# Bayesian Covariate-Dependent Graph Learning with a Dual Group Spike-and-Slab Prior

Zijian Zeng, Meng Li and Marina Vannucci

Department of Statistics, Rice University, Houston, Texas, USA

September 27, 2024

#### Abstract

Covariate-dependent graph learning has gained increasing interest in the graphical modeling literature for the analysis of heterogeneous data. This task, however, poses challenges to modeling, computational efficiency, and interpretability. The parameter of interest can be naturally represented as a three-dimensional array with elements that can be grouped according to two directions, corresponding to node level and covariate level, respectively. In this article, we propose a novel dual group spikeand-slab prior that enables multi-level selection at covariate-level and node-level, as well as individual (local) level sparsity. We introduce a nested strategy with specific choices to address distinct challenges posed by the various grouping directions. For posterior inference, we develop a tuning-free Gibbs sampler for all parameters, which mitigates the difficulties of parameter tuning often encountered in high-dimensional graphical models and facilitates routine implementation. Through simulation studies, we demonstrate that the proposed model outperforms existing methods in its accuracy of graph recovery. We show the practical utility of our model via an application to microbiome data where we seek to better understand the interactions among microbes as well as how these are affected by relevant covariates.

**Keywords:** Bayesian inference, Gaussian graphical model, global-local prior, human microbiome, variable selection.

# 1 Introduction

Gaussian graphical models have been applied in a wide variety of fields to recover the dependence structure among data (Lauritzen, 1996; Maathuis et al., 2018). The idea dates back to Dempster (1972), who proposed the covariance selection method that estimates conditional independencies based on the inverse covariance matrix (a.k.a., precision matrix or concentration matrix) by linking the absence of an edge in an undirected graph to a zero entry in the precision matrix. Expanding upon this idea, Meinshausen and Buhlmann (2006) showed that neighborhood selection for each node in the graph is equivalent to perform variable selection in a Gaussian linear model, turning the edge detection problem into variable selection for independent regressions. This approach has inspired numerous studies with a focus on using different selection methods to recover edges within a graph (Peng et al., 2009; Leday et al., 2017; Liu and Wang, 2017).

Recent work has demonstrated the value of incorporating covariates in the modeling of subject-specific graphs via Gaussian graphical regression models, in particular for characterizing and discovering interactions in complex biological systems and diseases such as cancer (Ni et al., 2019; Zhang and Li, 2023; Wang et al., 2022; Niu et al., 2023). Most of the existing literature has focused on covariate-adjusted mean structures in Gaussian graphical models, with either constant graphs across subjects or group-specific graphs depending on *discrete* covariates; for a comprehensive review of this rich literature, see Zhang and Li (2023) and Section 1.3 of Wang et al. (2022), with Osborne et al. (2022) providing a recent example. In this article, instead, we focus on modeling the dependence of the precision matrix on covariates, a framework referred to as *precision-on-scalar* regression. This covariate-dependent graph learning task is comparatively much less studied and poses challenges to modeling, computational efficiency, and interpretability. Partition-based Bayesian approaches to model covariate-dependent graphs are explored by (Niu et al., 2023), while Wang et al. (2022) consider an edge regression model for undirected graphs, which estimates conditional dependencies as a function of subject-level covariates, and employs shrinkage priors. Zhang and Li (2023) introduce bi-level sparsity, where element- and group-wise sparsity are encouraged by lasso and group lasso, respectively. Also, Ni et al. (2019) consider a conditional DAG model that allows the graph structure to vary with covariates. Their approach assumes a known hierarchical ordering of the nodes, a prior knowledge that may not always be available in practical settings.

Precision-on-scalar regression models are characterised by an ultra-dimensional parameter space, which can be viewed according to more than one grouping direction, e.g. node or covariate. It is desirable to have both node-level and covariate-level group sparsity, in addition to individual (local) level sparsity. This simultaneous sparsity at the local level and the two group levels is crucial for interpretable graphical models, particularly in the presence of many nodes and covariates. The majority of the existing works on heterogeneous graphs fail to model such structured sparsity, as they typically group parameters in one direction only, and there is a lack of efficient estimation strategies with easy parameter tuning to address the daunting computational challenges. Also, in the work of Zhang and Li (2023), the authors use lasso and group lasso to induce covariate-level sparsity by imposing node-level sparsity. However, relying on one group level to induce the sparsity of another restricts the ability to flexibly capture interactions between the two group levels. Here, we introduce a novel dual-group spike-and-slab prior as a general framework to encode group sparsity at both the covariate and the node level. At the covariate level, we allow for group (global) and individual (local) sparsity. Even though this general prior is complementary to a wide range of existing priors and empowers them into dual-group variants, modeling the two grouping directions in the context of graphical models has distinct challenges. To this end, we propose to use particular choices tailored to each grouping direction, leading to a dual-group spike-and-slab prior well suited for graphical models. We complete our proposed modeling construction with tuning-free posterior sampling, that aids model interpretability. Overall, we are not aware of any work in the Bayesian literature addressing multiple covariates with the aforementioned structured sparsity. Through simulation studies, we demonstrate that the proposed model outperforms the method of Zhang and Li (2023) in its accuracy of graph recovery. We also compare performances to the Bayesian sparse group selection method of Xu and Ghosh (2015).

As an illustration of the utility of our method, we consider an application to multivariate data arising from microbiome studies. The human microbiome has been implicated in many diseases including colorectal cancer, inflammatory bowel disease, and immunologically mediated skin diseases. Here, we apply the proposed method to real data from the Multi-'Omic Microbiome Study-Pregnancy Initiative (MOMS-PI) study, to estimate the interaction between microbes in the vagina, as well as the interplay between vaginal cytokines and microbial abundances, providing insight into mechanisms of host-microbial interaction during pregnancy. These factors influence the microbiome by introducing new organisms, changing the abundance of metabolites, or altering the pH of their environment. Identifying factors that lead to the prevalence of different microbes can improve the understanding of the importance and the function of the microbiome. Our method identifies a large number of microbiome interactions (edges) that are simultaneously influenced by multiple cytokines. It also highlights a subnetwork of multiple microbes that belong to the same family (phylum) and that appear to be consistently detected as having covariate-dependent interactions for various cytokines, which aligns with previous findings.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed prior construction and the sampling procedure. In Section 3, we conduct simulations and compare the proposed approach with existing methods. In Section 4, we apply the proposed model to a human microbiome study. In Section 5 we provide some concluding remarks.

## 2 Methods

#### 2.1 Gaussian Graphical Regression Models with Covariates

Let  $\mathbf{Y} = (Y^1, \ldots, Y^p)$  be a *p*-dimensional outcome vector and  $\mathbf{X} = (X^1, \ldots, X^q)$  a *q*dimensional covariate vector. We denote *N* independent and identically distributed observations by  $\mathbf{y}_n = (y_n^1, \ldots, y_n^p)$  and  $\mathbf{x}_n = (x_n^1, \ldots, x_n^q)$ , for  $n = 1, \ldots, N$ . For simplicity, we assume that the outcomes have been centered with zero mean. The covariate-dependent Gaussian graphical model can be written as

$$\boldsymbol{y}_{n}|\boldsymbol{x}_{n} \sim N_{p}\left(\boldsymbol{0}, \left[\boldsymbol{\Omega}\left(\boldsymbol{x}_{n}\right)\right]^{-1}\right),$$
 (1)

where  $\Omega(\boldsymbol{x}_n) = (\omega^{ij}(\boldsymbol{x}_n))_{i,j=1}^p$ . Similarly to the typical covariate-free setting studied in the Gaussian graphical model literature (Lauritzen, 1996), the covariate-dependent precision

matrix  $\Omega(\mathbf{X})$  encodes independence for node *i* and node *j* given the other nodes  $\mathbf{Y}^{-(i,j)}$ but in a covariate dependent manner as:  $\omega^{ij}(\mathbf{X}) = 0 \iff Y^i \perp Y^j | \mathbf{Y}^{-(i,j)}, \mathbf{X}$ . This adds flexibility to modeling the dependence structure of  $\mathbf{Y}$ .

Under the Gaussian assumption (1) the elements of the precision matrix,  $\omega^{ij}(\boldsymbol{x}_n)$ , are related to the coefficients in the linear regression of  $y_n^i$  on the other  $y_n^j$ ,  $1 \le i \ne j \le p$  as

$$y_n^i = \sum_{j \neq i}^p \theta^{ij} \left( \boldsymbol{x}_n \right) y_n^j + \epsilon_n^i \quad \epsilon_n^i \sim N\left( 0, \sigma_i^2(\boldsymbol{x}_n) \right),$$

where  $\theta^{ij}(\boldsymbol{x}_n) = -\frac{\omega^{ij}(\boldsymbol{x}_n)}{\omega^{ii}(\boldsymbol{x}_n)}, \ \sigma_i^2(\boldsymbol{x}_n) = \frac{1}{\omega^{ii}(\boldsymbol{x}_n)}$ . This model generalizes the standard treatment of Gaussian graphical models (Meinshausen and Buhlmann, 2006; Peng et al., 2009) to a covariate-dependent regime. To complete the model specification, we consider specifying  $\theta^{ij}(\boldsymbol{x}_n)$  and  $\omega^{ii}(\boldsymbol{x}_n)$  using interpretable structures. In particular, we assume  $\omega^{ii}(\boldsymbol{x}_n) = \omega^{ii}$ to be independent of covariates as in Wang et al. (2022) and Zhang and Li (2023), and assume a linear structure  $\theta^{ij}(\boldsymbol{x}_n) = \sum_{k=1}^q \beta_k^{ij} x_n^k$  for  $1 \leq i \neq j \leq p$ . These assumptions lead to the following model

$$y_{n}^{i} = \sum_{j \neq i}^{p} \sum_{k=1}^{q} \beta_{k}^{ij} x_{n}^{k} y_{n}^{j} + \epsilon_{n}^{i}, \quad \epsilon_{n}^{i} \sim N\left(0, \sigma_{i}^{2}\right), \quad 1 \le i \le p, \quad n = 1, ..., N.$$
(2)

## 2.2 Dual group spike-and-slab prior

The conditional regression in Eq. (2) models the effect of covariate  $x_n^k$  on edge (i, j) via coefficient  $\beta_k^{ij}$ . Therefore, sparsity of the regression coefficients  $\beta_k^{ij}$  induces sparsity of the covariate-dependent precision  $\Omega(\mathbf{x}_n)$ . Let us collect the coefficients  $\beta_k^{ij}$  in Eq. (2) into a *p*by-*p*-by-*q* array  $\mathcal{B} = (\beta_k^{ij})$ , for  $i, j = 1, \ldots, p$  and  $k = 1, \ldots, q$ , with diagonal elements  $\beta_k^{ii} = 0$ . The elements of this multi-dimensional array  $\mathcal{B}$  can be grouped in different ways, i.e., as nodelevel and covariate-level groupings. Simultaneous sparsity at the two group levels, as well as locally at individual level, can improve the interpretability and estimability of covariatedependent graphical models, particularly in the case of many nodes and covariates (Zhang and Li, 2023).

