Mathematical Biosciences and Engineering

https://www.aimspress.com/journal/mbe

MBE, 23(2): 499–546.
DOI: 10.3934/mbe.2026020
Received: 01 July 2025
Revised: 10 November 2025
Accepted: 25 November 2025
Published: 12 January 2026

Research article

Bayesian chain graph models to characterize microbe-environment dynamics

Yunyi Shen[1] **and Claudia Solís-Lemus**[2,*]

[1] Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA

[2] Department of Plant Pathology, Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, USA

* **Correspondence:** Email: solislemus@wisc.edu.

Abstract: Microbiome data require statistical models that can simultaneously decode microbes' reaction to the environment and interactions among microbes. While a multiresponse linear regression model seems like a straight-forward solution, we argue that treating it as a graphical model is problematic given that the regression coefficient matrix does not encode the conditional dependence structure between response and predictor nodes. This observation is especially important in biological settings when we have prior knowledge on the edges from specific experimental interventions that can only be properly encoded under a conditional dependence model. Here, we propose a chain graph model with two sets of nodes (predictors and responses) whose solution yields a graph with edges that indeed represent conditional dependence, thus agreeing with the experimenter's intuition on the average behavior of nodes under treatment. The solution to our model is sparse via the Bayesian linear regression (LASSO). In addition, we propose an adaptive extension so that different shrinkages can be applied to different edges to incorporate edge-specific prior knowledge. Our model is computationally inexpensive through an efficient Gibbs sampling algorithm and can account for binary, counting, and compositional responses via an appropriate hierarchical structure. We test the performance of our model in a variety of simulated datasets, thereby showing superior performance to state-of-the-art approaches. We further apply our model to human gut and soil microbial compositional datasets, and we highlight that CG-LASSO can estimate biologically meaningful network structures in the data. Our software is available as an R package at https://github.com/YunyiShen/CAR-LASSO.

Keywords: linear regression; compositional data; interaction network; microbiome; direct effects

1. Introduction

Microbial communities are among the main driving forces of all biogeochemical processes on Earth. On one side, many critical soil processes, such as mineral weathering and soil cycling of mineral-sorbed organic matter, are governed by mineral-associated microbes [1–5]. On another side, the plant and soil microbiome drive phenotype variations related to plant health and crop production [6, 7]. In addition, as evidenced by The Human Microbiome Project [8], the microbes that live in the human body are key determinants of human health and disease [9].

Understanding the composition of microbial communities and what environmental or experimental factors play a role in shaping this composition is crucial to comprehend biological processes in humans, soil, and plants alike, and to predict microbial responses to environmental changes. However, the interconnectivity of the microbes-environment is still not fully understood. One of the reasons for this gap in knowledge is the lack of statistical tools to simultaneously infer connections among microbes and their direct reactions to different environmental factors in a unified framework.

On the surface, a statistical solution to this problem seems straight-forward. On one side, a multiresponse linear regression model (sometimes denoted covariate-adjusted Gaussian graphical models [10]) with microbial abundances as responses and environmental features as predictors can allow us to estimate links between the environment and the microbial compositions. On the other side, graphical LASSO [11, 12] allows us to infer sparse partial correlations among the microbial abundances that represents the microbial community as a network. Thus, intuitively, the combination of a multiresponse linear regression model to infer links between environment and microbes, and a graphical LASSO model to infer interactions among microbes would provide a framework to investigate microbe-microbe and microbe-enviroment interactions. However, in this formulation the regression coefficients between responses (microbes) and predictors (environment) represent *marginal* effects rather than *direct* effects.

