [Lenkoski \(2013\)](#). The last one is the building block of our BDRJ, which is the generalization to the block structured graph case.

The augmented target is the joint distribution

$$p(\mathbf{K}, G, \tilde{\mathbf{W}}, G'|\mathbf{Y}) \propto p(\mathbf{K}, G|\mathbf{Y})q(G'|G)p(\tilde{\mathbf{W}}|G'), \tag{12}$$

where $G, G' \in \mathcal{B}$ and $\tilde{\mathbf{W}} \in \mathbb{P}_{G'}$, so that the marginal density of $(\mathbf{K}, G|\mathbf{Y})$ remains unchanged. Let $(\mathbf{K}^{[s]}, G^{[s]})$ be the current state of the chain, hence $\mathbf{K}^{[s]} \in \mathbb{P}_{G^{[s]}}$. Suppose, without loss of generality, $G' \in nbd_M^{\mathcal{B},+}(G_B^{[s]})$ which means that its multigraph

representative G'_B has been obtained by adding link (l, m) to $G_B^{[s]}$. Suppose, without loss of generality, the proposed graph G' is obtained from (4) by adding edge (l, m) to the multigraph representation of $G^{[s]}$. The set of links that are changing in \mathcal{G} is $L = \{(i, j) \mid i, j \in V, i < j, (i, j) \in E', (i, j) \notin E^{[s]}\}$. Let $l = |L|$ be its cardinality, which is arbitrary and, in general, different from one. Note that, by construction, L can not contain diagonal elements and that $v(G') = v(G^{[s]}) \cup L$. We call $V(L) = B_l \cup B_m$ the set of the vertices involved in the change.

The Block Double Reversible Jump consider switching between $(\mathbf{K}^{[s]}, G^{[s]}, \widetilde{\mathbf{W}}, G')$ to the alternative $(\mathbf{K}', G', \mathbf{W}^0, G^{[s]}), \mathbf{W}^0 \in \mathbb{P}_{G^{[s]}}$, by performing two Reversible Jump moves; a dimension increasing step from $(\mathbf{K}^{[s]}, G^{[s]})$ to (\mathbf{K}', G') according to posterior parameters $b + n$ and $D + U$ and a dimension decreasing step from $(\widetilde{\mathbf{W}}, G')$ to $(\mathbf{W}^0, G^{[s]})$ according to prior parameters b and D . Such transitions exploit the change of variable $\mathbf{K} \mapsto \Phi$, where Φ is an upper triangular matrix such that $\mathbf{K} = \Phi^T \Phi$. Some of its properties have been introduced in Section 3. Beside those, [Roverato \(2002\)](#) proved that the Jacobian of such a transformation is

$$J(\mathbf{K} \mapsto \Phi) = 2^p \prod_{i=1}^p \Phi_{ii}^{\nu_i^G}, \quad (13)$$

where $\nu_i^G = |\{j : j > i \text{ and } (i, j) \in E\}|$ is the sum of elements in i -th row of adjacency matrix, from position $i + 1$ up to the end. We also define $d_i^G = |\{j : j < i \text{ and } (i, j) \in E\}|$ that is the sum of elements in i -th row of adjacency matrix, from position 1 up to the $i - 1$ -th or, exploiting the symmetry of the matrix, the sum of elements in i -th column, from position 1 up to the $i - 1$ -th. It is clear that $\nu_i^G + d_i^G$ represents the number of neighbors of vertex i . The quantities one wants to compute in practice are $d_i^{G'} - d_i^{G^{[s]}}$ and $\nu_i^{G'} - \nu_i^{G^{[s]}}$, for each $i \in V(L)$. Suppose without loss of generality that $i \in B_l$, then we have

$$d_i^{G'} - d_i^{G^{[s]}} = |\{j \in B_m : j < i\}|, \quad (14)$$

$$\nu_i^{G'} - \nu_i^{G^{[s]}} = |\{j \in B_m : j > i\}|, \quad (15)$$

that are the number of indices smaller and greater than i in the other group involved in the change, respectively.

Moreover, we recall a well known property of the Cholesky decomposition that allows us to rapidly determine the determinant of \mathbf{K} . We have

$$\det(\mathbf{K}) = \prod_{i=1}^p \Phi_{ii}^2 \quad (16)$$

Note that this formulation involves only diagonal values of Φ that are free elements by definition. This implies, by (16), that $\det(\mathbf{K}') = \det(\mathbf{K}^{[s]})$. As explained in Section 3, after the change of variable, we set $\Phi'_{ij} = \Phi_{ij}$ for all $(i, j) \in \nu(G^{[s]})$. The free elements that are left are the ones in the set L that are proposed by perturbing the old values independently and with the same variance σ_g^2 , which is a tuning parameter. Namely, we set

$$\Phi'_{ij} = \Phi_{ij}^{[s]}, \quad \forall (i, j) \in E, \quad (17)$$

then, $\forall h \in L$ sample

$$\eta_h \sim N\left(\Phi_h^{[s]}, \sigma_g^2\right), \quad (18)$$

This defines all free elements of Φ' , the non-free elements may be determined through the completion operation. Hence Φ' is well defined and so is $\mathbf{K}' = (\Phi')^T \Phi'$.