Here, we introduce a novel dual-group spike-and-slab prior as a general framework to encode group sparsity at both the covariate and the node level. At the covariate level, we allow for group (global) and individual (local) sparsity. We complete our proposed modeling construction with tuning-free posterior sampling, that aids model interpretability. Even though this general prior is complementary to a wide range of existing priors and empowers them into dual-group variants, the two grouping directions in the context of graphical models have distinct challenges. In our construction, we allow covariate-dependent directed effects between two nodes to be asymmetric; symmetrizing  $\beta_k^{ij}$ , if desired, can be achieved by enforcing the constraint  $\beta_k^{ij} = \beta_k^{ji}$  as in Wang et al. (2022), or via posterior summary, as commonly done in the literature (Meinshausen and Buhlmann, 2006; Zhang and Li, 2023).

We start with a conventional spike-and-slab prior of the type:

$$\beta_k^{ij} | \delta_k^{ij}, \sigma_k^{ij} \sim \delta_k^{ij} N \left[ 0, \left( \sigma_k^{ij} \right)^2 \right] + (1 - \delta_k^{ij}) \delta_0, \tag{3}$$

where  $\delta_k^{ij} \in \{0,1\}$  is the overall selection indicator for a combination of nodes (i,j) and covariate  $k, \sigma_k^{ij}$  represents the prior variance of the slab distribution, and  $\delta_0$  is the Dirac mass at 0 (see (Tadesse and Vannucci, 2021) for a comprehensive treatment of this class of priors). To encode sparsity, we decompose the selection indicator  $\delta_k^{ij}$  into three parts:  $\delta_k^{ij} = \delta^{ij} \times \delta_k \times \gamma_k^{ij}$ , where  $\delta_k$  is the covariate-level effect,  $\delta^{ij}$  is the node-level effect, and  $\gamma_k^{ij}$ represents the local-level effect. For each (i, j, k), the marginal prior on  $\beta_k^{ij}$  is

$$\beta_k^{ij} | \delta^{ij}, \delta_k, \gamma_k^{ij}, \sigma_k^{ij} \sim \delta^{ij} \delta_k \gamma_k^{ij} N \left[ 0, \left( \sigma_k^{ij} \right)^2 \right] + \left( 1 - \delta^{ij} \delta_k \gamma_k^{ij} \right) \delta_0.$$

$$\tag{4}$$

One particular challenge in this *dual-group* approach is to build interpretable group structures into the prior on  $\beta_k^{ij}$  jointly across (i, j, k) beyond the marginal specification in (4), which not only should encode two group structures but also account for the distinct challenges posed by high-dimensional graphical models.

The two sets of group indicators  $(\delta^{ij})$  and  $(\delta_k)$  are symmetric in (4) in that no particular order between the two groups is enforced when combined with the local-level indicator  $\gamma_k^{ij}$ . The spike-and-slab specification allows us to consider any sequential order of them, leading to a notion of nested decomposition of the two groups. Without loss of generality, below we focus on the sequence of first  $\delta^{ij}$  then  $\delta_k$ , and discuss the different corresponding model structures along with our specification of each group sparsity. To this end, we let  $\tau_k^{ij} = \delta_k \gamma_k^{ij} \sigma_k^{ij}$  and reparameterize Eq (4) as

$$\beta_k^{ij} | \delta^{ij}, \tau_k^{ij} \sim \delta^{ij} N \left[ 0, \left( \tau_k^{ij} \right)^2 \right] + \left( 1 - \delta^{ij} \right) \delta_0, \tag{5}$$

and describe our dual-group structured prior specification below.

#### 2.2.1 Node-level group sparsity: outer-layer structured prior for scalar response

For a given pair of nodes (i, j), let the coefficient vector  $\mathbf{B}^{ij} = (\beta_k^{ij})_{1 \le k \le q}$  indicate the coefficients grouped based on the paired indices (i, j). The model in Eq. (2) leades to

$$y_n^i = \sum_{j \neq i}^p (\boldsymbol{x}_n^T \boldsymbol{B}^{ij}) y_n^j + \varepsilon_n^i,$$

where  $\mathbf{B}^{ij} = \mathbf{0}$  implies that node j has no effect on node i. The group sparsity at the node-level is described by the sparsity of vector  $\mathbf{B}^i = (\mathbf{B}^{ij})_{j\neq i}^{1\leq j\leq p}$  with group label j. One challenge in defining a prior on  $\mathbf{B}^{ij}$  is the need to achieve sparsity at both the group level and individual level, with the added difficulty when addressing dual group sparsity.

Note that the model above is a *high-dimensional linear regression model* with standard scalar response, for which a rich menu of group priors has been proposed, such as Stingo et al. (2011); Xu and Ghosh (2015); Bai et al. (2022). We propose to use the multivariate spike-and-slab prior for group sparsity in the outer layer of our model:

$$\boldsymbol{B}^{ij} = \operatorname{diag}\left(\tau_1^{ij}, ..., \tau_q^{ij}\right) \boldsymbol{b}^{ij}, \quad 1 \le i \ne j \le p,$$
(6)

$$\begin{cases} \boldsymbol{b}^{ij} | \delta^{ij} \sim \delta^{ij} MVN\left(\boldsymbol{0}_{q}, \boldsymbol{I}_{q}\right) + (1 - \delta^{ij}) \,\delta_{\boldsymbol{0}_{q}}, \\ \delta^{ij} | \pi^{i} \sim \text{Bernoulli}\left(\pi^{i}\right), \quad \pi^{i} \sim \text{Beta}\left(a^{i}, b^{i}\right), \end{cases}$$
(7)

where  $\mathbf{b}^{ij} = (b_1^{ij}, ..., b_q^{ij})^T$ . Conditional on  $\tau_k^{ij}$  for k = 1, ..., q, Eqs. (6) and (7) provide node-level selection for the paired indices (i, j) for all k. The node-level indicator  $\delta^{ij}$  models the group effect of node j on node i through all covariates. For every j, if  $\delta^{ij} = 0$  then effects  $\mathbf{b}^{ij}$  will be excluded from the model, implying that node j does not affect node i through any of the covariates, i.e.,  $\mathbf{B}^{ij} = \mathbf{0}$ . The parameter  $\pi^i$  can be interpreted as the prior probability that  $y^i$  is affected by the other nodes.

# 2.2.2 Covariate-level group and local sparsity: inner-layer structured prior for multivariate response

For a given covariate  $x^k$ , let the coefficient matrix  $\mathbf{B}_k = (\beta_k^{ij})^{1 \le i,j \le p}$  indicate the coefficients grouped based on the index k. One challenge in modeling the sparsity of matrix  $\mathbf{B}_k$  is to achieve simultaneous group sparsity and individual sparsity in an interpretable manner. This calls for a strategy different from the treatment of node-level sparsity. To see this, the independent regression system from Eq. (2) has the representation

$$\boldsymbol{y}_n = \sum_{k=1}^q \left( \boldsymbol{B}_k \boldsymbol{y}_n 
ight) x_n^k + \boldsymbol{\varepsilon}_n$$

which, unlike the preceding node-level representation, is a high-dimensional vector-on-scalar regression model with multivariate response, where  $\mathbf{B}_k = \mathbf{0}$  implies that covariate  $x^k$  has no influence on the precision matrix. We propose to jointly model  $(\delta_k, \gamma_k^{ij})$  by

$$\delta_k = I(\pi_k \ge d_k), \quad \gamma_k^{ij} | \pi_k \sim \text{Bernoulli}(\pi_k), \quad \pi_k \sim \text{Beta}(a_k, b_k).$$
(8)

This global-local structure has been recently advocated by Zeng et al. (2024) in a different setting when studying image-on-scalar regression, which is the inner layer of our prior. The global level indicator  $\delta_k$  represents the covariate-level selection, as  $\delta_k = 0$  zeros out  $\tau_k^{ij}$  for any  $1 \leq i \neq j \leq p$ , eliminating the covariate from the model. That is,  $\delta_k = 0$  implies that covariate  $x^k$  has no influence on any of the edges, hence the whole graph. At the local level  $\gamma_k^{ij}$  refers to the influence of the covariate on the pair of nodes (i, j). The importance of a covariate is characterized by the total number of pairs influenced by that covariate. The parameter  $\pi_k$ , which can be interpreted as the probability that  $x^k$  has an influence on the graph, is called *participation rate* in Zeng et al. (2024) and is estimated under the assumption that the covariate affects all pairs independently. The participation rate  $\pi_k$  also informs the selection by excluding those covariates expected to affect less than  $d_k \times 100\%$  pairs, saying  $\tau_k = 0$ , if  $\pi_k < d_k$ , leading to  $B_k = 0$ . This hard-threshold provides a probability-based selection rule and uses the local-level selection to inform the global-level selection. Without prior domain knowledge, Zeng et al. (2024) recommend  $d_k = 0.05$  (i.e., 5%) as a conventional probability threshold for sparse models.

Eqs. (6), (7), and (8) lead to a dual-group spike-and-slab prior. The prior sparsity encoded therein can be obtained by calculating the expectation of the selection indicator

$$\delta_{k}^{ij} = [\delta_{k} \times \gamma_{k}^{ij}] \times \delta^{ij} = [I(\pi_{k} \ge d_{k}) \times \gamma_{k}^{ij}] \times [\delta^{ij}]:$$

$$E[\delta_{k}^{ij}] = E[I(\pi_{k} \ge d_{k})\gamma_{k}^{ij}\delta^{ij}] = E_{\pi_{k}} [E[I(\pi_{k} \ge d_{k})\gamma_{k}^{ij}] |\pi_{k}] E_{\pi^{i}} [E[\delta^{ij}] |\pi^{i}]$$

$$= E_{\pi_{k}} [I(\pi_{k} \ge d_{k})\pi_{k}] E_{\pi^{i}} [\pi^{i}] = \int_{d_{k}}^{1} \pi_{k} \frac{1}{B(a_{k}, b_{k})} \pi_{k}^{a_{k}-1} (1-\pi_{k})^{b_{k}-1} d\pi_{k} \frac{a^{i}}{a^{i}+b^{i}}$$

$$= E[\pi_{k}] [1-F_{\text{Beta}_{k}}(d_{k})] E[\pi^{i}] = \frac{a^{i}}{a^{i}+b^{i}} \frac{a_{k}}{a_{k}+b_{k}} [1-F_{\text{Beta}_{k}}(d_{k})], \qquad (9)$$

where  $F_{\text{Beta}_k}(\cdot)$  is the cumulative distribution function of  $\text{Beta}(a_k + 1, b_k)$ . Eq. (9) offers a flexible way to incorporate prior beliefs on the graphical structure. For example, a possible belief is that the graph consists of a dense population level and a sparse covariate level, which corresponds to an always-included intercept term with imbalanced penalization for the intercept and covariates, as seen in Zhang and Li (2023). Although our model does not intentionally incorporate an intercept, it can adapt to this belief by incorporating  $x_n^1 = 1$ for all n and adjusting the prior parameters  $d_k$ ,  $a_k$  and  $b_k$ , i.e.,  $d_1 = 0$  and  $\frac{a_1}{a_1+b_1} > \frac{a_k}{a_k+b_k}$  for  $k \neq 1$ . Similarly, prior parameters  $a^i$  and  $b^i$  can be adjusted to adapt to beliefs on node sparsity. Without any prior information, a non-informative prior can be used, allowing the model to learn from the data.