The distinction between marginal effect and direct effect is crucial when we would like to biologically interpret the result or to include prior biological knowledge to the model [13]. For instance, penicillin has no biological effect on Gram-negative bacteria, yet it might still promote the abundance of such bacteria by inhibiting their Gram-positive competitors. In this example, penicillin has no *direct* effect on Gram-negative bacteria, but it may have a *marginal* effect on them (via interaction among microbes). Figure 1(B) illustrates this case. In this figure, X represents penicillin, Y_1 represents the relative abundance of Gram-positive bacteria, and Y_2 represents the relative abundance of Gram-negative bacteria. There is not a direct link between X (penicillin) and Gram-negative (Y_2), and there is a negative (blue) link between penicillin and Gram-positive bacteria (Y_1). Both microbes in this example are competitors, and thus, negatively correlated (with a blue link). In this figure, links correspond to *conditional* dependency and thus, the network can be denoted a *conditional* network. In contrast, a marginal network would display a spurious positive (red) link between penicillin (X) and Gram-negative bacteria (Y_2), since the relative abundance of such bacteria would likely increase by the inhibition of Gram-positive competitors. While the marginal network could be interesting in some settings, we argue that it is not as biologically meaningful as the conditional network here; the standard multiresponse linear regression with graphical LASSO (which produces marginal associations) is not adequate to investigate microbe-environment interactions, especially when the marginal effects of environmental variables are assumed to be sparse.

Here, we introduce a model framework to infer a sparse *conditional* network structure that represents both interactions among microbes and environmental effects. Specifically, our model simultaneously estimates the direct effect of a set of predictors (e.g., diet, weather, experimental treatments) that influence the responses (e.g., abundances of microbes) and connections among the responses. Our model is represented by a chain graph [14, 15] with two sets of nodes: predictors and responses (Figure 1). Directed edges between a predictor and a response represent conditional links, and undirected edges among responses represent correlations. We argue that conditional links, as those in the chain graph in Figure 1, are more biologically meaningful. For example, networks A and C can produce a similar marginal correlation structure between any two nodes. In network D, X and Y_2 are conditionally correlated, yet they might not have a marginal correlation. For example, if Y_1, Y_2 have a marginal variance of 1, a covariance of $\rho = -0.5$, the conditional regression coefficient between Y_1 and X conditioned on Y_2 is $\beta_1 = 2$, and the conditional regression coefficient between Y_2 and X conditioned on Y_1 is $\beta_2 = 1$, then we can show that the marginal regression coefficient between Y_2 and X when integrating out Y_1 is zero ($\rho\beta_1 + \beta_2 = 0$). While chain graph models are not new, we argue that they are under-utilized in microbiome studies, and our work will serve to illustrate their potential to elucidate ties between microbial interactions and experimental or environmental predictors.

Finally, we accompany our model with an open-source easy-to-use software: the novel R package `CARlasso` which is publicly available on GitHub (https://github.com/YunyiShen/CAR-LASSO). This package has a similar structure to existing regression tools in R for the ease of use among the microbiology community.

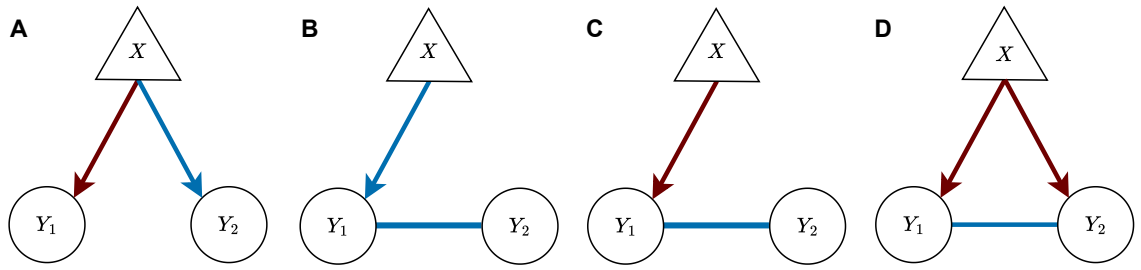

Figure 1. Microbe-microbe and microbe-environment interactions represented as chain graphs. In a chain graph, we use triangles to represent environmental predictors (here, X) and circles to represent the responses Y_1 and Y_2 (e.g., relative abundances for two microbes). Red edges correspond to positive conditional links between nodes while blue edges correspond to a negative conditional links.

