Supplementary Document for "Bayesian Graphical Models under Positivity Constraints: A Scalable Generalized Likelihood Approach"

Supplementary Material for *Statistics and Computing*

Swarnali Raha[1], Partha Sarkar[2*], Sirani M. Perera[3], Kshitij Khare[1]

[1]Department of Statistics, University of Florida, Gainesville, Florida, USA.

[2]Department of Statistics, Florida State University, Tallahassee, Florida, USA.

[3]Department of Mathematics, Embry-Riddle Aeronautical University, Daytona Beach, Florida, USA.

*Corresponding author(s). E-mail(s): ps24v@fsu.edu;

Contributing authors: swarnali.raha@ufl.edu; Sirani.Perera@erau.edu; kdkhare@ufl.edu;

This supplementary document contains proofs of lemmas introduced in the original text, detailed description of methodologies and numerical experiments on the diagonally dominant setting and additional experimental details.

S.1 Proof of the Lemmas

S.1.1 Proof of Lemma 2

Let $\xi = (\omega_{12}, \omega_{13}, \ldots, \omega_{1p}, \omega_{23}, \ldots, \omega_{p-1,p})^T$. denote the vector of off-diagonal entries of Ω, and let $\delta = (\omega_{11}, \ldots, \omega_{pp})^T$ denote the vector of diagonal entries. Let $\mathcal{L} = \{0,1\}^{\binom{p}{2}}$ be the collection of all sparsity patterns for ξ. For $l \in \mathcal{L}$, let d_l= number of active off-diagonal entries under the pattern l, and let $\mathcal{M}_l$ denote the corresponding subspace of off-diagonal vectors satisfying that pattern.

Using the spike-and-slab representation, the normalizing constant associated with the posterior kernel in Lemma 1 can be written as

$$Z = \sum_{l \in \mathcal{L}} q^{d_l} (1-q)^{\binom{p}{2} - d_l} \left(\frac{2}{\tau\sqrt{2\pi}} \right)^{d_l} Z_l,$$

where

$$Z_l = \int_{\mathcal{S}_l^*(p)} \exp\left\{ n \operatorname{tr}(\Omega) - \frac{n}{2} \operatorname{tr}(\Omega^2 S) - \frac{1}{2\tau^2} \|\xi_l\|_2^2 \right\} d\delta \, d\xi_l.$$

Here ξ_l denotes the subvector of active off-diagonal entries under the pattern l, and $\mathcal{S}_l^*(p)$ denotes the corresponding section of $\mathcal{S}^*(p)$ for which $\xi \in \mathcal{M}_l$. It is enough to show that $Z_l < \infty$ for every $l \in \mathcal{L}$, since $\mathcal{L}$ is finite. Fix $l \in \mathcal{L}$, and write $\theta_l = (\delta^T, \xi_l^T)^T \in A_l^*(p)$, where $A_l^*(p)$ is the corresponding vector parameter space induced by $\mathcal{S}_l^*(p)$. Since $\Omega \mapsto \operatorname{tr}(\Omega^2 S)$ is a quadratic form in the free parameters, there exists a nonnegative definite matrix $Q_l(S)$ such that $\operatorname{tr}(\Omega^2 S) = \theta_l^T Q_l(S) \theta_l$. Define

$$H_l = n Q_l(S) + \begin{pmatrix} 0_{p \times p} & 0 \\ 0 & \tau^{-2} I_{d_l} \end{pmatrix}.$$

Then

$$n \operatorname{tr}(\Omega) - \frac{n}{2} \operatorname{tr}(\Omega^2 S) - \frac{1}{2\tau^2} \|\xi_l\|_2^2 = n \mathbf{1}_p^T \delta - \frac{1}{2} \theta_l^T H_l \theta_l.$$

We next show that H_l is such that on the joint parameter vector $\theta_l = (\delta^T, \xi_l^T)^T \in A_l^*(p)$, we have $\theta_l^T H_l \theta_l > 0$ whenever $\theta_l \neq 0$.

For any θ_l, let $\Omega = \Omega(\theta_l)$ denote the corresponding symmetric matrix under the sparsity pattern l. Then

$$\theta_l^T H_l \theta_l = n \operatorname{tr}(\Omega^2 S) + \tau^{-2} \|\xi_l\|_2^2 = n \operatorname{tr}(\Omega S \Omega) + \tau^{-2} \|\xi_l\|_2^2.$$

Since S is nonnegative definite, $\operatorname{tr}(\Omega S \Omega) = \|S^{1/2} \Omega\|_F^2 \geq 0$, and hence $\theta_l^T H_l \theta_l \geq 0$ for every $\theta_l \in A_l^*(p)$. Suppose $\theta_l^T H_l \theta_l = 0$, which implies both nonnegative terms above must be zero. In particular, $\|\xi_l\|_2^2 = 0$, and hence $\xi_l = 0$. Since all inactive off-diagonal entries are already fixed at zero by the sparsity pattern l, it follows that all off-diagonal

entries of Ω are zero. Thus $\Omega = \text{diag}(\delta)$, and consequently,

$$0 = \text{tr}(\Omega S \Omega) = \sum_{j=1}^{p} s_{jj}\omega_{jj}^2.$$

Under the assumption $s_{jj} > 0$, for every $j = 1, \ldots, p$, we obtain $\omega_{11} = \cdots = \omega_{pp} = 0$. Hence $\delta = 0$, and therefore $\theta_l = 0$. This proves that

$$\theta_l^T H_l \theta_l > 0 \qquad \text{for every } \theta_l \neq 0.$$

Consequently,

$$Z_l = \int_{A_l^*(p)} \exp\left\{ n\mathbf{1}_p^T \delta - \frac{1}{2}\theta_l^T H_l \theta_l \right\} d\theta_l < \infty.$$

Therefore $Z_l < \infty$ for every $l \in \mathcal{L}$. Since the number of sparsity patterns can not exceed $2^{p(p-1)/2}$,

$$Z = \sum_{l \in \mathcal{L}} q^{d_l}(1-q)^{\binom{p}{2}-d_l} \left(\frac{2}{\tau\sqrt{2\pi}}\right)^{d_l} Z_l < \infty.$$

Hence the joint generalized posterior distribution of Ω, given $\mathbf{Y}$, is proper.