#### 2.2.3 Complete tuning-free prior specification

We place a prior  $\mathcal{F}_k^{ij}$  with positive support on  $\sigma_k^{ij}$ . In particular, we specify  $\mathcal{F}_k^{ij} = N^+(0, s_k^2)$ , ie. a truncated normal on the positive line. Combining this with Eq. (8), our prior on  $\tau_k^{ij}$  is

$$\begin{cases} \tau_k^{ij} = \tilde{\tau}_k^{ij} I\left(\pi_k \ge d_k\right) \\ \tilde{\tau}_k^{ij} |\gamma_k^{ij} \sim \gamma_k^{ij} N^+\left(0, s_k^2\right) + \left(1 - \gamma_k^{ij}\right) \delta_0 \\ \gamma_k^{ij} |\pi_k \sim \text{Bernoulli}(\pi_k), \quad \pi_k \sim \text{Beta}(a_k, b_k). \end{cases}$$
(10)

Finally, we complete the model by assigning conjugate priors  $s_k^2 \sim \text{InvGamma}(1, t)$ , with  $t \sim \text{Gamma}(a_t, b_t)$ , for  $k = 1 \dots, q$ , and inverse gamma priors on the error variances,  $\sigma_i^2 \sim \text{InvGamma}(a_{\sigma}^i, b_{\sigma}^i)$ , for  $i = 1, \dots, p$ . We set  $a_t = b_t = 0$ , which leads to a commonly used flat and improper prior on t, although other values can be used. These specifications, along with Eqs. (6), (7) and (10), complete our prior model.

We conclude this section by noting that in the presentation of our proposed dual-group spike-and-slab prior we have used the sequential order of nesting a vector-on-scalar regression into a scalar-on-scalar regression. In practice, one can also vary this order and substitute particular choices for each module with alternative structures, following the same nesting strategy to address multi-level sparsity.

### 2.3 Posterior Inference

We derive a tuning-free full Gibbs sampler for inference in the proposed model, which combines blocked Gibbs strategies based on the samplers used in Xu and Ghosh (2015) and Zeng et al. (2024). We describe the updates of the parameters below and provide detailed derivations in the supplementary materials.

• Update the covariate-level selection parameters  $\{\tau_k^{ij}, \tilde{\tau}_k^{ij}, \gamma_k^{ij}, \pi_k\}$ 

We rewrite the distribution of response node i in Eq. (2) by separating the parameters to be sampled as follows:

$$y_n^i| - \sim N\left(\underbrace{\sum_{\substack{s \neq k}} \sum_{\substack{l \neq i}} \beta_s^{il} y_n^l x_n^s}_{\text{denoted as } c_n^{1,ijk}} + \underbrace{\sum_{\substack{j' \notin \{i,j\}\\ \text{conditional on } s \neq k\\ \text{denoted as } c_n^{2,ijk}}}_{\text{conditional on } j' \notin \{i,j\}} + \beta_k^{ij'} y_n^{j'} x_n^k + \beta_k^{ij} y_n^j x_n^k, \sigma_i^2\right).$$
(11)

By denoting  $y_n^{ijk} = y_n^i - c_n^{1,ijk} - c_n^{2,ijk}$ . Eqs. (10) and (11) lead to the following conditional probabilities for the latent coefficients and indicators,

$$y_n^{ijk} |\tilde{\tau}_k^{ij}, \gamma_k^{ij} = 1, - \sim N\left(\tilde{\tau}_k^{ij} b_k^{ij} y_n^j x_n^k, \sigma_i^2\right)$$
$$\tilde{\tau}_k^{ij} |\gamma_k^{ij} = 1 \sim N^+\left(0, s_k^2\right)$$
$$y_n^{ijk} |\gamma_k^{ij} = 0, - \sim N\left(0, \sigma_i^2\right)$$
(12)

with corresponding Bayes factor, integrating out  $\tilde{\tau}_k^{ij},$  given as

$$\theta_{k}^{ij} = \frac{p\left(y_{\cdot}^{ijk} | \gamma_{k}^{ij} = 0, \tilde{\tau}_{k}^{ij} = 0\right) \times (1 - \pi_{k})}{\int p\left(y_{\cdot}^{ijk} | \gamma_{k}^{ij} = 1, \tilde{\tau}_{k}^{ij}\right) p\left(\tilde{\tau}_{k}^{ij}\right) d\tilde{\tau}_{k}^{ij} \times \pi_{k}}$$
$$= \frac{1 - \pi_{k}}{2\left(s_{k}^{2}\right)^{-\frac{1}{2}} \times \left(\tilde{\nu}_{ijk}^{2}\right)^{\frac{1}{2}} \exp\left\{\frac{1}{2}\frac{\tilde{m}_{ijk}^{2}}{\tilde{\nu}_{ijk}^{2}}\right\} \Phi\left(\frac{\tilde{m}_{ijk}}{\tilde{\nu}_{ijk}}\right) \times \pi_{k}},$$

where  $y_{\cdot}^{ijk}$  denotes  $\{y_n^{ijk}\}_{n=1}^N$ ,  $\Phi(\cdot)$  denotes the cumulative distribution function of the standard Normal distribution, and

$$\tilde{\nu}_{ijk}^2 = \left(\sum_{n=1}^N \left(y_n^j x_n^k\right)^2 \left(b_k^{ij}\right)^2 / \sigma_i^2 + 1/s_k^2\right)^{-1} \text{ and } \tilde{m}_{ijk} = \tilde{\nu}_{ijk}^2 b_k^{ij} \sum_{n=1}^N y_n^j x_n^k y_n^{ijk} / \sigma_i^2.$$

This Bayes factor allows to sample the local indicators  $\gamma_k^{ij}$  from

$$\gamma_k^{ij}| - \sim \operatorname{Bernoulli}\left(\frac{1}{1 + \theta_k^{ij}}\right).$$

If  $\gamma_k^{ij} = 1$ , Eq. (12) leads to the update  $\tilde{\tau}_k^{ij} | - \sim N^+ \left( \tilde{m}_{ijk}, \tilde{\nu}_{ijk}^2 \right)$ ; else if  $\gamma_k^{ij} = 0$ , we set  $\tilde{\tau}_k^{ij} = 0$ .

After updating all indicators for covariate  $x^k$ , we update

$$\pi_k | - \sim \text{Beta}\left(a_k + \sum_{1 \le i \ne j \le p} \gamma_k^{ij}, b_k + p(p-1) - \sum_{1 \le i \ne j \le p} \gamma_k^{ij}\right)$$

leading to  $\tau_k^{ij} = \tilde{\tau}_k^{ij} \delta_k = \tilde{\tau}_k^{ij} I(\pi_k \ge d)$ . This joint update of parameters and selection indicators avoids reversible jump (Savitsky et al., 2011).

• Update the node-level selection parameters  $\{b^{ij}, \delta^{ij}, \pi^i\}$  together with  $\{\beta_k^{ij}\}$ We rewrite the distribution of response node *i* in Eq. (2) by separating the parameters to be sampled as follows:

$$y_{n}^{i}| - \sim N\left(\sum_{j'\notin\{i,j\}}\sum_{k=1}^{q}\beta_{k}^{ij'}y_{n}^{j'}x_{n}^{k} + \sum_{k=1}^{q}\beta_{k}^{ij}y_{n}^{j}x_{n}^{k}, \sigma_{i}^{2}\right).$$
(13)

Denoting  $z_n^{ij} = y_n^i - \sum_{j' \notin \{i,j\}} \sum_{k=1}^q \beta_k^{ij'} y_n^{j'} x_n^k$ , eqs. (7) and (13) lead to the following conditional probability distributions for the latent coefficients and indicators

$$z_n^{ij} | \delta^{ij} = 1, - \sim N\left( \left( \boldsymbol{X}_n^{ij} \right)^T \boldsymbol{V}^{ij} \boldsymbol{b}^{ij}, \sigma_i^2 \right)$$
$$\boldsymbol{b}^{ij} | \delta^{ij} = 1 \sim MVN\left( \boldsymbol{0}_q, \boldsymbol{I}_q \right)$$
$$z_n^{ij} | \delta^{ij} = 0, - \sim N\left( 0, \sigma_i^2 \right)$$
(14)

where  $\boldsymbol{X}_{n}^{ij} = (y_{n}^{j}x_{n}^{1}, ..., y_{n}^{j}x_{n}^{q})^{T}$  and  $\boldsymbol{V}^{ij} = \operatorname{diag}\left(\tau_{1}^{ij}, ..., \tau_{q}^{ij}\right)$ . In addition, we denote  $\boldsymbol{Z}^{ij} = \left(z_{1}^{ij}, ..., z_{n}^{ij}\right)^{T}$  and  $\boldsymbol{X}^{ij} = \left(\boldsymbol{X}_{1}^{ij}, ..., \boldsymbol{X}_{n}^{ij}\right)^{T}$ . The Bayes factor of edge (i, j) can be

obtained by integrating out  $\boldsymbol{b}^{ij}$ ,

$$\theta^{ij} = \frac{p\left(\boldsymbol{Z}^{ij} | \delta^{ij} = 0, \boldsymbol{b}^{ij} = \boldsymbol{0}\right) \times (1 - \pi^{i})}{\int p\left(\boldsymbol{Z}^{ij} | \delta^{ij} = 1, \boldsymbol{b}^{ij}\right) p\left(\boldsymbol{b}^{ij}\right) d\boldsymbol{b}^{ij} \times \pi^{i}}$$
$$= \frac{1 - \pi^{i}}{\left|\tilde{\boldsymbol{\Sigma}}^{ij}\right|^{\frac{1}{2}} \exp\left\{\frac{1}{2}\left(\tilde{\boldsymbol{\mu}}^{ij}\right)^{T}\left(\tilde{\boldsymbol{\Sigma}}^{ij}\right)^{-1}\tilde{\boldsymbol{\mu}}^{ij}\right\} \times \pi^{i}},$$

where  $\tilde{\boldsymbol{\Sigma}}^{ij} = \left(\frac{1}{\sigma_i^2} \left(\boldsymbol{X}^{ij} \boldsymbol{V}^{ij}\right)^T \left(\boldsymbol{X}^{ij} \boldsymbol{V}^{ij}\right) + \boldsymbol{I}_q\right)^{-1}$  and  $\tilde{\boldsymbol{\mu}}^{ij} = \left(\frac{1}{\sigma_i^2} \left(\boldsymbol{Z}^{ij}\right)^T \boldsymbol{X}^{ij} \boldsymbol{V}^{ij} \tilde{\boldsymbol{\Sigma}}^{ij}\right)^T$ . This Bayes factor allows to sample the indicators  $\delta^{ij}$  as

$$\delta^{ij}| - \sim \operatorname{Bernoulli}\left(\frac{1}{1+\theta^{ij}}\right).$$

Then, if  $\delta^{ij} = 1$ , we update  $\mathbf{b}^{ij}|_{-} \sim MVN\left(\tilde{\boldsymbol{\mu}}^{ij}, \tilde{\boldsymbol{\Sigma}}^{ij}\right)$ ; otherwise if  $\gamma^{ij} = 0$ , we set  $\mathbf{b}^{ij} = \mathbf{0}$ . The update of  $(\beta_k^{ij})_{1 \leq k \leq q}$  is followed by  $\mathbf{B}^{ij} = \mathbf{V}^{ij} \mathbf{b}^{ij}$ .