2. Methods

2.1. Model specification

We define a chain graph model that takes a matrix of relative abundances and a matrix with environmental (or experimental) predictors as input. Relative abundances are the fraction of sequencing reads in a sample that are assigned to a specific microbial group relative to the total

number of reads, thus representing that microbe's share of the microbial community. The model infers interactions among microbes (as the precision matrix $\mathbf{\Omega}$) and direct effects among environmental predictors and microbial abundances (as the regression coefficient matrix $\mathbf{B}$), and represents both matrices as a chain graph with two types of nodes (Figure 1). Note that the base model is defined on Gaussian responses and does not directly apply to microbial abundances that are compositional. We extend the base normal model to compositional data via a hierarchical structure (see the Appendix A.4).

Let $\mathbf{Y}_i \in \mathbb{R}^k$ be a multivariate response with k entries (microbial units) for $i = 1, \ldots, n$ samples. Let $\mathbf{X}_i \in \mathbb{R}^{1\times p}$ be the row vector of environmental predictors for $i = 1, \ldots, n$ (i.e., the i^{th} row of the input matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$). We assume that the input matrix is standardized so that each column has a mean of 0 and the same standard deviation (set to be 1 in the simulations).

Let $\mathbf{Y}_i$ follow a Normal distribution with mean vector $\mathbf{\Omega}^{-1}(\mathbf{B}^T\mathbf{X}_i^T + \mu)$ and precision matrix $\mathbf{\Omega} \in \mathbb{R}^{k\times k}$, where $\mathbf{B} \in \mathbb{R}^{p\times k}$ corresponds to the regression coefficients that connects the responses ($\mathbf{Y}_i \in \mathbb{R}^k$) and the predictors ($\mathbf{X}_i \in \mathbb{R}^{1\times p}$) and $\mu \in \mathbb{R}^k$ corresponds to the intercept.

While the inclusion of $\mathbf{\Omega}$ in the mean function is unusual in statistical settings, we argue that it is necessary to draw the right biological interpretations – abundance of microbial taxa is determined by both the environmental effect and inter-taxa interactions. Indeed, in this parametrization, $\mathbf{B}$ encodes the *direct effect* of $\mathbf{X}$ on $\mathbf{Y}$ as B_{jq} is the coefficient of product between X_j and Y_q. If $B_{jq} = 0$, then X_j and Y_q are *conditionally independent*. This is analogous to the case of $\mathbf{\Omega}$ whose off-diagonal entries encode the conditional dependence between responses Y_q and $Y_{q'}$. To see the analogy, we provide an interpretation of the parameters in the same manner as in a univariate linear regression. Let $\mathbf{Y}_{-q}$ be the vector of responses without the qth component, let $\mathbf{Y}_{-(q',q)}$ be the vector of responses without the qth and q'th components, and let $\mathbf{X}_{-j}$ be the vector of predictors without the jth component.

$$\begin{aligned}\mathbb{E}[Y_q|X_j = x_j + 1, \mathbf{Y}_{-q}, \mathbf{X}_{-j}, \mathbf{B}, \mathbf{\Omega}] - \mathbb{E}[Y_q|X_j = x_j, \mathbf{Y}_{-q}, \mathbf{X}_{-j}, \mathbf{B}, \mathbf{\Omega}] = B_{jq}/(\omega_{qq}),\\ \mathbb{E}[Y_q|Y_{q'} = y_{q'} + 1, \mathbf{Y}_{-(q',q)}, \mathbf{X}, \mathbf{B}, \mathbf{\Omega}] - \mathbb{E}[Y_q||Y_{q'} = y_{q'}, \mathbf{Y}_{-(q,q')}, \mathbf{X}, \mathbf{B}, \mathbf{\Omega}] = \omega_{qq'}/(-\omega_{qq}).\end{aligned} \tag{2.1}$$

In particular, by fixing the values of all but one predictor and all of the other responses, an increase of one unit in X_j is associated with $B_{jq}/(\omega_{qq})$ unit increase in the expectation of Y_q, hence conditioned on the values of all other responses $Y_{q'}$.

