

# Block Structured Graph Priors in Gaussian Graphical Models



Alessandro Colombi

**Abstract** Gaussian graphical models are a powerful statistical tool to describe the concept of conditional independence between variables through a map between a graph and the family of multivariate normal models. The structure of the graph is unknown and has to be learned from the data. Inference is carried out in a Bayesian framework: thus, the structure of the precision matrix is constrained by the graph through a G-Wishart prior distribution. In this work we first introduce a prior distribution to impose a block structure in the adjacency matrix of the graph. Then we develop a Double Reversible Jump Monte Carlo Markov chain that avoids any G-Wishart normalizing constant calculation when comparing graphical models. The novelty of this procedure is that it looks for block structured graphs, hence proposing moves that add or remove not just a single link but an entire group of them.

**Keywords** Bayesian statistics · Double reversible jump · G-Wishart prior

## 1 Introduction

The increasing capacity of human beings of collecting large amount of data gave rise to the need of developing models to study how variables interact with one another. Benefits of such discoveries are well known, for example in clinical and genetic applications it is useful to understand how risk factors are related so that patient-specific therapies may be planned. See [5, 9, 27] for cancer applications. The same reasoning applies to problems in economics, for example [25] studied the interconnectedness of credit risk.

Probabilistic graphical modeling is a possible approach to the task of studying the dependence structure among a set of variables. It 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 probabilities

---

A. Colombi (✉)  
Università di Milano-Bicocca, Milan, Italy  
e-mail: [a.colombi10@campus.unimib.it](mailto:a.colombi10@campus.unimib.it)

is chosen to be Gaussian, those models are known as Gaussian graphical models [12]. This is the choice made throughout the paper, which is the most common in the literature.

Let  $\mathbf{X}$  be a  $p$ -random vector distributed as  $N_p(\mathbf{0}, \Sigma)$ .  $\Sigma$  is the covariance matrix and we assume  $\mathbf{X}$  to be centered without loss of generality. Let  $G = (V, E)$  be an undirected graph, where  $V = \{1, \dots, p\}$  is the set of nodes and  $E$  is the set of undirected edges.  $\mathbf{X}$  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 are conditionally independent given all the others. Moreover, under the normality assumption, the conditional independence relationship between variables can be represented in terms of the null elements of the precision matrix  $\mathbf{K} = \Sigma^{-1}$ . Therefore the following equivalence provides an interpretation of the graph

$$X_i \perp\!\!\!\perp X_j \mid \mathbf{X}_{-(ij)} \iff (i, j) \notin E \iff k_{ij} = 0, \quad (1)$$

where  $\mathbf{X}_{-(ij)}$  is the random vector containing all elements in  $\mathbf{X}$  except the  $i$ -th and the  $j$ -th. Each node is associated to one of the variables of interest and its links 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,  $G$  is unknown and it is the goal of the statistical inference, along with  $\mathbf{K}$ . Such a process is also known as structural learning. In a Bayesian framework, we set a G-Wishart prior distribution for the precision matrix  $\mathbf{K}$  [1, 20], which is attractive as it is conjugate to the likelihood. Since the graph  $G$  is considered to be a random variable having values in the space  $\mathcal{G}$  of all possible undirected graphs with  $p$  nodes, we need to specify a prior on it. A common practice is to choose a uniform distribution over  $\mathcal{G}$ . This is appealing for its simplicity but it assigns most of its mass to graphs with a “medium” number of edges [9]. On the other hand, it is known that an undirected graph is uniquely identified by its set of edges  $\mathcal{E}$ . Therefore it is simpler to define a prior on  $\mathcal{E}$ , which then naturally induces a prior over  $\mathcal{G}$ . In this setting, the most natural choice is to assign independent Bernoulli priors to each link. The Bernoulli parameters  $\theta$  could be different from edge to edge, but one usually assigns a common value. For example, [9] suggested to choose  $\theta = 2/(p - 1)$  to induce more sparsity in the graph. Scott and Carvalho [21] placed instead a Beta hyperprior on that parameter, a solution known as multiplicity correction prior. Similarly, [22] described a multivariate Bernoulli distribution where edges are not necessarily independent. A common feature of previously described priors is that they are non-informative. The only type of prior information they can include in the model is the expected sparsity.