After all indicators for node i are updated, the probability  $\pi^i$  can be updated by

$$\pi^i | - \sim \text{Beta}\left(a^i + \sum_{j \neq i} \delta^{ij}, b^i + (p-1) - \sum_{j \neq i} \delta^{ij}\right).$$

• Update the variances  $\{\sigma_i^2\}$ 

This is a conjugate update

$$\sigma_i^2 | - \sim \text{InvGamma}\left(\frac{N}{2} + a_\sigma, \frac{1}{2}\sum_{n=1}^N \left(y_n^i - \sum_{j\neq i}\sum_{k=1}^q \beta_k^{ij} y_n^j x_n^k\right)^2 + b_\sigma\right).$$

• Update  $\{s_k^2\}$  and t:

These are also conjugate updates. For each  $k \in [q]$ , we sample

$$s_k^2| - \sim \operatorname{InvGamma}\left(1 + \frac{1}{2}\sum_{1 \le i \ne j \le p} \gamma_k^{ij}, t + \frac{1}{2}\sum_{1 \le i \ne j \le p} \left(\tilde{\tau}_k^{ij}\right)^2\right).$$

After all  $s_k$  are updated, we sample

$$t|-\sim \operatorname{Gamma}\left(q+1,\sum_{k=1}^{q}\frac{1}{s_{k}^{2}}\right).$$

Given the MCMC samples, we perform posterior inference by calculating the marginal posterior probabilities of inclusion (MPPIs) of the indicators  $\delta_k^{ij}$ . Following the median

probability model (Barbieri and Berger, 2004; Zeng et al., 2024), we define the inclusion indicator  $\kappa_k^{ij} = 1$  if the MPPIs of  $\delta_k^{ij}$  is greater than 0.5. To infer the edges in the undirected graph, we use the "OR" rule (Meinshausen and Buhlmann, 2006; Zhang and Li, 2023), which concludes that the edge (i, j) is affected by covariate  $x_k$  if either  $\kappa_k^{ij} = 1$  OR  $\kappa_k^{ji} = 1$ . At the two group level, we conclude that the edge (i, j) exists if either  $\sum_k \kappa_k^{ij} \neq 0$  OR  $\sum_k \kappa_k^{ji} \neq 0$ , and that the covariate  $x^k$  is influential if  $\sum_{ij} \kappa_k^{ij} \neq 0$ . Bayesian false discovery rate control methods can also be utilized in determining the  $\kappa_k^{ij}$ 's (Newton et al., 2004).

## 3 Simulation Study

In this section, we conduct simulations and compare the proposed approach with selected covariate-dependent Gaussian graphical regression approaches.

## **3.1** Data Generation

We generate data from Eq (1) using  $\Omega(\mathbf{x}_n) = (\omega^{ij}(\mathbf{x}_n))_{i,j=1}^p = (\sum_{k=1}^q \beta_k^{ij} x_n^k)_{i,j=1}^p$ . We set the number of nodes to p = 25 and the number of covariates to q = 10, and introduce sparsity as follows. First, we generate a 25-by-25 random graph  $\mathcal{G}$  with sparsity-level at 0.4 and randomly divide the edges into four parts for  $\mathbf{B}_{1-4}$ . We set  $\mathbf{B}_{5-10} = \mathbf{0}$  for the empty covariates. The values of the non-zero entries  $\beta_k^{ij}$  are sampled from uniform distributions supported on the intervals  $[-0.5, -0.35] \cup [0.35, 0.5]$ . To generate a valid precision matrix, we follow Zhang and Li (2023) by first rescaling each row *i* by dividing by  $\frac{1}{2} \sum_j \sum_k |\beta_k^{ij}|$ and then averaging  $\beta_k^{ij}$  and  $\beta_k^{ji}$  to fill each entry of (i, j, k). We set  $X_1 = 1$  as the intercept and sample  $X_{2-10}$  from uniform[0, 1]. We use two different sample sizes n = 200, 500 and repeatedly generate data 50 times for each size to evaluate the model performances.

## 3.2 Comparison Study

We compare performances of the proposed method, for which we use the acronym DGSS, to Lasso regression (Tibshirani, 1996), the GMMReg method of Zhang and Li (2023) and the Bayesian sparse group selection method with spike and slab prior of Xu and Ghosh (2015). We implement Lasso by the R-package glmnet and GMMReg by the Matlab code from the authors' page. For both Lasso and GMMReg, we tune the regularization parameters by cross-validation. We denote the Bayesian sparse group selection with spike and slab prior as BSGSSS, implemented by the R-package MBSGS. Lasso only considers local sparsity, whereas the other methods also take group sparsity into account. Specifically, GMMReg uses node-level sparsity to induce covariate-level sparsity, BSGSSS accounts for node-level sparsity only, and the proposed DGSS decouples node-level and covariate-level sparsity, modeling them with the dual group spike-and-slab prior. For DGSS, we run 20,000 MCMC iterations with a burn-in of 10,000. We specify non-informative priors as  $\pi_k \sim \text{Beta}(1,1)$ ,  $\pi^i \sim \text{Beta}(1,1)$  and use the conventional sparsity threshold  $d_k = 0.05$  for all  $k = 1, \ldots, q$ . For BSGSSS, we increase the MCMC iterations from default 10,000 with a burn-in of 5,000 to 20,000 with a burn-in of 10,000 and keep all other parameters set to their default values.

We consider the covariate-dependent edge detection as a classification task with the presence of an edge within each precision coefficient being treated as a positive signal. The total number of parameters is p(p-1)q = 6,000, and on average,  $p(p-1) \times 0.4 = 240$  of them are signals, which may vary due to random graph generation. The covariate-dependent edge further provides inference to the node-level and covariate-level selections, as described in Section 2.3. We report the following four metrics for comparison: True Positive Rate (TPR), False Positive Rate (FPR), F1 score (F1), and Matthews correlation coefficient (MCC). These metrics are defined by:

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN}, \quad F1 = \frac{TP}{TP + 1/2 \times (FP + FN)},$$
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{TP + FP} \times \sqrt{TP + FN} \times \sqrt{TN + FP} \times \sqrt{TN + FN}},$$

where TP, FP, TN and FN represent numbers of True Positive, False Positive, True Negative and False Negative, respectively.

#### 3.3 Results

Figure 1 shows examples of the adjacency matrices corresponding to the precision coefficients  $B_k$ , k = 1, ..., q, for one simulated dataset with n = 200, 500. For both sample sizes,



(b) n = 500

Figure 1: Adjacency matrices corresponding to the precision coefficients  $B_{1:10}$  (left to right), from one simulated dataset with sample size n = 200 (top) and n = 500 (bottom). In each plot, the five rows are the true matrices (1st row), and estimated matrices by Lasso (2nd row), GMMReg (3rd row), BSGSSS (4th row) and DGSS (5th row).

we observe that Lasso and BSGSSS do not penalize the coefficients sufficiently, failing to eliminate covariates  $X_{5-10}$ . On the contrary, GMMReg penalizes the coefficients excessively, selecting too few edges. This phenomenon may be caused by their sparsity assumptions. Lasso and BSGSSS totally ignore the covariate-level sparsity by treating the problem as a high-dimensional linear regression at each node, ultimately failing to exclude those covariates with no impact on edges. On the other hand, GMMReg assumes a dense intercept and sparse covariates, and tunes the parameters via cross-validation, which fails to find an optimal penalty for both dense intercept and covariates under their assumption, ultimately limiting the number of edges selected. Compared to the three existing methods, the proposed DGSS achieves relatively good sparse estimates, identifying the important covariates with a reasonable sparsity level.

We now proceed to assess each method using the edge and covariate selection metrics introduced above. All results are reported in Table 1, averaged across 50 replicates. For covariate-dependent edge detection, performance metrics tend to be low for all methods, even with the relatively large sample size n = 500. In this scenario, Lasso only considers the local-level penalty, while the other methods, GMMReg, BSGSSS and DGSS include at least two level of selection/penalties. With sample size n = 200, GMMReg and BSGSSS outperform Lasso in terms of F1 and MCC, because their second level selection/penalty can efficiently exclude the empty coefficients. On the contrary, with the relatively large sample size, n = 500, their performances become comparable to the Lasso in terms of those two scores as their second level of selection/penalty does not fit the sparsity pattern in the data-generation process. Meanwhile the proposed DGSS method takes advantage of the multi-level selection and outperforms the other methods in terms of F1 and MCC.