In contrast, in the alternative approach of multiresponse linear regression, the regression coefficients denoted as $\tilde{\mathbf{B}} = \mathbf{B\Omega}$, are marginal effects, that is, $\mathbb{E}[Y_q|X_j = x_j + 1, \mathbf{X}_{-j}, \tilde{\mathbf{B}}, \mathbf{\Omega}] - \mathbb{E}[Y_q|X_j = x_j, \mathbf{X}_{-j}, \tilde{\mathbf{B}}, \mathbf{\Omega}] = \tilde{B}_{jq}$. Note that the crucial difference here is that we do not condition on $\mathbf{Y}_{-q}$ (all other microbes). Specifically within the microbiome the presence of $\mathbf{\Omega}^{-1}$ in the mean represents the ecological knowledge that the responses of a species (e.g., relative abundances of microbes) depend on both its reaction to the environment ($\mathbf{B}$) and interactions with other species ($\mathbf{\Omega}$). In our model, the regression coefficients matrix $\mathbf{B}$ encode conditional dependence among the responses (scaled by variance) and the predictors which, arguably has a more mechanistic interpretation [14, 16, 17]. The key difference between conditional and marginal effects lies in whether we condition on the abundances of other microbes (Y_{-q}). Using an antibiotic treatment as an example, the conditional effect represents a hypothetical scenario in which we hold all other microbial abundances and treatments fixed, and increase the antibiotic level by one unit. It captures the direct effect of the antibiotic on microbe q, independent of changes in the rest of the community. In contrast,

the marginal effect does not condition on other microbes. It reflects the overall effect of antibiotics on microbe q, thus encompassing both direct and indirect effects mediated through interactions among microbes. Because microbial communities are typically highly interconnected, it is generally unrealistic to assume that a treatment would have a zero marginal effect on any single microbe: doing so would imply that all indirect effects through the network of interactions perfectly cancel out.

For more theoretical details on the model specification, prior specification, computational implementation, and extensions, see Appendix A.4.

2.2. Method performance on simulated data

We simulate data under the six graphical structures in [12]: 1) an AR(1) model so that $\mathbf{\Omega}$ is tridiagonal; 2) an AR(2) model such that $\omega_{k,k'} = 0$ whenever $|k - k'| > 2$; 3) a block model so that there are two dense blocks along the diagonal; 4) a star model with every node connected to the first node; 5) a circle model so that the graph forms a circle, and 6) a full dense model. See Figure 2 for a visual description of the six precision matrices ($\mathbf{\Omega}$).

We vary the sparsity of the regression coefficients (environmental effects on microbes) $\mathbf{B}$ (denoted beta sparsity in the figures) with either 80% or 50% entries equal to zero with non-zero entries sampled from a standard Normal distribution. The multivariate response is sampled with dimension $k = 10, 30$ (microbes), $p = 5, 10$ predictors, and $n = 50$ samples. The design matrices are sampled from standard Normal distributions (more details in the Supplementary Material). Each simulation setting was repeated 50 times.

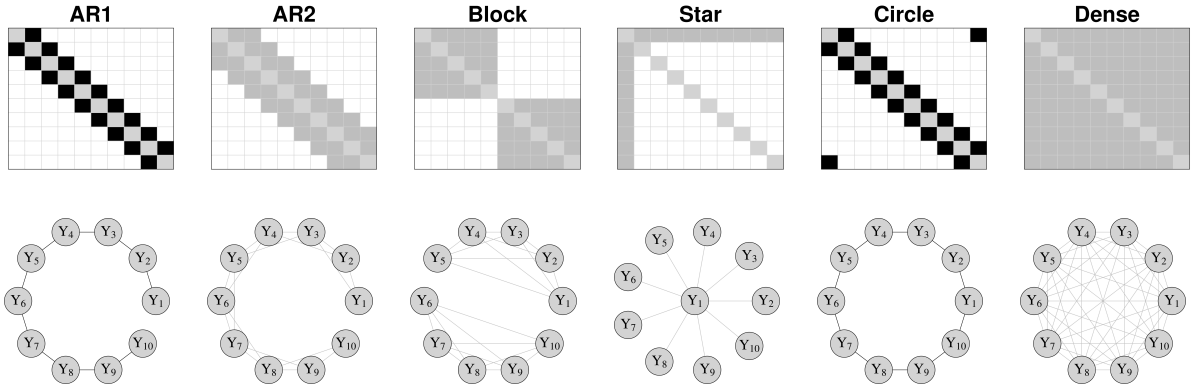

Figure 2. Precision matrices $\mathbf{\Omega}$ and corresponding graphical structure used in the simulations. These six models correspond to the six graphical structures in [12]. Positive entries are represented in gray and negative entries are represented in black.