The acceptance probability of such a move is $\min(1, R^+)$, with

$$R^+ = \frac{p(\mathbf{K}', G', \mathbf{W}^0, G^{[s]} | \mathbf{Y})}{p(\mathbf{K}^{[s]}, G^{[s]}, \tilde{\mathbf{W}}, G' | \mathbf{Y})} \frac{J(\mathbf{K}' \rightarrow \Phi') J(\mathbf{W}^0 \rightarrow \Phi^0)}{J(\mathbf{K}^{[s]} \rightarrow \Phi^{[s]}) J(\tilde{\mathbf{W}} \rightarrow \tilde{\Phi})} \frac{q(\mathbf{K}' | \mathbf{K}^{[s]})}{q(\mathbf{W}^0 | \tilde{\mathbf{W}})}. \quad (19)$$

As mentioned before, two Reversible Jumps are performed. In both cases, the change of dimension is equal to $l \geq 1$. Hence in (19) we should also consider the Jacobian of the transformation mapping $\left((\Phi^{[s]})^{\nu(G^{[s]})}, \eta_1, \dots, \eta_l\right) \rightarrow (\Phi')^{\nu(G')}$, which is a linear and its Jacobian is equal to one. Indeed, it can be expressed as

$$(\Phi')^{\nu(G')} = \mathbb{I}_{|\nu(G^{[s]})|+l}((\Phi^{[s]})^{\nu(G^{[s]})}, \eta_1, \dots, \eta_l)^T. \quad (20)$$

When developing the calculation, one can use (12) to factorize the first ratio in (19) as

$$\begin{aligned} \frac{p(\mathbf{K}', G', \mathbf{W}^0, G^{[s]} | \mathbf{Y})}{p(\mathbf{K}^{[s]}, G^{[s]}, \tilde{\mathbf{W}}, G' | \mathbf{Y})} &= \\ &= \frac{\pi(G')}{\pi(G^{[s]})} \frac{p(\mathbf{Y} | \mathbf{K}', G')}{p(\mathbf{Y} | \mathbf{K}^{[s]}, G^{[s]})} \frac{p(\mathbf{K}' | G')}{p(\mathbf{K} | G^{[s]})} \frac{p(\mathbf{W}^0 | G^{[s]})}{p(\tilde{\mathbf{W}} | G')} \frac{q(G^{[s]} | G')}{q(G' | G^{[s]})} = \\ &= \frac{\pi(G')}{\pi(G^{[s]})} \frac{|nbd_M^{\mathcal{B},+}(G_B^{[s]})|}{|nbd_M^{\mathcal{B},-}(G'_B)|} \frac{\exp\left\{-\frac{1}{2}\langle \mathbf{K}' - \mathbf{K}^{[s]}, D + U \rangle\right\}}{\exp\left\{-\frac{1}{2}\langle \tilde{\mathbf{W}} - \mathbf{W}^0, D \rangle\right\}}. \end{aligned} \quad (21)$$

Note that the two ratios of G-Wishart distributed variables allow to eliminate the presence of the normalizing constants.

Regarding the ratio of Jacobians, we have

$$\frac{J(\mathbf{K}' \rightarrow \Phi')}{J(\mathbf{K}^{[s]} \rightarrow \Phi^{[s]})} = \frac{2^p}{2^p} \frac{\prod_{i=1}^p (\Phi'_{ii})^{\nu_i^{G'} + 1}}{\prod_{i=1}^p (\Phi_{ii}^{[s]})^{\nu_i^G + 1}} = \prod_{i \in V(L)} \left(\Phi_{ii}^{[s]} \right)^{\nu_i^{G'} - \nu_i^G}, \quad (22)$$

where we used the fact that $\Phi_{i_0,i_0} = \Phi'_{i_0,i_0}$ by construction and $\nu_i(G) = \nu_i(G')$ for all $i \neq V(L)$. Analogously, one can show that $\frac{J(\mathbf{W}^0 \rightarrow \Phi^0)}{J(\tilde{\mathbf{W}} \rightarrow \tilde{\Phi})} = \prod_{i \in V(L)} (\Phi_{ii}^0)^{\nu_i^{G'} - \nu_i^G}$. The exponent, is readily computed using (15).

Finally, the last ratio in (19) is due to the randomness in the construction of the proposed and the auxiliary matrices. By definition, each term is just the ratio of independent multivariate Gaussian densities,

$$q(\mathbf{K}' | \mathbf{K}^{[s]}) = q(\eta_1, \dots, \eta_l) = \prod_{i=1}^l q_i(\eta_i) = \left(\frac{1}{\sqrt{2\pi\sigma_g^2}} \right)^l \exp \left\{ -\frac{1}{2\sigma_g^2} \sum_{h \in L} (\Phi'_h - \Phi_h)^2 \right\}, \quad (23)$$

where, for sake of clarity, in the last equality we explicitly wrote the ratio in terms of Φ' instead of η , but the two quantities coincide. The second ratio is similar.

Wrapping everything together, we end up with

$$R^+ = \frac{\pi(G')}{\pi(G^{[s]})} \frac{\exp \left\{ -\frac{1}{2} \langle \mathbf{K}' - \mathbf{K}^{[s]}, D + U \rangle \right\}}{\exp \left\{ -\frac{1}{2} \langle \tilde{\mathbf{W}} - \mathbf{W}^0, D \rangle \right\}} \prod_{i \in V(L)} \left(\frac{\Phi_{ii}^{[s]}}{\Phi_{ii}^0} \right)^{\nu_i^{G'} - \nu_i^G} \exp \left\{ \frac{1}{2\sigma_g^2} \sum_{h \in L} [(\Phi'_h - \Phi_h^{[s]})^2 - (\Phi_h^0 - \tilde{\Phi}_h)^2] \right\}. \quad (24)$$