Next, we evaluate the models' performance in edge detection for the *overall graph*, which we define as the graph where an individual edge is present if affected by any of the covariates  $x^k$ . This corresponds to group edge selection at the node level. There are p - 1 = 24 edges at the node group level, and on average  $(p - 1) \times 0.4 = 9.6$  of them are signals. Results in Table 1 show comparable performance across all methods. Although the GMMReg seems to have a higher F1 score than other methods when the sample size is n = 200, this appears to be due to higher TPR score at the cost of higher FPR score. As evidence, its MCC scores are close to those of other methods, and hence, we do not conclude that there is a significant outperformance. Finally, we look at the task of covariate selection based on the precision coefficient estimates  $\hat{B}_k$ . Four out of ten covariates are signals. We select  $X_k$  if  $\hat{B}_k \neq 0$ . From Table 1 we observe the failure of the Lasso and BSGSSS methods in this task. Without sufficient selection/penalties, these two methods include all covariates in almost all replicates, as also illustrated in Figures 1a and 1b, with the exception of 2 out of 100 cases. Although GMMReg tends to favor a dense precision coefficient for the intercept and sparse precision coefficients for the other covariates, its performance is similar to that of DGSS. The proposed DGSS, on the other hand, outperforms GMMReg in terms of all averaged metrics. Table 1: Performance for edge and covariate detection. Results are averaged over 50 repeated simulations, with standard errors reported in parentheses.

	sample size $n = 200$				sample size $n = 500$				
	Lasso	GMMReg	BSGSSS	DGSS	Lasso	GMMReg	BSGSSS	DGSS	
TPR	0.146	0.228	0.173	0.271	 0.429	0.303	0.508	0.679	
	(0.009)	(0.007)	(0.009)	(0.013)	(0.016)	(0.005)	(0.012)	(0.012)	
FPR	0.023	0.025	0.017	0.025	 0.043	0.028	0.062	0.066	
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	
F1	0.164	0.246	0.211	0.283	 0.341	0.307	0.337	0.414	
	(0.007)	(0.006)	(0.007)	(0.009)	(0.007)	(0.005)	(0.004)	(0.004)	
MCC	0.142	0.220	0.197	0.260	 0.318	0.279	0.322	0.417	
	(0.007)	(0.007)	(0.006)	(0.009)	(0.009)	(0.005)	(0.005)	(0.005)	

(a) Covariate-dependent edge detection

(b) Group edge detection	in th	e overall	graph
--------------------------	-------	-----------	-------

	sample size $n = 200$				sample size $n = 500$ LassoGMMRegBSGSSSDGSS $0.658$ $0.748$ $0.648$ $0.780$ $(0.014)$ $(0.006)$ $(0.011)$ $(0.010)$ $0.074$ $0.123$ $0.056$ $0.132$ $(0.005)$ $(0.007)$ $(0.003)$ $(0.005)$ $0.725$ $0.771$ $0.742$ $0.792$				
	Lasso	GMMReg	BSGSSS	DGSS		Lasso	GMMReg	BSGSSS	DGSS
TPR	0.345	0.552	0.266	0.412		0.658	0.748	0.648	0.780
	(0.016)	(0.011)	(0.011)	(0.013)		(0.014)	(0.006)	(0.011)	(0.010)
FPR	0.054	0.160	0.018	0.060		0.074	0.123	0.056	0.132
	(0.004)	(0.005)	(0.002)	(0.005)		(0.005)	(0.007)	(0.003)	(0.005)
F1	0.470	0.605	0.404	0.538		0.735	0.771	0.743	0.782
	(0.015)	(0.008)	(0.013)	(0.011)		(0.009)	(0.004)	(0.008)	(0.005)
MCC	0.382	0.412	0.380	0.433		0.622	0.636	0.639	0.650
	(0.012)	(0.011)	(0.010)	(0.010)		(0.010)	(0.009)	(0.010)	(0.008)

(c) Covariate selection

	sample size $n = 200$					sample size $n = 500$				
	Lasso	GMMReg	BSGSSS	DGSS	Las	so	GMMReg	BSGSSS	DGSS	
TPR	1	0.845	1	0.925	1		0.895	1	1	
	(0)	(0.030)	(0)	(0.019)	(0	)	(0.022)	(0)	(0)	
FPR	0.993	0.463	0.993	0.327	1		0.303	1	0.327	
	(0.005)	(0.036)	(0.005)	(0.031)	(0	)	(0.039)	(0)	(0.029)	
F1	0.573	0.663	0.573	0.776	0.5	$71^{-1}$	0.775	0.571	0.816	
	(0.001)	(0.018)	(0.001)	(0.020)	(0	)	(0.018)	(0)	(0.015)	
MCC	0.272	0.419	0.272	0.605			0.637		0.686	
	(0)	(0.035)	(0)	(0.038)	-		(0.031)	-	(0.026)	

## 4 Application to Microbiome Data

We demonstrate the proposed method with data from the Multi-Omic Microbiome Study: Pregnancy Initiative (MOMS-PI), a study funded by the NIH Roadmap Human Microbiome Project to understand the impact of the vaginal microbiome on pregnancy and the fetal microbiome. This study contains samples from multiple body sites, including mouth, skin, vagina and rectum, of 596 subjects throughout pregnancy and for a short term after childbirth. Previous research found the vaginal microbiome can change early in pregnancy and be predictive of pregnancy outcomes (Serrano et al., 2019; Fettweis et al., 2019).

#### 4.1 Data

Data from the MOMS-PI study is publicly available and can be found in the R package HMP2Data. Following Osborne et al. (2022), we focus on the interplay between microbial abundances and vaginal cytokines, a mechanism by which the host regulates the composition of the vaginal microbiome, and use the first baseline visit data of the n = 225 subjects whose microbiome and cytokine profiling of the vagina are available among the 596 subjects enrolled in the study. Furthermore, we consider p = 90 OTUs whose absolute abundance is greater than 1 in at least 10% of the subjects and use all the 29 available cytokines as covariates, adding an intercept term, which implies q = 30. We apply the centered log ratio transformation to normalize the abundance counts (Aitchison, 1982; Gloor et al., 2017; Lin and Peddada, 2020), as commonly done in Gaussian graphical modeling for microbiome data to satisfy the Gaussian assumption (Kurtz et al., 2015; Wilms and Bien, 2022). After transformation, we center the data such that each OTU has zero mean. For the covariates, we transform the data to the log scale and use the min-max normalization, so that values fall within the [0, 1] interval.

#### 4.2 Results

We use the same non-informative prior specifications and MCMC settings as in the simulation study. Given the results obtained in the simulation study, we restrict comparisons to the GMMReg and the proposed DGSS methods. On a server, with two 20-core 2.4 GHz Intel(R) Xeon CPUs, running the MCMC algorithm of our DGSS method, coded in Rcpp, took about 15 seconds per iteration.

Table 2 reports the number of covariate-dependent edges selected by DGSS and GMM-Reg, for each covariate and for the overall graph. Similar to the simulation study, we observe that GMMReg tends to select zero edges for almost all covariates, and a few more for the intercept (baseline). On the other hand, DGSS selects significantly more edges than GMM-Reg, not only for each covariate but also in the overall graph. Figure 2 shows the adjacency matrices of the graphs corresponding to the precision coefficients  $B_k$  of the four covariates with the most covariate-dependent edges and the four covariates with the least covariatedependent edges. We observe that a large number of edges are simultaneously influenced by multiple cytokines. Additionally, it appears that some edges remain within a block of the OTU 1-26 across covariates, which aligns with a finding from Osborne et al. (2022), as discussed next.

IL-12 FGF IFN-g Baseline Eotaxin G-CSF GM-CSF IL-10 IL-13 IL-15 IL-17A IL-1b IL-1ra IL-2IL-4 (p70) GMMReg 97 0 0 0 0 0 0 0 0 0 0 0 0 0 1 DGSS 172201104 105190167 16310418913214612320419587 MCP-1 MIF MIF PDGF RAN-TES TNF FGF VEGF IP-10 IL-17  $IL_{-5}$ IL-6 IL-7IL-8 IL-9 (MCAF (1a) (1b) (bb)(a) basic GMMReg 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 DGSS 154165130156 170 192186 100 147 167 127 38 93 2125

Table 2: Number of selected covariate-dependent edges

Figure 3 shows the adjacency matrix of the overall graph selected by DGSS, together with a plot showing the commonly selected edges with the graph selected in Osborne et al. (2022). OTUs are grouped based on their phylum (Firmicutes, Actinobacteria, Bacteroidetes, Proteobacteria, Fusobacteria, and TM7). Interestingly, even though Osborne et al. (2022) used a different Bayesian approach from our method, based on a latent Gaussian graphical model with separate variable selection priors for both covariate-dependent mean and covariateindependent precision, we found that a large number of edges in the overall graph selected by DGSS were also detected by their method (98 out of 271 edges). In particular, the common edges highlight the subnetwork formed by OTUs 1 - 26 within *Firmicutes*, which appears to be the area consistently detected as having covariate-dependent edges for various cytokines. Jointly, these findings may imply the existence of a subnetwork within the



Figure 2: Adjacency matrices corresponding to the precision coefficients  $B_k$  of the four covariates with the most covariate-dependent edges (first row) and of the four covariates with the least covariate-dependent edges (second row), labeled by each covariate's name, with the number of edges indicated in parentheses.

*Firmicutes* that is widely affected by cytokines. Additionally, the proposed method selects more inter-phylum edges compared to the model from Osborne et al. (2022), which capture correlations among different OTUs across phyla, suggesting more complex latent effects of microbiome during pregnancy.



Figure 3: Adjacency matrix of the overall graph selected by DGSS (a) and the common edges selected by both DGSS and Osborne et al. (2022) (b).

# 5 Concluding remarks

We have considered the framework of covariate-dependent Gaussian graphical modeling for learning heterogeneous graphs and proposed a dual group spike-and-slab prior that achieves simultaneous local sparsity and bi-directional group sparsity. The proposed prior accomplishes covariate-level selection, inferred by the local-level selection, on grouped precision coefficients sliced in one direction and the node-level selection on grouped coefficients sliced in another direction. Our approach has led to a parsimonious model for covariate-dependent precision matrices with improved interpretability. For posterior inference, we have designed a Gibbs sampler to automatically tune the hyper-parameters while incorporating their uncertainty, leads to interpretable and flexible selection results. Through simulation studies, we have demonstrated that the proposed model outperforms existing methods in its accuracy of graph recovery. We have applied our model to microbiome data to estimate the interaction between microbes in the vagina, as well as the interplay between vaginal cytokines and microbial abundances, providing insight into mechanisms of host-microbial interaction during pregnancy.

There are several interesting future directions to extend our model. First, the model can be expanded to incorporate a covariate-adjusted mean. A potential challenge here is the increased computational complexity due to a larger parameter space. Secondly, although our focus in on Gaussian graphical models, the structured sparsity we consider can be useful for other models with ultra-dimensional parameter spaces, such as arrays, that exhibit various grouping directions. Finally, approximation methods such as Variational Expectation Maximization may be worth investigating, as they improve the scalability of the method and allow for its application to larger datasets.

## Supplementary Material

The supplementary material includes detailed derivations of the Gibbs sampler in Section 2.3. R code and scripts to reproduce the results from the simulation study and the real data applications, with main functions coded in Rcpp, will be made available on Github upon acceptance of the paper.