On the simulated data, we compare the performance of 12 methods that generally fall into two categories: 1) methods that estimate both $\mathbf{B}$ and $\mathbf{\Omega}$ (or that can get the parameters by transformation), and 2) methods that only estimate $\mathbf{\Omega}$.

The methods that estimate both $\mathbf{B}$ and $\mathbf{\Omega}$ are as follows:

1) CG-LASSO: our proposed model, denoted as `CG-LASSO` or `CG` in the figures;
2) Adaptive CG-LASSO: our proposed model with different shrinkage parameters for different entries in $\mathbf{B}$ and $\mathbf{\Omega}$ (`CG-ALASSO` and `CG-A` in the figures);

3) SRG-LASSO: Bayesian LASSO with standard mean-covariance parametrization, that is, sparsity is on $\tilde{\mathbf{B}} = \mathbf{B}\boldsymbol{\Omega}^{-1}$ and $\boldsymbol{\Omega}$ (see Section A.1) (`SRG-LASSO` and `SRG` in the figures);
4) Bayesian multiresponse regression with conjugate priors, (`multireg` in the figures), and
5) Bayesian multiresponse regression with conjugate priors that assume the marginal mean is 0 similar to Graphical LASSO (`multireg_mu0` in the figures).

Note that methods 3–5 are based on the idea of fitting the marginal model (similar to covariate-adjusted GGMs, e.g., [18, 19]) and getting the chain graph parameter by transforming $\mathbf{B} = \tilde{\mathbf{B}}\boldsymbol{\Omega}$. While methods 4–5 put do not provide sparsity assumptions, method 3 indeed imposes sparsity on $\tilde{\mathbf{B}}$ rather than on $\mathbf{B}$. All of our comparisons in the simulations are based on the conditional parameter $\mathbf{B}$; however see the results on real data (Section 2.3) for comparisons among the marginal, the conditional, and the conditional obtained from the marginal parameters.

The methods that only estimate $\boldsymbol{\Omega}$ are as follows:

6) Graphical LASSO in [12], (`GLASSO` in the figures);
7) Adaptive Graphical LASSO: adaptive version in [12] (`GALASSO` in the figures);
8) Augmented Graphical LASSO: Graphical LASSO including responses and predictors assumed Normally distributed (denoted as `GLASSO-aug` in the figures);
9) Adaptive version of Augmented Graphical LASSO (`GALASSO-aug` in the figures);
10) Bayesian multiresponse regression with conjugate priors (Wishart prior on the precision matrix and a Normal prior on the mean) that assume the marginal mean is 0 (similar to Graphical LASSO), but using all the responses and predictors as "responses" in the model and no predictors (`multireg_mu0-aug` in the figures);
11) Calculate the inverse of the empirical covariance matrix (`ad-hoc` in the figures), and
12) Calculate the inverse of the empirical covariance matrix with both responses and predictors (`ad-hoc-aug` in the figures).

As in [12] and [20], we set the hyperparameters of the Gamma hyperprior for the shrinkage parameters of both $\mathbf{B}$ and $\boldsymbol{\Omega}$ as $r = 0.01, \delta = 10^{-6}$ for the adaptive versions, and $r = 1, \delta = 0.01$ for the non-adaptive versions. In the multiresponse regression models (methods 8-10), we consider any edge with weight $< 1 \times 10^{-3}$ to be 0.