In this work we propose a prior for the graph that aims to be informative, according to the prior information available for the application at hand. Since the graph describes the conditional dependence structure of variables involved in complex and high-dimensional phenomena, it is unrealistic to assume that prior knowledge is available for one-to-one relationships between the observed quantities. It is instead more reasonable to assume that variables may be grouped in smaller subsets. This

is common in biological application where the groups may be families of bacteria [17], or genomics where groups of genes are known to be part of a common process. Also in market basket analysis products and customers can easily be grouped; see, for instance [6].

We propose a class of priors, called *block graph priors*, that encodes such information and imposes a block structure in the adjacency matrix that describes the graph. We allow variables in different groups only to be fully connected or not connected at all. Therefore, the goal is no longer in looking for all possible relationships between nodes but on deriving the underlying pattern between groups.

We introduce a Reversible Jump sampler that leverages the structure induced by our new prior. In particular, we generalize the procedure by Lenkoski [13]. The resulting method is called Block Double Reversible Jump (BDRJ for short). Its main feature is that it modifies, at each step of the chain, an entire block of links to guarantee a block structure that is always compatible with our hypotheses.

The remainder of the paper is organized as follows. Section 2 introduces the block structured graph priors and Sect. 3 provides the sampling strategy. In Sect. 4 we present a simulation study along with a comparison against an existing approach. Finally, we conclude with a brief discussion in Sect. 5.

## 2 Block Structured Graph Priors

The starting point for our proposed model is that we assume the  $p$  observed variables to be grouped, a priori, in  $M$  mutually exclusive groups. 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*.

We aim to study relationships between groups of variables. Therefore the usual graph representation  $G = (V, E)$ , where  $V$  is the set of nodes and  $E$  is set of links, is redundant. Indeed we assume that groups are given and links have to satisfy a precise block structure. As a consequence, we synthesize those information by defining a new space of undirected graphs whose nodes represent the chosen groups of variables and links represent the structure of relationships between them. Namely, let  $V_B = \{B_1, \dots, B_M\}$  be a partition of  $V$  in  $M$  groups that are 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,

$$E_B \subset \mathcal{E}_B = \left\{ (l, m) | l, m \in V_B, \wedge l < m, (l, l) | l \in V_B, \wedge n_l > 1 \right\}. \quad (2)$$

In graph theory, graphs that have self-loops are called *multigraphs*. Finally, let  $\mathcal{G}_B$  be the set of all possible multigraphs  $G_B$  having  $V_B$  as set of nodes. In the following, we want to clarify the relationship between this space and  $\mathcal{G}$ .

Consider  $G_B \in \mathcal{G}_B$  and  $G \in \mathcal{G}$ . By definition, the set of nodes of the first multigraph is obtained by grouping together the nodes of the second graph. What about



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

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

$$\begin{aligned}
 V &= \{B_{l,h}, l = 1, \dots, n_h, h = 1, \dots, M\} = \{1, \dots, p\} \\
 \text{if } (l, m) \in E_B &\Rightarrow (i, j) \in E \forall i \in B_l, \quad \forall j \in B_m \\
 \text{if } (l, m) \notin E_B &\Rightarrow (i, j) \notin E \forall i \in B_l, \quad \forall j \in B_m
 \end{aligned} \tag{3}$$

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

Nevertheless,  $\rho$  is not surjective which implies that there are graphs that do not have a representative in  $\mathcal{G}_B$ . Indeed, only those graphs with 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 our modelling assumptions. Let us consider the image of  $\rho$ , denoted by  $\mathcal{B}$ . It is 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  $\rho^{-1}$ . We say that  $G \in \mathcal{B}$  is the block graph representation of the multigraph  $G_B \in \mathcal{G}_B$ . This synthesised representation of block graphs allows us to work in a space where we can use standard tools of graphical analysis. In a different setting, [4] adopts a similar approach to model the conditional dependence across Markov processes.