## References

- Aitchison, J. (1982). The statistical analysis of compositional data. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 44(2):139–177.
- Bai, R., Moran, G. E., Antonelli, J. L., Chen, Y., and Boland, M. R. (2022). Spike-and-slab group lassos for grouped regression and sparse generalized additive models. *Journal of the American Statistical Association*, 117(537):184–197.
- Barbieri, M. M. and Berger, J. O. (2004). Optimal predictive model selection. Annals of Statistics, 32:870–897.
- Dempster, A. P. (1972). Covariance selection. *Biometrics*, 28(1):157–175.
- Fettweis, J. M., Serrano, M. G., Brooks, J. P., Edwards, D. J., Girerd, P. H., Parikh, H. I., Huang, B., Arodz, T. J., Edupuganti, L., Glascock, A. L., Xu, J., Jimenez, N. R., Vivadelli,

S. C., Fong, S. S., Sheth, N. U., Jean, S., Lee, V., Bokhari, Y., Lara, A. M., Mistry, S. D., Duckworth, R. A., Bradley, S. P., Kopardé, V. N., Orenda, X. V., Milton, S. H., Rozycki, S. K., Matveyev, A. V., Wright, M. L., Huzurbazar, S. V., Jackson, E. M., Smirnova, E., Korlach, J., Tsai, Y.-C., Dickinson, M. R., Brooks, J. L., Drake, J. I., Chaffin, D. O., Sexton, A. L., Gravett, M. G., Rubens, C. E., Wijesooriya, N. R., Hendricks-Muñoz, K. D., Jefferson, K. K., Strauss, J. F., and Buck, G. A. (2019). The vaginal microbiome and preterm birth. *Nature Medicine*, 25:1012–1021.

- Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, 8.
- Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., and Bonneau, R. (2015). Sparse and compositionally robust inference of microbial ecological networks. *PLoS Computational Biology*, 11.
- Lauritzen, S. L. (1996). Graphical Models. Oxford University Press.
- Leday, G. G. R., de Gunst, M. C. M., Kpogbezan, G. B., van der Vaart, A., van Wieringen, W. N., and van de Wiel, M. A. (2017). Gene network reconstruction using global-local shrinkage priors. *The Annals of Applied Statistics*, 11 1:41–68.
- Lin, H. and Peddada, S. D. (2020). Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ Biofilms and Microbiomes*, 6.
- Liu, H. and Wang, L. (2017). Tiger: A tuning-insensitive approach for optimally estimating Gaussian graphical models. *Electronic Journal of Statistics*, 11(1):241–294.
- Maathuis, M., Drton, M., Lauritzen, S., and Wainwright, M. (2018). Handbook of Graphical Models. CRC Press.
- Meinshausen, N. and Buhlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, 34:1436–1462.
- Newton, M. A., Noueiry, A., Sarkar, D., and Ahlquist, P. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics*, 5(2):155–176.

- Ni, Y., Stingo, F. C., and Baladandayuthapani, V. (2019). Bayesian graphical regression. Journal of the American Statistical Association, 114(525):184–197.
- Niu, Y., Ni, Y., Pati, D., and Mallick, B. K. (2023). Covariate-assisted bayesian graph learning for heterogeneous data. *Journal of the American Statistical Association*, pages 1–15.
- Osborne, N., Peterson, C. B., and Vannucci, M. (2022). Latent network estimation and variable selection for compositional data via variational EM. *Journal of Computational* and Graphical Statistics, 31(1):163–175.
- Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104:735–746.
- Savitsky, T. D., Vannucci, M., and Sha, N. (2011). Variable selection for nonparametric Gaussian process priors: Models and computational strategies. *Statistical Science*, 26 1:130–149.
- Serrano, M. G., Parikh, H. I., Brooks, J. P., Edwards, D. J., Arodz, T. J., Edupuganti, L., Huang, B., Girerd, P. H., Bokhari, Y., Bradley, S. P., Brooks, J. L., Dickinson, M. R., Drake, J. I., Duckworth, R. A., Fong, S. S., Glascock, A. L., Jean, S., Jimenez, N. R., Khoury, J., Kopardé, V. N., Lara, A. M., Lee, V., Matveyev, A. V., Milton, S. H., Mistry, S. D., Rozycki, S. K., Sheth, N. U., Smirnova, E., Vivadelli, S. C., Wijesooriya, N. R., Xu, J., Xu, P., Chaffin, D. O., Sexton, A. L., Gravett, M. G., Rubens, C. E., Hendricks-Muñoz, K. D., Jefferson, K. K., Strauss, J. F., Fettweis, J. M., and Buck, G. A. (2019). Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. Nature Medicine, 25:1001–1011.
- Stingo, F. C., Chen, Y. A., Tadesse, M. G., and Vannucci, M. (2011). Incorporating biological information into linear models: a Bayesian approach to the selection of pathways and genes. *The Annals of Applied Statistics*, 5 3:1978–2002.
- Tadesse, M. and Vannucci, M. (2021). Handbook of Bayesian Variable Selection. CRC PRESS.

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 58:267–288.
- Wang, Z., Baladandayuthapani, V., Kaseb, A. O., Amin, H. M., Hassan, M. M., Wang, W., and Morris, J. S. (2022). Bayesian edge regression in undirected graphical models to characterize interpatient heterogeneity in cancer. *Journal of the American Statistical Association*, 117:533–546.
- Wilms, I. and Bien, J. (2022). Tree-based node aggregation in sparse graphical models. Journal of Machine Learning Research, 23(243):1–36.
- Xu, X. and Ghosh, M. (2015). Bayesian variable selection and estimation for group lasso. Bayesian Analysis, 10:909–936.
- Zeng, Z., Li, M., and Vannucci, M. (2024). Bayesian image-on-scalar regression with a spatial global-local spike-and-slab prior. *Bayesian Analysis*, 19(1):235–260.
- Zhang, J. and Li, Y. (2023). High dimensional Gaussian graphical regression models with covariates. Journal of the American Statistical Association, 118(543):2088–2100.

## Supplementary materials

## S1. Markov Chain Monte Carlo Sampling (MCMC)

In this section, we provide the detailed derivations for the Gibbs sampler used in the main paper.

• Update the covariate-level selection parameters  $\{\tau_k^{ij}, \tilde{\tau}_k^{ij}, \gamma_k^{ij}, \pi_k\}$ Rewriting the likelihood of  $y_n^i$  from the covariate-level group perspective, we have that

the mean part is

$$\underbrace{\sum_{s \neq k} \sum_{l \neq i} \beta_s^{il} y_n^l x_n^s}_{\text{conditional on } s} + \sum_{j \neq i} \beta_k^{ij} y_n^j x_n^k}_{\text{conditional on } s}$$

denoted as  $c_n^{1,ijk}$ 

leading to the conditional distribution

$$(y_n^i - c_n^{1,ijk})| - \sim N\left(\sum_{j \neq i} \beta_k^{ij} y_n^j x_n^k, \sigma_i^2\right).$$

Similarly, with  $\beta_k^{ij} = \tau_k^{ij} b_k^{ij}$ , we have

$$(y_n^i - c_n^{1,ijk})| - \sim N \left( \underbrace{\sum_{\substack{j' \notin \{i,j\} \\ \text{ conditional on } j' \notin \{i,j\} \\ \text{ denoted as } c_n^{2,ijk}}}_{\beta_k^{ij'} y_n^{j'} x_n^k} + \tau_k^{ij} b_k^{ij} y_n^j x_n^k, \sigma_i^2 \right).$$

Denoting  $y_n^{ijk} = y_n^i - c_n^{1,ijk} - c_n^{2,ijk}$ , we have that the distribution of the latent coefficients conditional upon the indicators is

$$\begin{split} y_n^{ijk} |\tilde{\tau}_k^{ij}, \gamma_k^{ij} &= 1, - \sim N\left(\tilde{\tau}_k^{ij} b_k^{ij} y_n^j x_n^k, \sigma_i^2\right) \\ y_n^{ijk} |\gamma_k^{ij} &= 0, - \sim N\left(0, \sigma_i^2\right). \end{split}$$

Following Zeng et al. (2024), we integrate out the latent coefficients, obtaining

$$\begin{split} p\left(\gamma_{k}^{ij}=1 \mid -\right) \\ &= \frac{\int p\left(\gamma_{k}^{ij}=1, \tilde{\tau}_{k}^{ij} \mid -\right) d\tilde{\tau}_{k}^{ij}}{p\left(\gamma_{k}^{ij}=0, \tilde{\tau}_{k}^{ij}=0 \mid -\right) + \int p\left(\gamma_{k}^{ij}=1, \tilde{\tau}_{k}^{ij} \mid -\right) d\tilde{\tau}_{k}^{ij}} \\ &= \frac{1}{1+\theta_{k}^{ij}}, \end{split}$$

where the Bayes factor is

$$\begin{aligned} \theta_k^{ij} &= \frac{p\left(\gamma_k^{ij} = 0, \tilde{\tau}_k^{ij} = 0|-\right)}{\int p\left(\gamma_k^{ij} = 1, \tilde{\tau}_k^{ij}|-\right) d\tilde{\tau}_k^{ij}} \\ &= \frac{\frac{1}{p(y_{\cdot}^{ijk})} p\left(y_{\cdot}^{ijk}|\gamma_k^{ij} = 0, \tilde{\tau}_k^{ij} = 0\right) \times (1 - \pi_k)}{\frac{1}{p(y_{\cdot}^{ijk})} \int p\left(y_{\cdot}^{ijk}|\gamma_k^{ij} = 1, \tilde{\tau}_k^{ij}\right) p\left(\tilde{\tau}_k^{ij}\right) d\tilde{\tau}_k^{ij} \times \pi_k} \\ &= \frac{p\left(y_{\cdot}^{ijk}|\gamma_k^{ij} = 0, \tilde{\tau}_k^{ij} = 0\right) \times (1 - \pi_k)}{\int p\left(y_{\cdot}^{ijk}|\gamma_k^{ij} = 1, \tilde{\tau}_k^{ij}\right) p\left(\tilde{\tau}_k^{ij}\right) d\tilde{\tau}_k^{ij} \times \pi_k} \\ &= \left(\frac{\int p\left(y_{\cdot}^{ijk}|\gamma_k^{ij} = 1, \tilde{\tau}_k^{ij}\right) p\left(\tilde{\tau}_k^{ij}\right) d\tilde{\tau}_k^{ij}}{p\left(y_{\cdot}^{ijk}|\gamma_k^{ij} = 0, \tilde{\tau}_k^{ij} = 0\right)}\right)^{-1} \times \frac{1 - \pi_k}{\pi_k}. \end{aligned}$$

Next, we estimate  $\theta_k^{ij}$ . First we note that

$$p\left(y_{\cdot}^{ijk}|\gamma_{k}^{ij}=0,\tilde{\tau}_{k}^{ij}=0\right) = \left(2\pi\sigma_{i}^{2}\right)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2}\sum_{n=1}^{N}\left(y_{n}^{ijk}\right)^{2}/\sigma_{i}^{2}\right\}.$$