S.1.2 Proof of Lemma 3

We write

$$\text{tr}(\Omega^2 S) = \sum_{i=1}^{p} \left[s_{ii}\omega_{ii}^2 + 2\omega_{ii} \sum_{j \neq i} \omega_{ij} s_{ij} \right] + \sum_{1 \leq i < j \leq p} (s_{ii} + s_{jj})\omega_{ij}^2 + 2 \sum_{i=1}^{p} \sum_{\substack{1 \leq j < k \leq p \\ j \neq i,\ k \neq i}} \omega_{ij}\omega_{ik} s_{jk}.$$

Using this and the expression of $\pi_D(\Omega|\mathbf{Y})$ in equation (4), we obtain the following conditional posterior distributions.

$$\begin{aligned} & \pi(\omega_{ij} \mid \Omega_{-(ij)}, \mathbf{Y}) \\ & \propto \exp\left\{ -\frac{n}{2}(a_{ij}\omega_{ij}^2 + 2b_{ij}\omega_{ij}) \right\} \left(I_{\{0\}}(\omega_{ij}) + \frac{2q}{(1-q)\sqrt{2\pi\tau^2}} I_{(-\infty,0)}(\omega_{ij}) \right) \\ & = \quad I_{\{0\}}(\omega_{ij}) + c_{ij} \frac{\sqrt{na_{ij}}}{\sqrt{2\pi}} \exp\left\{ -\frac{na_{ij}}{2} \left(\omega_{ij} + \frac{b_{ij}}{a_{ij}} \right)^2 \right\} I_{(-\infty,0)}(\omega_{ij}) \end{aligned}$$

with a_{ij}, b_{ij} and c_{ij} as defined earlier in equation (5). Writing $q_{ij} = \frac{c_{ij}}{1+c_{ij}}$, we obtain the conditional posterior distribution in equation (5) for $1 \leq i < j \leq p$.

Similarly, the conditional posterior distributions for the diagonal entries of Ω are computed as follows. For $i = 1, 2, \ldots, p$,

$$\begin{aligned}\pi(\omega_{ii} \mid \Omega_{-(ii)}, \mathbf{Y}) &\propto \exp\left(n\omega_{ii} - \frac{n}{2}s_{ii}\omega_{ii}^2 - nd_i\omega_{ii}\right) I_{(0,\infty)}(\omega_{ii}) \\ &\propto \exp\left[-\frac{ns_{ii}}{2}\left(\omega_{ii} - \frac{1-d_i}{s_{ii}}\right)^2\right] I_{(0,\infty)}(\omega_{ii})\end{aligned}$$

with $d_i = \sum_{i' \neq i} \omega_{ii'} s_{ii'}$. Therefore, the posteriors of the diagonal entries of Ω are as in equation (6).

S.1.3 Proof of Lemma 4

First, note that step 3 of algorithm 1 has computational complexity $\mathcal{O}(np^2)$. However, this step is performed only once in the beginning of the algorithm. Steps 5-22 are repeated in each of the T iterations of the Gibbs sampler.

Among these, steps 8-16 sample the off-diagonal entries of Ω and require $\mathcal{O}(p)$ computations for the vector multiplication in computing b in step 8, and $\mathcal{O}(1)$ computation for all the other steps. These steps with complexity $\mathcal{O}(p)$ are performed for each of the $\binom{p}{2}$ distinct off-diagonal entries.

Finally, steps 17-18 also require $\mathcal{O}(p)$ computations, and are repeated $\mathcal{O}(p)$ times, constituting $\mathcal{O}(p^2)$ complexity. Therefore, in total, the D-Gibbs algorithm (algorithm 1) has a per-iteration computational complexity $\mathcal{O}(p^3)$. □

S.1.4 Proof of Lemma 5

The augmented posterior distribution of Ω, conditioned on the augmentation variable R and the observation matrix $\mathbf{Y}$ can be written as

$$\pi(\Omega \mid R, \mathbf{Y}) \propto \pi(R \mid \Omega, \mathbf{Y}) L_D(\Omega \mid \mathbf{Y}) \pi(\Omega) \tag{S1}$$

where,

$$\begin{aligned}\pi(R \mid \Omega, \mathbf{Y}) &\propto \exp\left(-\frac{1}{2}\operatorname{tr}\left[(kI - nS)^{-1}(R - \Omega(kI - nS))^T(R - \Omega(kI - nS))\right]\right) \\ &\propto \exp\left(-\frac{k}{2}\operatorname{tr}(\Omega^2) + \frac{n}{2}\operatorname{tr}(\Omega^2 S) + \operatorname{tr}(\Omega R)\right)\end{aligned} \tag{S2}$$

Plugging $\pi(R \mid \Omega, \mathbf{Y})$ in (S1) in the expression of $\pi(\Omega \mid R, \mathbf{Y})$ in (S2), we obtain

$$\begin{aligned}&\pi(\Omega \mid R, \mathbf{Y}) \\ &\propto \exp\left(-\frac{k}{2}\operatorname{tr}(\Omega^2) + \operatorname{tr}(\Omega R) + n\operatorname{tr}(\Omega)\right)\end{aligned}$$

$$\times \prod_{i<j} \left((1-q) I_{\{0\}}(\omega_{ij}) + \frac{2q}{\tau\sqrt{2\pi}} e^{-\frac{\omega_{ij}^2}{2\tau^2}} I_{(-\infty,0)}(\omega_{ij}) \right)$$

$$= \exp\left[-\frac{1}{2} \left(\sum_{i<j} 2k\omega_{ij}^2 - 2\sum_{i<j} \omega_{ij}(R_{ij}+R_{ji}) \right) + n\sum_{i=1}^{p} \omega_{ii} - \frac{k}{2}\sum_{i=1}^{p} \omega_{ii}^2 + \sum_{i=1}^{p} \omega_{ii} R_{ii} \right]$$
$$\times \prod_{i<j} \left((1-q) I_{\{0\}}(\omega_{ij}) + \frac{2q}{\tau\sqrt{2\pi}} e^{-\frac{\omega_{ij}^2}{2\tau^2}} I_{(-\infty,0)}(\omega_{ij}) \right)$$

For $1 \le i < j \le p$, the posterior distribution of ω_{ij}, conditioned on R and $\mathbf{Y}$ can be computed as follows. Defining $k^* = 2k + \frac{1}{\tau^2}$, we have

$$\begin{aligned}
&\pi(\omega_{ij} \mid R, \mathbf{Y}, q, \tau^2) \\
&\propto e^{-\frac{1}{2}\left[2k\omega_{ij}^2 - 2(R_{ij}+R_{ji})\omega_{ij}\right]} \left[(1-q) I_{\{0\}}(\omega_{ij}) + q\frac{2}{\tau\sqrt{2\pi}} e^{-\frac{\omega_{ij}^2}{2\tau^2}} I_{(-\infty,0)}(\omega_{ij}) \right] \\
&\propto \exp\left[-\frac{1}{2} \left\{ \left(2k + \frac{1}{\tau^2} \right) \omega_{ij}^2 - 2(R_{ij}+R_{ji})\omega_{ij} \right\} \right] \\
&\qquad\qquad \times \left[I_{\{0\}}(\omega_{ij}) + \frac{2q}{(1-q)} \frac{1}{\tau\sqrt{2\pi}} I_{(-\infty,0)}(\omega_{ij}) \right] \\
&= \exp\left[-\frac{k^*}{2} \left(\omega_{ij}^2 - 2\frac{R_{ij}+R_{ji}}{k^*}\omega_{ij} \right) \right] \left[I_{\{0\}}(\omega_{ij}) + \frac{2q}{(1-q)} \frac{1}{\tau\sqrt{2\pi}} I_{(-\infty,0)}(\omega_{ij}) \right] \\
&= I_{\{0\}}(\omega_{ij}) + c_{ij} \cdot \frac{\sqrt{k^*}}{\sqrt{2\pi}} \exp\left[-\frac{k^*}{2} \left(\omega_{ij} - \frac{R_{ij}+R_{ji}}{k^*} \right)^2 \right] I_{(-\infty,0)}(\omega_{ij})
\end{aligned}$$