In particular, such a representation allows us to introduce a class of priors that encodes the knowledge about the partition of the nodes. We place zero mass probability on all those graphs that belong to  $\mathcal{G} \setminus \mathcal{B}$ , which is the set of all those graphs that do not satisfy our block structure constraint. Then, we place a standard prior, say  $\pi_B(\cdot)$ , over  $\mathcal{G}_B$ , which is possible as it is a space of undirected multigraphs where links can be considered to be independent with one another. Finally, we map the results in  $\mathcal{B}$  using  $\rho^{-1}$ . Namely

$$\pi(G) \propto \begin{cases} \pi_B(\rho^{-1}(G)), & \text{if } G \in \mathcal{B} \\ 0, & \text{if } G \in \mathcal{G} \setminus \mathcal{B}. \end{cases} \tag{4}$$

We refer to those priors as *block graph priors*. In this work, we consider a *block-Bernoulli prior*,  $\pi(G)$ , that is obtained by applying (4) to  $\pi_B(G_B) = \theta^{|E_B|}(1 - \theta)^{\binom{M}{2} - |E_B|}$ , that is the Bernoulli prior where each link has prior probability of inclusion  $\theta$ , which is fixed a priori. The reasoning used to define block priors is similar to priors described in [7]. However, in this case we are not limiting the learning of the graph to the class of the decomposable ones but to block structured graphs. Moreover, this limitation is not due to computational limitations but because prior knowledge is available. In the next section we present a method to learn such a structure. In principle, one can still apply such prior to limit the analysis to the class of decomposable block graphs to exploit their properties. However, in this work we do not make such assumption and we present a method that is valid also for non-decomposable graphs.

### 3 Sampling Strategy

One of the difficulties in the development of efficient methods for structural learning is the presence of the G-Wishart prior distribution. Given a random matrix  $\mathbf{K}$ , we say that  $\mathbf{K} | G, b, D \sim \text{G-Wishart}(b, D)$  if its density is

$$P(\mathbf{K} | G, b, D) = I_G(b, D)^{-1} |\mathbf{K}|^{\frac{b-2}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{K}D)\right\} \mathbb{1}_{\mathbb{P}_G}(\mathbf{K}), \quad (5)$$

where  $b$  and  $D$  are fixed hyperparameters,  $\mathbb{P}_G$  is the space of all  $p \times p$  symmetric and positive definite matrices whose null elements are associated to links absent in graph  $G$  and

$$I_G(b, D) = \int_{\mathbb{P}_G} |\mathbf{K}|^{\frac{b-2}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{K}D)\right\} d\mathbf{K}, \quad (6)$$

is an intractable normalizing constant. Numerical methods to approximate such a constant [1] are unstable in high dimensional problems [9, 15]. Several techniques that avoid any calculation of  $I_G(b, D)$  are available in the literature, but an exhaustive review of them is beyond the goals of this work. In the following, we limit ourselves to present how our proposed method, called Block Double Reversible Jump (BDRJ for short). It generalizes the procedure by Lenkoski [13] to get a Reversible Jump chain defined over the joint space of graph and precision matrix that visits only the subspace  $\mathcal{B}$  of block structured graphs. Note that if one is interested only in decomposable graph models, the normalizing constant  $I_G(b, D)$  can be computed explicitly and it would be enough to use a standard Metropolis-Hastings algorithm without resorting to the usage of the Reversible Jump technique presented in the remaining part of this paper.

We denote the current state of the chain by  $(\mathbf{K}^{[s]}, G^{[s]})$ , with  $\mathbf{K}^{[s]} \in \mathbb{P}_{G^{[s]}}$ . The proposed state  $(\mathbf{K}', G')$  is constructed in two subsequent steps; in Sect. 3.1 we describe the proposal for the new graph  $G'$  and then in Sect. 3.2 we discuss how to get the

proposed precision matrix  $\mathbf{K}' \in \mathbb{P}_{G'}$ . Once that  $(\mathbf{K}', G')$  has been drawn, we accept or reject the whole state with a Metropolis-Hastings step.

### 3.1 Construction of Proposed 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  $G$ . Namely,  $nb d_p(G) = nb d_p^+(G) \cup nb d_p^-(G)$  where  $nb d_p^+(G)$  and  $nb d_p^-(G)$  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  $G' \in nb d_p(G^{[s]})$  is said to be a local move.