Therefore,

$$\begin{split} &\int p\left(y_{\cdot}^{ijk}|\gamma_{k}^{ij}=1,\tilde{\tau}_{k}^{ij}\right)p\left(\tilde{\tau}_{k}^{ij}\right)d\tilde{\tau}_{k}^{ij} \\ &= \int \left(2\pi\sigma_{i}^{2}\right)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2}\sum_{n=1}^{N}\left(y_{n}^{ijk}-y_{n}^{j}x_{n}^{k}b_{k}^{ij}\tilde{\tau}_{k}^{ij}\right)^{2}/\sigma_{i}^{2}\right\} \\ &\times 2\left(2\pi s_{k}^{2}\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\tilde{\tau}_{k}^{ij}\right)^{2}/s_{k}^{2}\right\}\mathbbm{1}\left(\tilde{\tau}_{k}^{ij}\geq0\right)d\tilde{\tau}_{k}^{ij} \\ &= 2\left(2\pi s_{k}^{2}\right)^{-\frac{1}{2}}\underbrace{\left(2\pi\sigma_{i}^{2}\right)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2}\sum_{n=1}^{N}\left(y_{n}^{jik}\right)^{2}/\sigma_{i}^{2}\right\}}_{=p\left(y_{\cdot}^{ijk}|\gamma_{k}^{ij}=0,\tilde{\tau}_{k}^{ij}=0\right)} \\ &\times \int \exp\left\{-\frac{1}{2}\left[\left(\sum_{n=1}^{N}\left(y_{n}^{j}x_{n}^{k}\right)^{2}\left(b_{k}^{ij}\right)^{2}/\sigma_{i}^{2}+1/s_{k}^{2}\right)\left(\tilde{\tau}_{k}^{ij}\right)^{2}-2\left(b_{k}^{ij}\sum_{n=1}^{N}y_{n}^{j}x_{n}^{k}y_{n}^{ijk}/\sigma_{i}^{2}\right)\tilde{\tau}_{k}^{ij}\right]\right\} \\ &\mathbbm{1}\left(\tilde{\tau}_{k}^{ij}\geq0\right)d\tilde{\tau}_{k}^{ij}. \end{split}$$

Letting  $\tilde{\nu}_{ijk}^2 = \left(\sum_{n=1}^N \left(y_n^j x_n^k\right)^2 \left(b_k^{ij}\right)^2 / \sigma_i^2 + 1/s_k^2\right)^{-1}$  and  $\tilde{m}_{ijk} = \tilde{\nu}_{ijk}^2 b_k^{ij} \sum_{n=1}^N y_n^j x_n^k y_n^{ijk} / \sigma_i^2$ , we obtain the ratio

$$\begin{split} &\frac{\int p\left(y_{\cdot}^{ijk}|\gamma_{k}^{ij}=1,\tilde{\tau}_{k}^{ij}\right)p\left(\tilde{\tau}_{k}^{ij}\right)d\tilde{\tau}_{k}^{ij}}{p\left(y_{\cdot}^{ijk}|\gamma_{k}^{ij}=0,\tilde{\tau}_{k}^{ij}=0\right)} \\ &=& 2\left(2\pi s_{k}^{2}\right)^{-\frac{1}{2}}\left(2\pi\tilde{\nu}_{ijk}^{2}\right)^{\frac{1}{2}}\exp\left(\frac{1}{2}\frac{\tilde{m}_{ijk}^{2}}{\tilde{\nu}_{ijk}^{2}}\right) \\ &\times\int \left(2\pi\tilde{\nu}_{ijk}^{2}\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\left(\tilde{\tau}_{k}^{ij}\right)^{2}-2\tilde{\tau}_{k}^{ij}\tilde{m}_{ijk}+\left(\tilde{m}_{ijk}\right)^{2}\right]/\tilde{\nu}_{ijk}^{2}\right\}\mathbb{1}\left(\tilde{\tau}_{k}^{ij}\geq0\right)d\tilde{\tau}_{k}^{ij} \\ &=& 2\left(s_{k}^{2}\right)^{-\frac{1}{2}}\times\left(\tilde{\nu}_{ijk}^{2}\right)^{\frac{1}{2}}\exp\left\{\frac{1}{2}\frac{\tilde{m}_{ijk}^{2}}{\tilde{\nu}_{ijk}^{2}}\right\}\times\Phi\left(\frac{\tilde{m}_{ijk}}{\tilde{\nu}_{ijk}}\right), \end{split}$$

where the last line follows from the fact that the integral to be evaluated is associated with the truncated normal kernel,  $N^+(\tilde{m}_{ijk}, \tilde{\nu}_{ijk}^2)$ , which leads to the result  $\Phi(\tilde{m}_{ijk}/\tilde{\nu}_{ijk})$ . Substituting the ratio above yields

$$\theta_{k}^{ij} = \left(\frac{\int p\left(y_{\cdot}^{ijk} | \gamma_{k}^{ij} = 1, \tilde{\tau}_{k}^{ij}\right) p\left(\tilde{\tau}_{k}^{ij}\right) d\tilde{\tau}_{k}^{ij}}{p\left(y_{\cdot}^{ijk} | \gamma_{k}^{ij} = 0, \tilde{\tau}_{k}^{ij} = 0\right)}\right)^{-1} \times \frac{1 - \pi_{k}}{\pi_{k}}$$
$$= \frac{1 - \pi_{k}}{2\left(s_{k}^{2}\right)^{-\frac{1}{2}} \times \left(\tilde{\nu}_{ijk}^{2}\right)^{\frac{1}{2}} \exp\left\{\frac{1}{2}\frac{\tilde{m}_{ijk}^{2}}{\tilde{\nu}_{ijk}^{2}}\right\} \Phi\left(\frac{\tilde{m}_{ijk}}{\tilde{\nu}_{ijk}}\right) \times \pi_{k}}.$$

Hence, we sample each  $\gamma_k^{ij}$  as

$$\gamma_k^{ij}| - \sim \text{Bernoulli}\left(\frac{1}{1 + \theta_k^{ij}}\right)$$

Then, if  $\gamma_k^{ij} = 1$ , we update  $\tilde{\tau}_k^{ij} | - \sim N^+ (\tilde{m}_{ijk}, \tilde{\nu}_{ijk}^2)$ ; else, if  $\gamma_k^{ij} = 0$ , we set  $\tilde{\tau}_k^{ij} = 0$ . After updating all indicators for covariate  $x^k$ , we update

$$\pi_k | - \sim \text{Beta}\left(a_k + \sum_{1 < i \neq j \le p} \gamma_k^{ij}, b_k + p(p-1) - \sum_{1 < i \neq j \le p} \gamma_k^{ij}\right)$$

,

leading to  $\tau_k^{ij} = \tilde{\tau}_k^{ij} \delta_k = \tilde{\tau}_k^{ij} I(\pi_k \ge d)$ .

• Update the node-level selection parameters  $\{b^{ij}, \delta^{ij}, \pi^i\}$  together with  $\{\beta_k^{ij}\}$ Rewriting the likelihood of  $y_n^i$  from the node-level group perspective, we have the mean part

$$\sum_{j' \notin \{i,j\}} \sum_{k} \beta_k^{ij'} y_n^{j'} x_n^k + \sum_{k} \beta_k^{ij} y_n^j x_n^k.$$
With  $\beta_k^{ij} = \tau_k^{ij} b_k^{ij}$ , we denote  $z_n^{ij} = y_n^i - \sum_{j' \notin \{i,j\}} \sum_k \beta_k^{ij'} y_n^{j'} x_n^k$ , leading to  $z_n^{ij} |- \sim N\left(\sum_k b_k^{ij} \tau_k^{ij} y_n^j x_n^k, \sigma_i^2\right)$ 

and

$$\begin{aligned} z_n^{ij} | \delta^{ij} &= 1, - \sim N\left( \left( \boldsymbol{X}_n^{ij} \right)^T \boldsymbol{V}^{ij} \boldsymbol{b}^{ij}, \sigma_i^2 \right) \\ z_n^{ij} | \delta^{ij} &= 0, - \sim N\left( 0, \sigma_i^2 \right), \end{aligned}$$

where  $\boldsymbol{X}_{n}^{ij} = (y_{n}^{j}x_{n}^{1}, ..., y_{n}^{j}x_{n}^{q})^{T}$  and  $\boldsymbol{V}^{ij} = \text{diag}\left(\tau_{1}^{ij}, ..., \tau_{q}^{ij}\right)$ . In addition, we denote  $\boldsymbol{Z}^{ij} = \left(z_{1}^{ij}, ..., z_{n}^{ij}\right)^{T}$  and  $\boldsymbol{X}^{ij} = \left(\boldsymbol{X}_{1}^{ij}, ..., \boldsymbol{X}_{n}^{ij}\right)^{T}$ , leading to the vector form formulations

$$\boldsymbol{Z}^{ij}|\delta^{ij} = 1, - \sim MVN\left(\boldsymbol{X}^{ij}\boldsymbol{V}^{ij}\boldsymbol{b}^{ij}, \sigma_i^2\boldsymbol{I}_n\right)$$
$$\boldsymbol{Z}^{ij}|\delta^{ij} = 0, - \sim MVN\left(\boldsymbol{0}_n, \sigma_i^2\boldsymbol{I}_n\right).$$

Similarly, we integrate out  $\boldsymbol{b}^{ij}$ 

$$p\left(\delta^{ij} = 1 \mid -\right)$$

$$= \frac{\int p\left(\delta^{ij} = 1, \mathbf{b}^{ij} \mid -\right) d\mathbf{b}^{ij}}{p\left(\delta^{ij} = 0, \mathbf{b}^{ij} = \mathbf{0} \mid -\right) + \int p\left(\delta^{ij} = 1, \mathbf{b}^{ij} \mid -\right) d\mathbf{b}^{ij}}$$

$$= \frac{1}{1 + \theta^{ij}},$$

where the Bayes factor is

$$\begin{aligned} \theta^{ij} &= \frac{p\left(\delta^{ij} = 0, \mathbf{b}^{ij} = \mathbf{0}\right| -)}{\int p\left(\delta^{ij} = 1, \mathbf{b}^{ij}\right| -) d\mathbf{b}^{ij}} \\ &= \frac{\frac{1}{p(\mathbf{Z}^{ij})} p\left(\mathbf{Z}^{ij} | \delta^{ij} = 0, \mathbf{b}^{ij} = \mathbf{0}\right) \times (1 - \pi^{i})}{\frac{1}{p(\mathbf{Z}^{ij})} \int p\left(\mathbf{Z}^{ij} | \delta^{ij} = 1, \mathbf{b}^{ij}\right) p\left(\mathbf{b}^{ij}\right) d\mathbf{b}^{ij} \times \pi^{i}} \\ &= \frac{p\left(\mathbf{Z}^{ij} | \delta^{ij} = 0, \mathbf{b}^{ij} = \mathbf{0}\right) \times (1 - \pi^{i})}{\int p\left(\mathbf{Z}^{ij} | \delta^{ij} = 1, \mathbf{b}^{ij}\right) p\left(\mathbf{b}^{ij}\right) d\mathbf{b}^{ij} \times \pi^{i}} \\ &= \left(\frac{\int p\left(\mathbf{Z}^{ij} | \delta^{ij} = 1, \mathbf{b}^{ij}\right) p\left(\mathbf{b}^{ij}\right) d\mathbf{b}^{ij}}{p\left(\mathbf{Z}^{ij} | \delta^{ij} = 0, \mathbf{b}^{ij} = \mathbf{0}\right)}\right)^{-1} \times \frac{(1 - \pi^{i})}{\pi^{i}}. \end{aligned}$$