Writing $c_{ij} = \frac{2q}{(1-q)} \cdot \frac{1}{\tau\sqrt{k^*}} \cdot \exp\left[\frac{(R_{ij}+R_{ji})^2}{2k^*} \right] \Phi\left(-\frac{R_{ij}+R_{ji}}{\sqrt{k^*}} \right)$, and $q_{ij}^* = \frac{c_{ij}}{1+c_{ij}}$, we have

$$\pi(\omega_{ij} \mid R, \mathbf{Y}, \tau^2, q) \propto (1-q_{ij}^*)\delta_0 + q_{ij}^* \mathcal{N}\left(\frac{R_{ij}+R_{ji}}{k^*}, \frac{1}{k^*}; -\infty, 0 \right).$$

Also, for $i = 1, 2, \ldots, p$,

$$\begin{aligned}
\pi(\omega_{ii} \mid R_{ii}, \mathbf{Y}) &\propto \exp\left[(R_{ii}+n)\omega_{ii} - \frac{k}{2}\omega_{ii}^2 \right] I_{(0,\infty)}(\omega_{ii}) \\
&\propto \exp\left[-\frac{k}{2} \left(\omega_{ii} - \frac{R_{ii}+n}{k} \right)^2 \right] I_{(0,\infty)}(\omega_{ii})
\end{aligned}$$

Therefore, the conditional posterior distribution of the diagonal entries of Ω, for $i = 1, 2, \ldots, p$ are given by

$$\omega_{ii} \mid R_{ii}, \mathbf{Y} \overset{\text{ind}}{\sim} \mathcal{N}\left(\frac{R_{ii}+n}{k}, \frac{1}{k}; 0, \infty\right)$$

Hence the proof. □

S.1.5 Proof of Lemma 6

The moment generating function (mgf) of X is given by —

$$\begin{aligned}
M_X(P) &= E\left[\exp\left\{\text{tr}(P^T X)\right\}\right] \\
&= E\left[\exp\left\{\sqrt{k}\,\text{tr}\left(P^T\{V_1(I_p - UU^T) + V_2 A^T U^T\}\right)\right\}\right] \\
&= E\left[\exp\left\{\sqrt{k}\,\text{tr}\left((I_p - UU^T)P^T V_1\right)\right\}\exp\left\{\sqrt{k}\,\text{tr}\left(A^T U^T P^T V_2\right)\right\}\right] \\
&= M_{V_1}(\sqrt{k}P(I_p - UU^T))M_{V_2}(\sqrt{k}PUA), \text{ since } V_1, V_2 \text{ are independent} \\
&= \exp\left[\frac{k}{2}\,\text{tr}\left((I_p - UU^T)P^T P(I_p - UU^T)\right)\right]\exp\left[\frac{k}{2}\,\text{tr}\left(A^T U^T P^T PUA\right)\right] \\
&= \exp\left[\frac{k}{2}\,\text{tr}\left((I_p - UU^T)P^T P\right)\right], \text{ using } AA^T = I_n - U^T U, \\
&= \exp\left[\frac{1}{2}\,\text{tr}\left((kI_p - nS)P^T P\right)\right]
\end{aligned}$$

Using the uniqueness of the MGF, $X \sim \mathcal{MN}_{p\times p}(0, I_p, kI_p - nS)$.

After precomputing A, each draw of X requires $\mathcal{O}(p^2)$ and $\mathcal{O}(np)$ flops to compute samples V_1 and V_2 respectively, and computing $V_1(I_p - UU^T) + V_2 A^T U^T$ requires an additional $\mathcal{O}(np^2)$ flops Therefore, one round of sampling X has a computational complexity of $\mathcal{O}(np^2)$, when $p > n$. □

S.2 Additional Details on Graphical Models Under Diagonal Dominance Constraint

In this section, we extend the generalized Bayesian framework developed under the positivity constraint $\mathcal{S}^*(p)$ to the diagonally dominant class $\mathcal{S}^{**}(p)$. As discussed in Section 3, $\mathcal{S}^{**}(p)$ consists of symmetric matrices with positive diagonal entries, non-positive off-diagonal entries, and strict diagonal dominance, which guarantees positive definiteness. Diagonally dominant precision matrices arise naturally in several applications, including conditional autoregressive (CAR) models, multiresolution spatial graphical regression (Chen et al. 2026), and genetic regulatory networks (Zavlanos et al. 2011). These considerations motivate the development of scalable Bayesian procedures under the joint positivity and diagonal-dominance constraints. Building on the framework developed in the preceding sections, we develop a generalized Bayesian

approach for this restricted class using the D-trace generalized likelihood introduced in Section 2.2.

S.2.1 Model Description

In this section, we specify the prior distribution for Ω under the constraints defining $\mathcal{S}^{**}(p)$ and derive the corresponding joint generalized posterior distribution. As in Section 3, sparsity is induced through spike-and-slab prior components for the off-diagonal entries of Ω, while flat prior components are used for the diagonal entries. Unlike the setting considered in Section 3, the diagonal-dominance constraints couple the entries of Ω, and therefore the resulting prior is no longer componentwise independent. Specifically, we define

$$f(\Omega) \propto \left\{ \prod_{i=1}^{p} f_1(\omega_{ii}) \prod_{1 \le i < j \le p} f_2(\omega_{ij}) \right\} I_{\mathcal{S}^{**}(p)}(\Omega),$$

where f_1 denotes the flat prior component for the diagonal entries and f_2 denotes the spike-and-slab prior component for the off-diagonal entries. Specifically, f_2 consists of a point mass at zero and a truncated normal slab $\mathcal{N}(0, \tau^2; -\infty, 0)$ with slab inclusion probability $q \in (0, 1)$. The indicator $I_{\mathcal{S}^{**}(p)}(\Omega)$ further restricts the support of the joint prior to matrices satisfying the positivity and diagonal dominance constraints. Under the generalized likelihood based on the D-trace loss, the priors described above, and the diagonal dominance and positivity constraints, the joint generalized posterior density of Ω is calculated as given below.

$$\begin{aligned} \pi_D^{DD}(\Omega \mid \mathbf{Y}) \propto \exp &\left\{ n \sum_{i=1}^{p} \omega_{ii} - \frac{n}{2} \operatorname{tr}(\Omega^2 S) \right\} \\ &\times \prod_{\substack{i,j=1 \\ i<j}}^{p} \left[(1-q) I_{\{0\}}(\omega_{ij}) + q \frac{2}{\tau\sqrt{2\pi}} \exp\left\{ -\frac{\omega_{ij}^2}{2\tau^2} \right\} \right] I_{\mathcal{S}^{**}(p)}(\Omega). \end{aligned} \tag{S3}$$

Lemma S1 (Posterior propriety) Let $S = n^{-1}\mathbf{Y}^T\mathbf{Y}$, and suppose that $s_{ii} > 0$ for all $i = 1, \ldots, p$. Then the joint generalized posterior distribution of Ω conditional on $\mathbf{Y}$, given in (S3), is proper; that is, its normalizing constant is finite.