The proposed BDRJ approach is innovative because we derive 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  $\mathcal{G}$ . Suppose  $G^{[s]} \in \mathcal{B}$ , we propose a new graph  $G' \in \mathcal{B}$  by first drawing its multigraph representation  $G'_B \in \mathcal{G}_B$  from

$$q(G'_B | G^{[s]}) = \frac{1}{2} \text{Unif} \left( nb d_M^{\mathcal{B},+} \left( \rho^{-1}(G^{[s]}) \right) \right) + \frac{1}{2} \text{Unif} \left( nb d_M^{\mathcal{B},-} \left( \rho^{-1}(G^{[s]}) \right) \right), \quad (7)$$

where  $nb d_M^{\mathcal{B}}(G_B^{[s]})$  is the one-edge-away neighbourhood of  $G_B^{[s]} = \rho^{-1}(G^{[s]})$  with respect to the space of multigraphs  $\mathcal{G}_B$ . Addition and removal moves are chosen with the same probability. Given this choice,  $q(G'_B | G^{[s]})$  chooses, with uniform probability, which link is to be added (or removed). Finally  $\rho$  is applied once again to map the resulting multigraph back in  $\mathcal{B}$  to obtain  $G'$ , i.e. setting  $G' = \rho(G'_B)$ . A closer look at (7) 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 usual space  $\mathcal{G}$ .

### 3.2 Construction of Proposed Precision Matrix $\mathbf{K}'$

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. The method by Wang and Li [26] based on the partial analytical structure of the G-Wishart appears to be an efficient choice. However, it 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 [20], can be handled in practice only if at each step of the graph only one link of the graph is modified. Instead, the proposal distribution presented in Sect. 3.1 modifies an arbitrary number of links. Hence, it is complicated, if possible at all, to generalize the method by Wang and Li [26] to such framework. As a consequence, we rely on a generalization of the Reversible Jump mechanism by Lenkoski [13]. The idea is that it is possible to guarantee the positive definiteness of

$\mathbf{K}'$  and the zero constraints imposed by  $G'$  just by working on the Cholesky decomposition matrix  $\Phi^{[s]}$  of  $\mathbf{K}^{[s]}$ . Indeed, [20] and [1] showed that the zero constraints imposed by  $G^{[s]}$  on the off-diagonal elements of  $\mathbf{K}^{[s]}$  induce a precise structure and properties on  $\Phi^{[s]}$ . Let  $\nu(G^{[s]}) = \{(i, j) \mid i, j \in V, i = j \text{ or } (i, j) \in E^{[s]}\}$  be the set of the diagonal elements and the links belonging to  $G^{[s]}$  and define the set of *free elements* of  $\Phi^{[s]}$  as  $\Phi^{\nu(G^{[s]})} = \{\phi_{ij} \mid i, j \in \nu(G^{[s]})\}$ . The remaining entries, that we simply refer to as non-free elements, are uniquely determined through the completion operation [1, Prop. 2] as a function of the free elements.

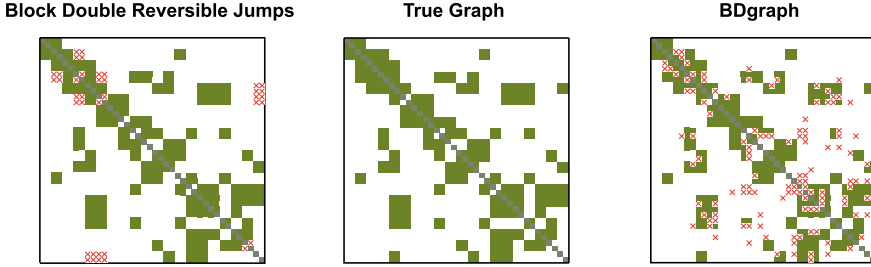
Suppose the proposed graph  $G'$  is obtained by adding edge  $(l, m)$  to the multigraph representation of  $G^{[s]}$ . The set of links that are changing in  $\mathcal{G}$  is  $L = \{(i, j) \mid i, j \in V, i < j, (i, j) \in E', (i, j) \notin E^{[s]}\}$ . Its cardinality  $l = |L|$  is arbitrary and, 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  $\nu(G') = \nu(G^{[s]}) \cup L$ . 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  $\sigma_g^2$ . Namely, draw  $\eta_h \sim^{ind} N(\phi_h^{[s]}, \sigma_g^2)$  and set  $\phi'_h = \eta_h$  for each  $h \in L$ . Then, it is enough to derive all non free elements of  $\Phi'$  through completion operation and finally to set  $\mathbf{K}' = (\Phi')^T \Phi$ . Note that, by doing so, we are generating a random variable  $\eta$  of length  $l$  that matches the dimension gap between  $\mathbf{K}$  and  $\mathbf{K}'$ . As usual, the dimension decreasing case is deterministically defined in terms of the dimension increasing one.