We next estimate  $\theta^{ij}$ . First, we note that

$$p\left(\boldsymbol{Z}^{ij}|\delta^{ij}=0,\boldsymbol{b}^{ij}=\boldsymbol{0}\right)=\left(2\pi\sigma^{2}\right)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2\sigma_{i}^{2}}\left(\boldsymbol{Z}^{ij}\right)^{T}\boldsymbol{Z}^{ij}\right\}.$$

Therefore,

$$\begin{split} &\int p\left(\boldsymbol{Z}^{ij}|\delta^{ij}=1,\boldsymbol{b}^{ij}\right)p\left(\boldsymbol{b}^{ij}\right)d\boldsymbol{b}^{ij} \\ &= \int \left(2\pi\sigma_{i}^{2}\right)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2\sigma_{i}^{2}}\left(\boldsymbol{Z}^{ij}-\boldsymbol{X}^{ij}\boldsymbol{V}^{ij}\boldsymbol{b}^{ij}\right)^{T}\left(\boldsymbol{Z}^{ij}-\boldsymbol{X}^{ij}\boldsymbol{V}^{ij}\boldsymbol{b}^{ij}\right)\right\} \\ &\times (2\pi)^{-\frac{q}{2}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{b}^{ij}\right)^{T}\boldsymbol{b}^{ij}\right\}d\boldsymbol{b}^{ij} \\ &= \underbrace{\left(2\pi\sigma_{i}^{2}\right)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2\sigma_{i}^{2}}\left(\boldsymbol{Z}^{ij}\right)^{T}\boldsymbol{Z}^{ij}\right\}}_{=p(\boldsymbol{Z}^{ij}|\delta^{ij}=0,\boldsymbol{b}^{ij}=0)} \left(2\pi)^{-\frac{q}{2}}\left(2\pi\right)^{\frac{q}{2}}\left|\tilde{\boldsymbol{\Sigma}}^{ij}\right|^{\frac{1}{2}}\exp\left\{\frac{1}{2}\left(\tilde{\boldsymbol{\mu}}^{ij}\right)^{T}\left(\tilde{\boldsymbol{\Sigma}}^{ij}\right)^{-1}\tilde{\boldsymbol{\mu}}^{ij}\right\} \\ &\times \int \left(2\pi\right)^{-\frac{q}{2}}\left|\tilde{\boldsymbol{\Sigma}}^{ij}\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{b}^{ij}\right)^{T}\left(\frac{1}{\sigma_{i}^{2}}\left(\boldsymbol{X}^{ij}\boldsymbol{V}^{ij}\right)^{T}\left(\boldsymbol{X}^{ij}\boldsymbol{V}^{ij}\right)+\boldsymbol{I}_{q}\right)\boldsymbol{b}^{ij} \\ &\quad -2\frac{1}{\sigma_{i}^{2}}\left(\boldsymbol{Z}^{ij}\right)^{T}\boldsymbol{X}^{ij}\boldsymbol{V}^{ij}\tilde{\boldsymbol{\Sigma}}^{ij}\left(\tilde{\boldsymbol{\Sigma}}^{ij}\right)^{-1}\boldsymbol{b}^{ij}+\left(\tilde{\boldsymbol{\mu}}^{ij}\right)^{T}\left(\tilde{\boldsymbol{\Sigma}}^{ij}\right)^{-1}\tilde{\boldsymbol{\mu}}^{ij}\right]\right\}. \\ \text{Letting } \tilde{\boldsymbol{\Sigma}}^{ij} = \left(\frac{1}{\sigma_{i}^{2}}\left(\boldsymbol{X}^{ij}\boldsymbol{V}^{ij}\right)^{T}\left(\boldsymbol{X}^{ij}\boldsymbol{V}^{ij}\right)+\boldsymbol{I}_{q}\right)^{-1} \text{ and } \tilde{\boldsymbol{\mu}}^{ij} = \left(\frac{1}{\sigma_{i}^{2}}\left(\boldsymbol{Z}^{ij}\right)^{T}\boldsymbol{X}^{ij}\boldsymbol{V}^{ij}\tilde{\boldsymbol{\Sigma}}^{ij}\right)^{T}, \end{split}$$

we obtain the ratio

$$\frac{\int p\left(\boldsymbol{Z}^{ij}|\delta^{ij}=1,\boldsymbol{b}^{ij}\right)p\left(\boldsymbol{b}^{ij}\right)d\boldsymbol{b}^{ij}}{p\left(\boldsymbol{Z}^{ij}|\delta^{ij}=0,\boldsymbol{b}^{ij}=\boldsymbol{0}\right)}$$
$$=\left|\tilde{\boldsymbol{\Sigma}}^{ij}\right|^{\frac{1}{2}}\exp\left\{\frac{1}{2}\left(\tilde{\boldsymbol{\mu}}^{ij}\right)^{T}\left(\tilde{\boldsymbol{\Sigma}}^{ij}\right)^{-1}\tilde{\boldsymbol{\mu}}^{ij}\right\}$$

,

where the last line follows from the fact that the integral to be evaluated is the probability of a  $MVN\left(\tilde{\mu}^{ij}, \tilde{\Sigma}^{ij}\right)$  random vector, i.e., 1. Substituting the ratio above yields

$$\theta^{ij} = \left(\frac{\int p\left(\mathbf{Z}^{ij}|\delta^{ij}=1, \mathbf{b}^{ij}\right) p\left(\mathbf{b}^{ij}\right) d\mathbf{b}^{ij}}{p\left(\mathbf{Z}^{ij}|\delta^{ij}=0, \mathbf{b}^{ij}=\mathbf{0}\right)}\right)^{-1} \times \frac{(1-\pi^{i})}{\pi^{i}}$$
$$= \frac{1-\pi^{i}}{\left|\tilde{\boldsymbol{\Sigma}}^{ij}\right|^{\frac{1}{2}} \exp\left\{\frac{1}{2}\left(\tilde{\boldsymbol{\mu}}^{ij}\right)^{T}\left(\tilde{\boldsymbol{\Sigma}}^{ij}\right)^{-1}\tilde{\boldsymbol{\mu}}^{ij}\right\} \times \pi^{i}}.$$

Hence, we first sample each  $\delta^{ij}$  by

$$\delta^{ij}| - \sim \text{Bernoulli}\left(\frac{1}{1+\theta^{ij}}\right)$$

Then, if  $\delta^{ij} = 1$ , we update  $\boldsymbol{b}^{ij} | - \sim N\left(\tilde{\boldsymbol{\mu}}^{ij}, \tilde{\boldsymbol{\Sigma}}^{ij}\right)$ ; else, if  $\delta^{ij} = 0$ , we set  $\boldsymbol{b}^{ij} = \mathbf{0}_q$ . After updating all indicators for node *i*, we update

$$\pi^{i}| - \sim \operatorname{Beta}\left(a^{i} + \sum_{j \neq i} \delta^{ij}, b^{i} + (p-1) - \sum_{j \neq i} \delta^{ij}\right),$$

and

$$(\beta_k^{ij})_{1\leq k\leq q}=\boldsymbol{B}^{ij}=\boldsymbol{V}^{ij}\boldsymbol{b}^{ij}.$$

• Update the variances  $\{\sigma_i^2\}$ 

The posterior of  $\sigma_i^2$  is:

$$p(\sigma_{i}^{2}|-) \propto (\sigma_{i}^{2})^{-\frac{N}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma_{i}^{2}}\sum_{n=1}^{N}\left(y_{n}^{i}-\sum_{j\neq i}\sum_{k=1}^{q}\beta_{k}^{ij}y_{n}^{j}x_{n}^{k}\right)^{2}\right\} \times (\sigma_{i}^{-2})^{a_{\sigma}+1} \exp\left\{-\sigma_{i}^{-2}b_{\sigma}\right\} \\ \propto (\sigma_{i}^{-2})^{\frac{N}{2}+a_{\sigma}+1} \exp\left\{-\sigma_{i}^{-2}\left[\frac{1}{2}\sum_{n=1}^{N}\left(y_{n}^{i}-\sum_{j\neq i}\sum_{k=1}^{q}\beta_{k}^{ij}y_{n}^{j}x_{n}^{k}\right)^{2}+b_{\sigma}\right]\right\},$$

which is an Inverse-Gamma distribution

$$\sigma_i^2 | - \sim \text{InvGamma}\left(\frac{N}{2} + a_\sigma, \frac{1}{2}\sum_{n=1}^N \left(y_n^i - \sum_{j \neq i}\sum_{k=1}^q \beta_k^{ij} y_n^j x_n^k\right)^2 + b_\sigma\right).$$

# • Update $\{s_k^2\}$ and t:

We have conjugate updates:

$$\begin{split} p\left(s_{k}^{2}|-\right) &= \prod_{\substack{1 \leq i \neq j \leq p \\ \gamma_{k}^{ij}=1}} 2\left(2\pi s_{k}^{2}\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{\left(\tilde{\tau}_{k}^{ij}\right)^{2}}{s_{k}^{2}}\right\} \mathbb{1}\left(\tilde{\tau}_{k}^{ij} \geq 0\right) \\ &\times \frac{t^{1}}{\Gamma(1)}\left(s_{k}^{2}\right)^{-2} \exp\left\{-\frac{t}{s_{k}^{2}}\right\} \\ &\propto \left(s_{k}^{2}\right)^{-\left(\frac{1}{2}\sum_{1 \leq i \neq j \leq p}\gamma_{k}^{ij}+1+1\right)} \exp\left\{-s_{k}^{-2}\left[\frac{1}{2}\sum_{1 \leq i \neq j \leq p}\left(\tilde{\tau}_{k}^{ij}\right)^{2}+t\right]\right\}, \end{split}$$

which is an Inverse-Gamma distribution

$$s_k^2 | - \sim \operatorname{InvGamma}\left(1 + \frac{1}{2} \sum_{1 \le i \ne j \le p} \gamma_k^{ij}, t + \frac{1}{2} \sum_{1 \le i \ne j \le p} \left(\tilde{\tau}_k^{ij}\right)^2\right).$$

The posterior of t is:

$$p(t|-) = \prod_{k=1}^{q} \frac{t^{1}}{\Gamma(1)} \left(s_{k}^{2}\right)^{-2} \exp\left\{-\frac{t}{s_{k}^{2}}\right\}$$
$$\propto t^{q} \exp\left(-t \sum_{k=1}^{q} \frac{1}{s_{k}^{2}}\right),$$

which is a Gamma distribution

$$t|- \sim \operatorname{Gamma}\left(q+1, \sum_{k=1}^{q} \frac{1}{s_k^2}\right).$$