The proof follows immediately from the proof of Lemma 2 after replacing $\mathcal{S}^{*}(p)$ by its subset $\mathcal{S}^{**}(p)$, and is therefore omitted. Consequently, the generalized posterior distribution of Ω under the joint positivity and diagonal dominance constraints is proper and can be used for posterior sampling.

S.2.2 Sampling from the Posterior

Although direct sampling from the joint posterior distribution in (S3) is intractable, the full conditional distributions admit closed-form truncated-normal representations. Consequently, following the strategy developed in Section 4.1, we construct a componentwise direct Gibbs sampler (D-Gibbs) for posterior computation.

Lemma S2 (Conditional Posterior Distribution) The conditional posterior distributions of the off-diagonal entries of Ω, under the M-matrix and diagonal dominance constraints, given the remaining entries, are characterized by the following posterior densities. For $1 \le i < j \le p$,

$$\omega_{ij} \mid \mathbf{Y}, \Omega_{-(ij)} \sim (1 - q_{ij}^D)\delta_0 + q_{ij}^D \, \mathcal{N}\Big(-\frac{b_{ij}}{a_{ij}}, \, \frac{1}{n a_{ij}}; \, -\alpha_{ij}^D, 0\Big), \tag{S4}$$

where

$$a_{ij} = s_{ii} + s_{jj} + \frac{1}{n\tau^2}, \, b_{ij} = \sum_{j' \neq j} \omega_{ij'} s_{jj'} + \sum_{i' \neq i} \omega_{i'j} s_{ii'}, \, \alpha_{ij}^D = \min\Big(\sum_{j' \neq j} \omega_{ij'}, \sum_{i' \neq i} \omega_{i'j}\Big),$$

and $q_{ij}^D = \frac{c_{ij}^D}{1 + c_{ij}^D}$ with

$$c_{ij}^D = \frac{2q}{(1-q)\sqrt{n a_{ij} \tau^2}} \exp\left(\frac{n b_{ij}^2}{2 a_{ij}}\right) \left[\Phi\left(b_{ij}\sqrt{\frac{n}{a_{ij}}}\right) - \Phi\left(b_{ij}\sqrt{\frac{n}{a_{ij}}} - \alpha_{ij}^D \sqrt{n a_{ij}}\right)\right],$$

and s_{ij} denotes the (i,j)-th entry of $S = \frac{1}{n}\mathbf{Y}^\top \mathbf{Y}$. For any matrix A, $A_{-(ij)}$ denotes the vector of the entries in the upper triangle except the (i,j)-th entry.

The conditional posterior distributions of the diagonal entries of Ω, given the remaining entries, under the specified constraints, are given by

$$\omega_{ii} \mid \mathbf{Y}, \Omega_{-(ii)} \sim \mathcal{N}\Big(\frac{1 - d_i}{s_{ii}}, \, \frac{1}{n s_{ii}}; \, \beta_i^D, \infty\Big), \tag{S5}$$

where $d_i = \sum_{i' \neq i} \omega_{ii'} s_{ii'}, \; \beta_i^D = \sum_{j \neq i} |\omega_{ij}| = -\sum_{j \neq i} \omega_{ij}, \quad \text{for } i = 1, 2, \ldots, p.$

A schematic of the componentwise direct Gibbs sampler has been provided in Algorithm S1.

The updates required to maintain the diagonal-dominance constraints can be implemented incrementally after each entry of Ω is sampled. In particular, the updates of α and β involve only scalar corrections to previously computed quantities and therefore do not alter the leading-order computational cost. Consequently, a complete Gibbs sweep still requires $O(p^3)$ operations. See Steps 20 and 24 of Algorithm S1 for implementation details.

We next develop an augmented Gibbs sampler (A-Gibbs) for the present setting by introducing the auxiliary matrix R as in Section 4.2. Recall that, conditional on $\mathbf{Y}$ and Ω,

$$R \mid \mathbf{Y}, \Omega \sim \mathcal{MN}_{p \times p}\big(\Omega(kI_p - nS), \, I_p, \, kI_p - nS\big).$$

Algorithm S1 Direct Gibbs Sampler

procedure D-GIBBS($\Omega^{init}, \mathbf{Y}, T, q, \tau^2$)

$\Omega^{(0)} = \Omega^{init}$,

$S = n^{-1}\mathbf{Y}^T\mathbf{Y}$

$\alpha = ((\alpha_{ij}))_{i,j=1}^{p}$ with $\alpha_{ij} = \sum_{j' \neq j} \omega_{ij}$, $\beta = (\beta_i)_{i=1}^{p}$ with $\beta_i = \sum_{i' \neq i} |\omega_{ij}|$

for $iter = 1, 2, \ldots, T$ **do**

$\Omega^{new} \leftarrow \Omega^{(iter-1)}$

for $i = 1, 2, \ldots, p-1$ **do** ▷ **Updating off-diagonals of** Ω

for $j = i+1, \ldots, p$ **do**

$a \leftarrow s_{ii} + s_{jj} + 1/(n\tau^2), \quad b \leftarrow (\Omega_{.i}^{new})^T S_{.j} + (\Omega_{.j}^{new})^T S_{.i} - a\omega_{ij}^{new}$

$l \leftarrow -\min(\alpha_{ij}, \alpha_{ji})$, $H_\Phi \leftarrow \Phi\big(b\sqrt{\frac{n}{a}}\big) - \Phi\big(b\sqrt{\frac{n}{a}} + \alpha_{ij}\sqrt{na}\big)$

$P(0) \leftarrow 1, \quad P(1) \leftarrow \frac{2q}{(1-q)\sqrt{na\tau^2}} \exp\Big(\frac{nb^2}{2a}\Big) H_\Phi$

if $P(1) = \infty$ **then**

$\omega_{ij}^{new} \leftarrow \mathcal{N}\left(-\frac{b}{a}, \frac{1}{na}; l, 0\right)$, $\omega_{ji}^{new} \leftarrow \omega_{ij}^{new}$

else

$P \leftarrow P/\mathrm{sum}(P)$

$\omega_{ij}^{new} \sim P(0)\delta_0 + P(1)\mathcal{N}\left(-\frac{b}{a}, \frac{1}{na}; l, 0\right)$, $\omega_{ji}^{new} \leftarrow \omega_{ij}^{new}$

end if

$\Delta_1 = \omega_{ij}^{new} - \Omega_{ij}^{(iter-1)}$

$\beta_i \leftarrow \beta_i - \Delta_1$, $\beta_j \leftarrow \beta_j - \Delta_1$ ▷ **Updating** β