## 4 Simulation Study

We compare our performances to the *Birth and Death* approach (BDMCMC for short) proposed by Mohammadi and Wit [14] and available in the R package `BDgraph` [16].

All final estimates, both from BDRJ and BDMCMC outputs, were obtained by controlling the *Bayesian False Discovery Rate*, as presented in [18] and [3]. Performances are assessed in terms of the standardized Structural Hamming Distance (Std-SHD, [23]) and the  $F_1$ -score [2, 19]. The first one prefers lower values, the second one higher values. Following the same approach of [14, 24], precision matrix estimation is measured using one half of the Stein loss score (SL) [8] which is equal to the Kullback-Leibler divergence [10] between  $N_p(\mathbf{0}, \mathbf{K}_{true}^{-1})$  and  $N_p(\mathbf{0}, \widehat{\mathbf{K}}^{-1})$ .

In the first experiment, we set  $p = 40$ ,  $n = 500$  and  $M = p/2$  groups of equal size, which leads to off-diagonal blocks of size  $2 \times 2$ . The true underlying graph is itself a block structured graph (see Fig. 2), while the true precision matrix was sampled by drawing from a G-Wishart  $(3, \mathbf{I}_p)$ .  $\sigma_g^2$  was set equal to 0.5 after a little tuning phase. 400,000 iterations were run plus 100,000 extra iterations as burn-in period that were discarded. A simple visual inspection of Fig. 2 suggests that BDRJ is more precise than BDMCMC. The number of misclassified edges is rather low, Std-SHD = 0.0243, and it is well balanced between false positiveness (10) and negativeness (9). Many true discoveries are achieved and indeed it has  $F_1$ -score = 0.954.



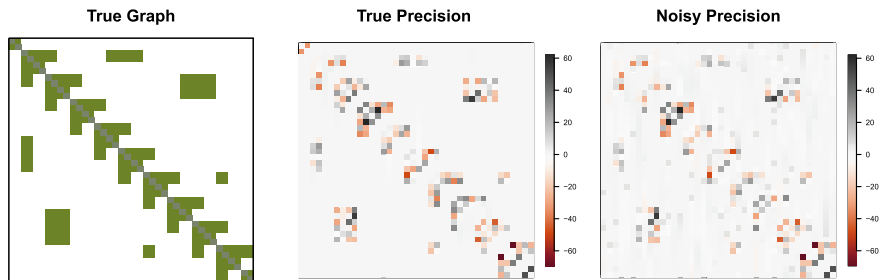
**Fig. 2** The adjacency matrices of true underlying graph (middle panel), the BDRJ one (leftmost panel) and the one obtained using `BDgraph` (rightmost panel). Squares represent the included links, crosses stand for edges that are wrongly classified

BDMCMC does not recognize the block structure of the true graph, actually it does not even look for such a structure because the prior information can not be included in the model. It estimates the probability of inclusion of every possible link independently from the others. This entails more errors in the final estimate as well as a less informative structure of the graph. It would be hard to explain why there are missing edges within some structures that are clearly blocked ones. We repeated the same experiments for 18 different dataset: the true underlying graphs were randomly generated by sampling from (4) with different sparsity indices  $\theta$  uniformly distributed in  $[0.2, 0.6]$ . The mean values, along with the standard deviations, for the  $F_1$ -scores are 0.845(0.13) and 0.80(0.03) and for the Std-SHD we have 0.053(0.04) and 0.060(0.02), respectively for BDRJ and BDMCMC. We see that BDRJ is more unstable with respect to BDMCMC. This is probably due to the fact that we used the same  $\sigma_g^2$  for all dataset, without tuning it every time. However both indices prefer BDRJ.