$(\alpha_{i.})_{-j} \leftarrow (\alpha_{i.})_{-j} + \Delta_1$, $(\alpha_{j.})_{-i} \leftarrow (\alpha_{j.})_{-i} + \Delta_1$ ▷ **Updating** α

end for

$d \leftarrow 1 - (\Omega_{.i}^{new})^T S_{.i} + \omega_{ii}^{new} s_{ii}$ ▷ **Updating diagonals of** Ω

$\omega_{ii}^{new} \leftarrow \mathcal{N}(\frac{d}{s_{ii}}, \frac{1}{ns_{ii}}; \beta_i, \infty)$

$\Delta_2 \leftarrow \omega_{ii}^{new} - \Omega_{ii}^{(iter-1)}, \alpha_{i.} \leftarrow \alpha_{i.} + \Delta_2$ ▷ **Updating** α

end for

Repeat Steps 22 - 24 for $i = p$

$\Omega^{(iter)} \leftarrow \Omega^{new}$

end for

return $\Omega^{(1)}, \Omega^{(2)}, \ldots, \Omega^{(T)}$

end procedure

When $p > n$, samples from this matrix-normal distribution can be generated efficiently by exploiting its Gram structure through the fast sampling scheme of Lemma 6. Under this augmentation, the conditional distributions retain tractable forms analogous to those derived in Section 4.2, with the diagonal-dominance constraints appearing through the truncation limits.

Lemma S3 (Conditional posterior distribution Under Augmentation) Under the augmented formulation in (7), the conditional posterior densities of the entries of Ω, conditioned on R, $\mathbf{Y}$, and the remaining entries of Ω are as follows. For $1 \leq i < j \leq p$,

$$\omega_{ij} \mid R, \mathbf{Y}, \Omega_{-(ij)} \sim (1 - q_{ij}^{*D})\delta_0 + q_{ij}^{*D}\mathcal{N}\left(\frac{R_{ij}+R_{ji}}{k^*}, \frac{1}{k^*}; -\alpha_{ij}^D, 0\right),$$
$$\omega_{ii} \mid R, \mathbf{Y}, \Omega_{-(ii)} \sim \mathcal{N}\left(\frac{R_{ii}+n}{k}, \frac{1}{k}; \beta_i^D, \infty\right), \tag{S6}$$

where

$$\alpha_{ij}^D = \min\left(\sum_{j' \neq j} \omega_{ij'}, \sum_{i' \neq i} \omega_{i'j}\right),\ \beta_i^D = -\sum_{j \neq i} \omega_{ij},\ k^* = 2k + \frac{1}{\tau^2},\ q_{ij}^{*D} = \frac{c_{ij}^D}{1 + c_{ij}^D},$$

and

$$c_{ij}^D = \frac{2q}{1-q}\frac{1}{\tau\sqrt{k^*}}\exp\left\{\frac{(R_{ij}+R_{ji})^2}{2k^*}\right\}\left[\Phi\left(-\frac{R_{ij}+R_{ji}}{\sqrt{k^*}}\right) - \Phi\left(-\frac{R_{ij}+R_{ji}}{\sqrt{k^*}} - \alpha_{ij}^D\sqrt{k^*}\right)\right].$$

Unlike the positivity-constrained formulation considered in Section 4.2, conditioning on R does not render the entries of Ω mutually independent, since the diagonal-dominance constraints induce dependence through the truncation bounds.

Lemma S2 and Lemma S3 can be established along similar lines as the proofs of Lemma 3 and Lemma 5, respectively; see Sections S.1.2 and S.1.5 in the online resource document. Their proofs are therefore omitted. A schematic of the resulting A-Gibbs sampler is provided in Algorithm S2 below.

The conditional-independence structure exploited in Section 4.2 is lost under diagonal dominance. As a result, the computational complexity of A-Gibbs increases from $O(np^2)$ to $O(p^3)$ when $n \ll p$. This additional cost is the price of enforcing diagonal dominance and thereby guaranteeing positive-definite precision matrix draws. Nevertheless, in our numerical experiments (Section S.3.3), the D-Gibbs sampler exhibits occasional instability over a small number of iterations, particularly for larger n, whereas A-Gibbs remains comparatively stable across the (n, p) configurations considered. Moreover, as discussed in detail in Section 4.3, A-Gibbs facilitates broader exploration through the augmented parameter space, while D-Gibbs performs more localized updates in the original parameter space. Motivated by these complementary behaviors, similar to Section 4.3, we propose an Interweaving Gibbs sampler that combines both mechanisms by inserting one D-Gibbs update within a sequence of A-Gibbs updates. A schematic of the resulting interweaving Gibbs sampler is provided in Algorithm S3.

Once the post-burn-in posterior samples of Ω are obtained from any of the three proposed algorithms, the graph structure can be estimated using the majority-voting approach and final precision matrix estimate is then obtained by averaging over the post burn-in samples as described in Section 4.5.

Algorithm S2 Augmented Gibbs Sampler

procedure A-GIBBS($\Omega^{init}, \mathbf{Y}, T, \varepsilon, q, \tau^2$)

$\Omega^{(0)} = \Omega^{init}$

$S = n^{-1}\mathbf{Y}^T\mathbf{Y}$, $\lambda =$ Maximum eigenvalue of nS

$\alpha = ((\alpha_{ij}))_{i,j=1}^{p}$ with $\alpha_{ij} = \sum_{j' \neq j} \omega_{ij}$, $\beta = (\beta_i)_{i=1}^{p}$ with $\beta_i = \sum_{i' \neq i} |\omega_{ij}|$

$k \leftarrow \varepsilon\lambda$, $k^* \leftarrow 2k + \tau^{-2}$

for iter $= 1, 2, \ldots, T$ **do**

if $p > n$ **then** ▷ **Sampling Augmented variable** R

$R \leftarrow \mathcal{MN}_{p\times p}\big(\Omega^{(iter-1)}(kI_p - nS),\, I_p,\, kI_p - nS\big)$ using Lemma 6.

else

$R \leftarrow \mathcal{MN}_{p\times p}\big(\Omega^{(iter-1)}(kI_p - nS),\, I_p,\, kI_p - nS\big)$ directly

end if

$\Omega^{new} \leftarrow \Omega^{(iter-1)}$

for $i = 1, 2, \ldots, p-1$ **do** ▷ **Updating off-diagonals of** Ω

for $j = i+1, \ldots, p$ **do**

$l \leftarrow -\min(\alpha_{ij}, \alpha_{ji})$, $H_\Phi \leftarrow \Phi\Big(-\frac{R_{ij}+R_{ji}}{\sqrt{k^*}}\Big) - \Phi\Big(-\frac{R_{ij}+R_{ji}}{\sqrt{k^*}} + \alpha_{ij}\sqrt{k^*}\Big)$

$P(0) \leftarrow 1, \quad P(1) \leftarrow 2\sqrt{\frac{1}{\tau^2 k^*}}\frac{q}{1-q}\exp\Big[\frac{(R_{ij}+R_{ji})^2}{2k^*}\Big] H_\Phi$

if $P(1) = \infty$ **then**

$\omega_{ij}^{new} \leftarrow \mathcal{N}\left(\frac{R_{ij}+R_{ji}}{k^*}, \frac{1}{k^*}; l, 0\right)$, $\omega_{ji}^{new} \leftarrow \omega_{ij}^{new}$

else

$P \leftarrow P/\mathrm{sum}(P)$

$\omega_{ij}^{new} \sim P(0)\delta_0 + P(1)\mathcal{N}\left(\frac{R_{ij}+R_{ji}}{k^*}, \frac{1}{k^*}; l, 0\right)$, $\omega_{ji}^{new} \leftarrow \omega_{ij}^{new}$

end if

$\Delta_1 = \omega_{ij}^{new} - \Omega_{ij}^{(iter-1)}$

$\beta_i \leftarrow \beta_i - \Delta_1$, $\beta_j \leftarrow \beta_j - \Delta_1$ ▷ **Updating** β

$(\alpha_{i.})_{-j} \leftarrow (\alpha_{i.})_{-j} + \Delta_1$, $(\alpha_{j.})_{-i} \leftarrow (\alpha_{j.})_{-i} + \Delta_1$ ▷ **Updating** α

end for

$\omega_{ii}^{new} \leftarrow \mathcal{N}(\frac{R_{ii}+n}{k}, \frac{1}{k}; \beta_i, \infty)$ ▷ **Updating diagonals of** Ω

$\Delta_2 \leftarrow \omega_{ii}^{new} - \Omega_{ii}^{(iter-1)}, \alpha_{i.} \leftarrow \alpha_{i.} + \Delta_2$ ▷ **Updating** α

end for

Repeat Steps 27-28 for $i = p$

$\Omega^{(iter)} \leftarrow \Omega^{new}$

end for

return $\Omega^{(1)}, \Omega^{(2)}, \ldots, \Omega^{(T)}$

end procedure

Algorithm S3 Interweaving Gibbs Sampler

procedure I-GIBBS($\Omega^{init}, \mathbf{Y}, T, \varepsilon, q, \tau^2, N$)

$\Omega^{(0)} = \Omega^{init}$

$S = n^{-1}\mathbf{Y}^T\mathbf{Y}$, $\lambda =$ Maximum eigenvalue of nS

$\alpha = ((\alpha_{ij}))_{i,j=1}^{p}$ with $\alpha_{ij} = \sum_{j' \neq j} \omega_{ij}$, $\beta = (\beta_i)_{i=1}^{p}$ with $\beta_i = \sum_{i' \neq i} |\omega_{ij}|$

for $iter = 1, 2, \ldots, T$ **do**

if iter $\equiv 0 \pmod{N}$ **then**

Perform Steps 6-27 of D-GIBBS in Algorithm S1 with $\Omega^{init} = \Omega^{(iter-1)}$

else

Perform Steps 7-31 of A-GIBBS in Algorithm S2 with $\Omega^{init} = \Omega^{(iter-1)}$

end if

$\Omega^{(iter)} \leftarrow$ sampled matrix in Steps 6-10

end for

return $\Omega^{(1)}, \Omega^{(2)}, \ldots, \Omega^{(T)}$

end procedure

S.2.3 Numerical Experiments

We evaluate the proposed samplers for graph recovery and precision matrix estimation under diagonal dominance, comparing them with GLASSO, CLIME, and DDGL, the diagonally dominant analogue of GGL. Since the true precision matrix corresponding to the line-graph generating model in Section 5 is diagonally dominant, we use the same line-graph setting and synthetic data-generation scheme considered earlier. For each fixed (n, p) configuration, 50 independent datasets are generated with $6n/5$ observations, of which n are used for model fitting and the remaining $n/5$ serve as test data for tuning. For the proposed samplers, the slab inclusion probability q is selected by grid search over $\{1/p, 1/p^{1.5}, 1/p^2\}$ using the held-out data, as described in Section 4.4, while the regularization parameters of the benchmark methods are selected by cross-validation over the ranges recommended by their respective authors. The resulting MCC, F-score, and relative estimation errors are reported in Tables S3, S4, and S5, respectively.

It is worth noting that, the proposed Gibbs samplers consistently achieve better sparsity recovery and precision matrix estimation than the benchmark methods, with substantial improvements in MCC, F-score, and relative estimation error. Among the three proposed samplers, I-Gibbs generally performs best, outperforming both D-Gibbs and A-Gibbs across the considered settings. As discussed in Section S.3.2, D-Gibbs exhibits instability in a few simulation runs, particularly for $n = 500$, which is reflected in the relative estimation errors reported in Table S5. By alternating between D-Gibbs and A-Gibbs updates, the interweaving I-Gibbs sampler mitigates this instability while retaining the complementary local and global exploration of the two

Table S1: *Matthews correlation coefficient (MCC)* for the benchmark and proposed algorithms across six (n, p) configurations, where n and p denote the sample size and number of variables, respectively, under the Line Graph generating model under **diagonal dominance constraint** on the precision matrix.

(n,p)	GLASSO	CLIME	DDGL	D-Gibbs	A-Gibbs	I-Gibbs
$(100, 20)$	0.339362	0.162940	0.575641	0.936973	0.925525	**0.938925**
$(100, 50)$	0.344382	0.366485	0.486751	**0.921028**	0.916477	0.919972
$(100, 100)$	0.338153	0.469339	0.420468	0.896818	0.883941	**0.896887**
$(100, 200)$	0.340614	0.545414	0.357547	0.846959	0.811773	**0.852325**
$(500, 100)$	0.324436	0.442767	0.390768	0.909015	0.951251	**0.956816**
$(500, 250)$	0.328980	0.472171	0.309920	0.950397	0.965262	**0.965424**

Table S2: *F-scores* for the benchmark and proposed algorithms across six (n, p) configurations, where n and p denote the sample size and number of variables, respectively, under the Line Graph generating model under **diagonal dominance constraint** on the precision matrix.

(n,p)	GLASSO	CLIME	DDGL	D-Gibbs	A-Gibbs	I-Gibbs
$(100, 20)$	0.346516	0.223613	0.569840	0.942625	0.931548	**0.944377**
$(100, 50)$	0.269845	0.290592	0.422301	0.923891	0.919400	**0.924868**
$(100, 100)$	0.235832	0.384320	0.324350	0.898747	0.885914	**0.898773**
$(100, 200)$	0.227745	0.482143	0.242386	0.846669	0.813571	**0.846946**
$(500, 100)$	0.222446	0.350902	0.290109	0.909273	0.952057	**0.956747**
$(500, 250)$	0.208240	0.372885	0.187231	0.949986	0.964967	**0.965110**

Table S3: *Relative Estimation Errors* (in Frobenius norm) for the benchmark and proposed algorithms across six (n, p) configurations, where n and p denote the sample size and number of variables, respectively, under the Line Graph generating model under **diagonal dominance constraint** on the precision matrix.