The second experiment is inspired by a simulation study presented in [11] that aims to learn a graph and precision matrix under a noisy setting. The true underlying graph  $G$  is displayed in Fig. 3. We sample  $\mathbf{K}_{\text{true}}|G \sim \text{G-Wishart}(3, \mathbf{I}_p)$  and set  $\mathbf{K}_{\text{noisy}}$  to be a random perturbation of  $\mathbf{K}_{\text{true}}$ : every possible value is perturbed, with probability  $s$ , by adding a random noise  $0.1u$ . Here  $u \sim \text{Unif}(-k^*, k^*)$ , where  $k^* = \max_{i < j} |k_{ij}^{\text{true}}|$ . Finally, data are generated from  $N_p(\mathbf{0}, \mathbf{K}_{\text{noisy}}^{-1})$ . To investigate the behaviour under different volumes of noise,  $s = 0.10, 0.20, 0.25$ , we repeat each experiment 15 times. Results are reported in Table 1.

BDRJ outperforms BDMCMC on every dataset and with respect to all indices we considered. Its robustness is due to the fact that to conclude that a whole block has to be inserted in the final graph a single, isolated link is not enough. Those isolated values are not compatible with the block structured graph that BDRJ is looking for, therefore they are rightly ignored. On the other hand, BDMCMC does not look for any particular structure, hence it does not recognize the perturbed values as noise.





**Fig. 3** The true underlying graph (leftmost panel) used to generate the true precision matrix  $\mathbf{K}_{\text{true}}$  (middle panel). The rightmost panel is  $\mathbf{K}_{\text{noisy}}$  (obtained with  $s = 0.25$ ). For plotting purposes, we removed the diagonal in both precision matrices

**Table 1**  $F_1$ -score, Std-SHD and Stein loss SL under different noise levels  $s$

		$s = 0.1$	$s = 0.2$	$s = 0.25$
F1-score	BDRJ	<b>0.75 (0.035)</b>	<b>0.70 (0.043)</b>	<b>0.68 (0.043)</b>
	BDMCMC	0.44 (0.022)	0.405 (0.028)	0.066 (0.042)
SHD	BDRJ	<b>0.039 (0.004)</b>	<b>0.047 (0.043)</b>	<b>0.049 (0.005)</b>
	BDMCMC	0.063 (0.002)	0.36 (0.042)	0.071 (0.003)
SL	BDRJ	<b>0.24 (0.030)</b>	<b>0.32 (0.041)</b>	<b>0.37 (0.060)</b>
	BDMCMC	1.00 (0.032)	1.05 (0.036)	1.08 (0.059)

Values in bold are the ones preferred by the corresponding index

## 5 Discussion

In the setting of graphical models, this work proposed a new class of priors, called *block graph priors*. They allow to include in the model the prior knowledge available about the partition of the nodes. We also introduced a new sampling strategy that leverage these priors to look only for a block structured graph, whose block, if included, have to be complete. In some applications, as the number of variables grows, the importance of each possible dependence loses of interest as it is more natural, and more interpretable, to understand the general structure of dependencies. This is the case of genomics applications as genes may be grouped in pathways, therefore a block structured graph is expected and more interpretable. Another example is market basket analysis which aims to find patterns of association between retailed items so that they can be bundled together to the end of delivering an appealing offer. Finally, we compared our model, on synthetic data, with BDMCMC. In both experiments, BDRJ estimates are better in terms of Std-SHD,  $F_1$ -score and SL.

As future developments, we aim to further develop the BDRJ technique, expand the simulation study by investigating the behaviour of BDRJ when the underlying graph has incomplete blocks and to assess its performances in real world applications. Moreover, experiments are performed using groups of only two nodes. Larger groups imply larger jumps of the chain in the state space and therefore they are less likely to

be accepted. We aim to better investigate the behaviour of our methodology in such cases. We would also like to understand if the proposed methodology could be also extended to Gaussian structured chain graph models for modelling the DAG model induced by the chain components. Finally, we would like to add flexibility to the model by allowing for a random partition of the nodes.

**Acknowledgements** The author is grateful to Raffaele Argiento (Università di Bergamo), Lucia Paci and Alessia Pini (both affiliated to Università Cattolica del Sacro Cuore) for the valuable advice received in the development of the work and during the preparation of the manuscript as well.