(n,p)	GLASSO	CLIME	DDGL	D-Gibbs	A-Gibbs	I-Gibbs
$(100, 20)$	0.402249	0.580012	0.247147	0.165747	0.161152	**0.160990**
$(100, 50)$	0.537159	0.435991	0.297830	0.180423	**0.172346**	0.174056
$(100, 100)$	0.577636	0.416646	0.341433	0.201587	0.196951	**0.193663**
$(100, 200)$	0.564507	0.395472	0.370931	0.229232	0.237152	**0.228886**
$(500, 100)$	0.389555	0.130876	0.119532	0.368408	0.074493	**0.074262**
$(500, 250)$	0.388881	0.160640	0.136523	0.228786	**0.079839**	0.080578

samplers, leading to improved overall performance. Moreover, the diagonally-dominant formulations consistently outperform their positivity-constrained counterparts across nearly all settings and evaluation metrics. This behavior is expected, since the data-generating precision matrix satisfies diagonal dominance, allowing the additional structural constraint to improve estimation efficiency and graph recovery accuracy.

S.3 Additional Experimental Results

S.3.1 Proportion of Positive Definite Samples in Proposed Gibbs Samplers

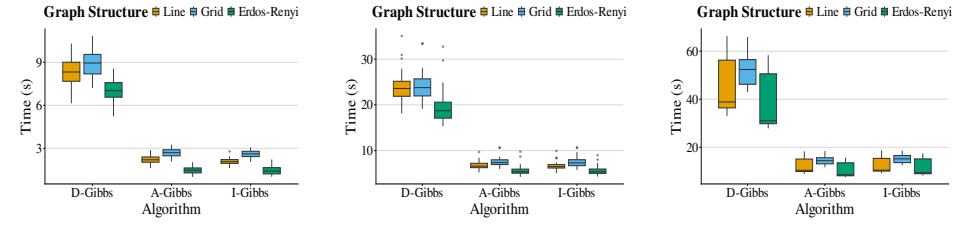
As discussed in Section 3, the proposed Gibbs samplers do not guarantee that either posterior draws or point estimates are positive definite. This relaxation, however, yields substantial computational gains by enabling efficient posterior sampling through Gibbs updates. In Table S1, we report the proportion of positive definite posterior draws under the simulation settings considered in Section 5.1. For the purely augmented sampler (A-Gibbs) and the interweaving sampler (I-Gibbs), a substantial fraction of posterior draws are positive definite in almost all settings. In contrast, the corresponding proportions for the direct Gibbs sampler (D-Gibbs) vary significantly- sometimes they are high, and sometimes they are substantially lower. These findings suggest that augmentation not only improves sparsity selection, but also increases the likelihood that posterior draws satisfy positive definiteness, a desirable structural property of precision matrices.

Table S4: *Proportion of Positive Definite samples* out of the 3000 samples for the proposed Gibbs samplers. The proportion is reported for three data generating models, namely, Line graph, Grid graph and Erdős-Rényi Graph (with 5% edge-density), and six (n, p) configurations.

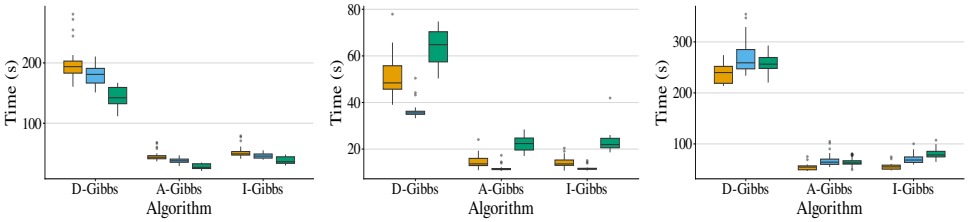
Model	(n, p)	**D-Gibbs**	**A-Gibbs**	**I-Gibbs**
Grid Graph	$(100, 20)$	0.977773	0.987680	0.969973
	$(100, 50)$	0.968927	0.990193	0.989800
	$(100, 100)$	0.944287	0.728107	0.985040
	$(100, 200)$	0.806647	0.924553	0.943500
	$(500, 100)$	1	0.999980	1
	$(500, 250)$	1	1	1
Erdős-Rényi	$(100, 20)$	0.981633	0.984380	0.988627
	$(100, 50)$	0.967980	0.992453	0.984093
	$(100, 100)$	0.544607	0.534313	0.949600
	$(100, 200)$	1	1	1
	$(500, 100)$	0.999100	0.998467	0.999980
	$(500, 250)$	1	1	1
Line Graph	$(100, 20)$	0.928573	0.943060	0.955793
	$(100, 50)$	0.607133	0.681673	0.738800
	$(100, 100)$	0.157527	0.508880	0.405813
	$(100, 200)$	0.130827	0.864593	0.373833
	$(500, 100)$	1	1	1
	$(500, 250)$	1	1	1

S.3.2 Comparison of Execution Time of D-Gibbs, A-Gibbs and I-Gibbs

As discussed previously in Lemma 4 and 7, the computational complexity of D-Gibbs is $\mathcal{O}(p^3)$, and that of A-Gibbs is $\mathcal{O}(\min\{np^2, p^3\})$, which creates a computational bottleneck for D-Gibbs and provides A-Gibbs a significant advantage over it, especially when $p \gg n$. Furthermore, even when $n > p$, for which D-Gibbs and A-Gibbs have the same asymptotic time complexity, Figure S1 shows that D-Gibbs incurs a higher wall-clock runtime. This difference is primarily attributable to the dependence among the full conditional distributions of the entries of Ω, which necessitates sequential updates in D-Gibbs. In contrast, the proposed data augmentation scheme enables joint updates in A-Gibbs through optimized matrix operations implemented using efficient C++ linear algebra libraries. We later introduced I-Gibbs to periodically induce the local behavior of D-Gibbs in the chain and avoid the computational burden simultaneously. Below, we provide a comparison of the average time needed to execute the 3 Gibbs samplers in line, grid and Erdős-Rényi model for all the 6 combinations of (n, p). The average is taken over the 50 datasets, and the full comparison is provided in Figure S1.



(a) $(n, p) = (100, 20)$ (b) $(n, p) = (100, 50)$ (c) $(n, p) = (100, 100)$



(d) $(n, p) = (100, 200)$ (e) $(n, p) = (500, 100)$ (f) $(n, p) = (500, 250)$

Fig S1 Comparison of average runtimes (in seconds) of the 3000 iterations of D-Gibbs, A-Gibbs and I-Gibbs samplers, averaged over 50 replications for 3 different models (line, grid and Erdős-Rényi graph) under 6 different values of (n, p).