## References

1. Atay-Kayis, A., Massam, H.: A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika* **92**, 317–335 (2005)
2. Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A., Nielsen, H.: Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* **16**(5), 412–424 (2000)
3. Codazzi, L., Colombi, A., Gianella, M., Argiento, R., Paci, L., Pini, A.: Gaussian graphical modeling for spectrometric data analysis. *Comput. Stat. Data Anal.* (2022)
4. Cremaschi, A., Argiento, R., De Iorio, M., Shirong, C., Chong, Y.S., Meaney, M.J., Kee, M.Z.: Seemingly unrelated multi-state processes: a Bayesian semiparametric approach. *arXiv preprint arXiv:2106.03072* (2021)
5. Dobra, A., Hans, C., Jones, B., Nevins, J.R., Yao, G., West, M.: Sparse graphical models for exploring gene expression data. *J. Multivariate Anal.* **90**(1), 196–212 (2004)
6. Giudici, P., Castelo, R.: Improving Markov chain Monte Carlo model search for data mining. *Mach. Learn.* **50**(1–2), 127–158 (2003)
7. Giudici, P., Green, P.: Decomposable graphical Gaussian model determination. *Biometrika* **86**(4), 785–801 (1999)
8. James, W., Stein, C.: Estimation with quadratic loss. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 361–379 (1961)
9. Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C., West, M.: Experiments in stochastic computation for high-dimensional graphical models. *Stat. Sci.* **20**, 388–400 (2005)
10. Kullback, S., Leibler, R.A.: On information and sufficiency. *Ann. Math. Stat.* **22**(1), 79–86 (1951)
11. Kumar, S., Ying, J., de Miranda Cardoso, J.V., Palomar, D.P.: A unified framework for structured graph learning via spectral constraints. *J. Mach. Learn. Res.* **21**(22), 1–60 (2020)
12. Lauritzen, S.L.: *Graphical Models*. Oxford University Press, Oxford (1996)
13. Lenkoski, A.: A direct sampler for G-Wishart variates. *Statistics* **2**(1), 119–128 (2013)
14. Mohammadi, A., Wit, E.C.: Bayesian structure learning in sparse Gaussian graphical models. *Bayesian Anal.* **10**(1), 109–138 (2015)
15. Mohammadi, R., Massam, H., Letac, G.: Accelerating Bayesian structure learning in sparse gaussian graphical models. *J. Am. Stat. Assoc.* **0**(0), 1–14 (2021)
16. Mohammadi, R., Wit, E.C.: BDgraph: an R package for Bayesian structure learning in graphical models. *J. Stat. Software* **89**(3), 1–30 (2019). <https://doi.org/10.18637/jss.v089.i03>
17. Osborne, N., Peterson, C., Vannucci, M.: Latent network estimation and variable selection for compositional data via variational EM. *J. Comput. Graph. Stat.* **31**(1), 1–22 (2021)
18. Peterson, C., Stingo, F.C., Vannucci, M.: Bayesian inference of multiple Gaussian graphical models. *J. Am. Stat. Assoc.* **110**(509), 159–174 (2015)
19. Powers, D.M.W.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2**, 2229–3981 (2011)

20. Roverato, A.: Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scand. J. Stat.* **29**(3), 391–411 (2002)
21. Scott, J., Carvalho, C.: Feature-inclusion stochastic search for Gaussian graphical models. *J. Comput. Graph. Statist.* **17**(4), 790–808 (2008)
22. Scutari, M.: On the prior and posterior distributions used in graphical modelling. *Bayesian Anal.* **8**(3), 505–532 (2013)
23. Tsamardinos, I., Brown, L.E., Aliferis, C.F., Moore, A.W.: The max-min hill-climbing Bayesian network structure learning algorithm. *Mach. Learn.* **65**(1), 31–78 (2006)
24. Wang, H.: Sparse seemingly unrelated regression modelling: applications in finance and econometrics. *Comput. Stat. Data Anal.* **54**(11), 2866–2877 (2010)
25. Wang, H.: Scaling it up: stochastic search structure learning in graphical models. *Bayesian Anal.* **10**(2), 351–377 (2015)
26. Wang, H., Li, S.Z.: Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electron. J. Stat.* **6**, 168–198 (2012)
27. Xia, Y., Cai, T., Cai, T.T.: Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions. *J. Am. Stat. Assoc.* **113**(521), 328–339 (2018)