Figure S1 clearly shows the advantage of the A-Gibbs and I-Gibbs sampler over the D-Gibbs sampler in terms of their runtimes for all the settings, supporting the order of magnitude difference between the computational complexities of D-Gibbs and A-Gibbs. The runtimes of the A-Gibbs and the I-Gibbs sampler are quite similar for

smaller p, and although for larger values of p, the runtime for I-Gibbs is slightly higher than A-Gibbs, it is still negligible. This is expected, as I-Gibbs replaces only 1% of A-Gibbs iterations by D-Gibbs ones.

S.3.3 Comparison of Area Under Receiver Operating Characteristic Curve (AUROC)

The Receiver Operating Characteristic (ROC) curve plots the true-positive rate against the false-positive rate across classification thresholds. The area under the ROC curve (AUROC) summarizes overall discriminatory performance, with values closer to 1 indicating better classification.

To compute the AUROC, we consider a sequence of 50 hyperparameter values. For each of the 18 simulation settings (three data-generating models and six pairs of (n, p) values), we evaluate AUROC over the 50 hyperparameter values, and average them across 50 independently generated datasets. We restrict the ROC curve to the region FPR ≤ 0.15 and compute the corresponding partial AUROC. To facilitate comparisons on the unit scale, the resulting partial AUROC is normalized by dividing by 0.15. The normalized AUROC values for GLASSO, CLIME, D-Gibbs, A-Gibbs, and I-Gibbs are reported in Table S2. GGL is intentionally omitted from this comparison because its intrinsic sparsity constraints enforce sparsity across the range of hyperparameter values and often prevent the FPR from reaching the 0.15 cutoff, and therefore preclude a directly comparable partial-AUROC calculation.

Across all 18 settings the methods generally achieved high AUROC values. The sole exception occurred under the Erdős–Rényi generating model with $(n, p) = (100, 200)$, where every method performed worse than random guessing. Among the proposed samplers, **I-Gibbs exhibited particularly strong performance** for the **grid** and **Erdős–Rényi** models. Although CLIME generally yielded slightly higher AUROC values in the line-graph setting, the differences were negligible and typically appeared only beyond the third decimal place. A similar pattern holds for the few instances where GLASSO achieved a modest edge. Within our proposed family of Gibbs samplers, D-Gibbs generally exhibited the weakest performance, A-Gibbs yielded substantial improvements over D-Gibbs, and I-Gibbs provided a further improvement over A-Gibbs.

S.4 Additional Details on Financial Data Analysis in Section 5.2

In this section, we report the marginal inclusion probabilities for all possible edges in the graphs estimated using our proposed Gibbs samplers. We partition the interval [0,1] into four equal-width intervals: [0,0.25), [0.25,0.5), [0.5,0.75), and [0.75,1]. The edges are further classified as within-sector or between-sector. After discarding the first 2000 iterations as burn-in, we compute the marginal inclusion probability of each edge as the proportion of the remaining 1000 posterior samples in which that edge is included. Table S6 reports the resulting inclusion probabilities for each edge class.

Table S5: *AUROC (Area Under the Receiver Operating Characteristic curve)* for the benchmark and proposed algorithms across three data-generating models, namely, Line Graph, Grid Graph, and Erdős–Rényi Graph (5% edge density), and six (n, p) configurations.

Model	(n,p)	GLASSO	CLIME	D-Gibbs	A-Gibbs	I-Gibbs
Grid Graph	$(100, 20)$	0.554125	0.816922	0.780880	0.833244	**0.842288**
	$(100, 50)$	0.755612	0.784037	0.747884	0.836661	**0.851225**
	$(100, 100)$	0.819931	0.819204	0.697720	0.815764	**0.828695**
	$(100, 200)$	**0.860535**	0.809257	0.696910	0.812945	0.832086
	$(500, 100)$	0.950085	0.994948	0.990655	0.993990	**0.996140**
	$(500, 250)$	0.983930	0.994093	0.984323	0.991919	**0.994223**
Erdős-Rényi	$(100, 20)$	0.936876	0.974328	0.815850	0.983847	**0.986017**
	$(100, 50)$	0.686814	0.749381	0.647868	0.758833	**0.769337**
	$(100, 100)$	0.548982	0.512198	0.375121	0.541621	**0.552572**
	$(100, 200)$	**0.284322**	0.217657	0.121556	0.273471	0.277320
	$(500, 100)$	0.841647	0.925679	0.878911	0.929387	**0.940927**
	$(500, 250)$	0.529413	0.492594	0.313160	0.527501	**0.535772**
Line Graph	$(100, 20)$	0.811230	0.957063	0.947920	**0.975754**	0.974224
	$(100, 50)$	0.928225	**0.979901**	0.950205	0.974808	0.976180
	$(100, 100)$	0.960837	**0.981858**	0.948411	0.970938	0.973246
	$(100, 200)$	**0.954007**	0.948304	0.915276	0.931987	0.939708
	$(500, 100)$	0.983510	**0.999860**	0.995035	0.990166	0.997524
	$(500, 250)$	0.996227	**0.999957**	0.997976	0.999267	0.999080

D-Gibbs shows a clear trade-off between sensitivity to *within-sector* structure and control of spurious *between-sector* connections. It assigns more *within-sector* edges to the high-probability intervals [0.5,0.75) and [0.75,1] than A-Gibbs and I-Gibbs, indicating stronger recovery of within-sector connectivity. This gain, however, comes at the cost of a substantially larger number of high-probability *between-sector* edges. In contrast, A-Gibbs and I-Gibbs are more selective, retaining fewer between-sector connections while preserving the dominant within-sector structure. This sharper separation between within-sector and between-sector connectivity is reflected in their markedly higher modularity compared with D-Gibbs.

Table S6: Table of *number of edges* belonging to different ranges of *Proportion of within-sector edges*. The proportion of *within-sector* and *between-sector* edges out of the 1000 posterior samples (after discarding the first 2000 samples as burn-in) has been divided into 4 separate classes, and the number of edges in the estimated precision matrix belonging to each of the classes has been reported for all 3 proposed Gibbs samplers.

	D-Gibbs		A-Gibbs		I-Gibbs	
	Within-sector	**Between-sector**	**Within-sector**	**Between-sector**	**Within-sector**	**Between-sector**
0 – 0.25	8031	31073	8353	32107	8155	32075
0.25 – 0.5	64	285	4	2	10	9
0.5 – 0.75	43	100	2	2	9	4
0.75 – 1	434	725	213	72	398	95

References

Chen, L., Acharyya, S., May, A. M., Udager, A. M., Keller, E. T. & Baladandayuthapani, V. (2026), 'Multi-resolution Spatial Graphical Regression Models for Hierarchical Spatial Transcriptomics Data', *arXiv preprint arXiv:2605.16804*.

Zavlanos, M. M., Julius, A. A., Boyd, S. P. & Pappas, G. J. (2011), 'Inferring stable genetic networks from steady-state data', *Automatica* **47**(6), 1113–1122.