To evaluate the performance of the methods, we compute the Frobenius loss of the estimate of $\mathbf{B}$ and the Stein's loss of the estimate of $\boldsymbol{\Omega}$. We use Stein's loss for $\boldsymbol{\Omega}$ since it is the KL-divergence when the mean vector is 0. In addition, we evaluate the reconstruction of the graphical structures based on the Matthews correlation coefficient (MCC) [21], which ranges from -1 to 1, with 1 representing a perfect prediction.

Finally, we calculate the proportion of true positive edges and false positive edges in the reconstructed graphs (both $\mathbf{B}$ and $\boldsymbol{\Omega}$). The true positive rate is calculated as the proportion of times a true edge is reconstructed, and the false positive rate is calculated as the proportion of times an edge appears in the estimated graph that is not present in the true graph. The false positive rate is presented as a negative quantity in the figures aligned with standard network reconstruction practices [22].

2.3. Application to real microbiome data

2.3.1. Human gut microbiome compositional data

The microbiome of older people displays greater inter-individual variations than that of younger adults. The study in [23] collected the faecal microbiome composition of 178 elderly subjects, together with the subjects' residence type (in the community, day-hospital, rehabilitation, or in long-term residential care) and diet (data at [24]). Researchers studied the correlation between microbes and other measurements. They found that the individual microbiomes of people in long-stay care were significantly less diverse, and this loss of diversity might associate with increased frailty. They clustered microbes based on co-abundances and performed dimension reduction techniques to infer relationships between composition and health. However, co-abundances might not appropriately infer interactions because of the existence of other microbes and the environment [25]. Partial correlations between microbes and the environment or other microbes are more ecologically meaningful. Here, we infer the partial correlation between environments and among microbes in those elderly subjects by reconstructing a sparse network via the adaptive CG-LASSO model.

We use the MG-RAST server [26] for profiling, with an e-value of 5, and identity of 60%, an alignment length of 15 bp, and a minimal abundance of 10 reads. Unclassified hits are not included in the analysis. A genus with more than 0.5% (human) or 1% (soil) relative abundance in more than 50 samples is selected as the focal genus, while all other genus serve as the reference group.

We reconstruct the weighted graph using the conditional regression coefficient between any two nodes. The $\alpha-$centrality [27] is used to identify the importance of the nodes. A weighted adjacency matrix is constructed with the posterior mean of the conditional regression coefficients of those that showed significance with the horseshoe method described in Section A.3.3.

2.3.2. Soil microbiome compositional data

The objective of this study [28, 29] is to examine the soil microbial community composition and structure of both bacteria and fungi at a microbially-relevant scale. The researchers isolated soil aggregates from three land management systems in central Iowa to test if the aggregate-level microbial responses are related to plant community and management practices. The clean dataset has 120 samples with 17 genera under consideration. We focus on the bacteria to further evaluate the partial association among them and the environmental factors.

We use the MG-RAST server [26] with the same settings as used for the human gut microbiome data described in Section 2.3.1. In addition, a weighted adjacency matrix is constructed with the posterior mean of the conditional regression coefficients of those that showed significance with the horseshoe method described in Section A.3.3.

3. Results

3.1. Method performance on simulated data

3.1.1. Performance on the inference of **B**

We select the four best performing methods based on the MCC plus a Bayesian multiresponse linear regression as a reference (multireg) to plot the true and false positive rates in the estimation

of the regression coefficient edges (**B**). Figure 3 shows the accuracy of reconstruction for **B** for the simulation study with $k = 10$ nodes, $p = 10$ predictors and a sparsity level of 0.8 (see Figures A1–A3 for other number of nodes, predictors and sparsity). Adaptive CG-LASSO (CG-A) produces the lowest false positive rate (represented as light blue entries) compared to the other methods, with the Bayesian multiresponse linear regression having the highest false positive rates. As shown in Figure 3, the circle model is particularly difficult for SRG-LASSO, with higher false positive rates than other precision matrices.

Our proposed models (CG-LASSO and adaptive CG-LASSO) outperform the other models to reconstruct **B**, as evaluated by the MCC in Figure A4 for $k = 10$. For $k = 30$, both CG-LASSO and SRG-LASSO have comparable performances (Figure A5), but adaptive CG-LASSO continues to outperform all other methods in almost every combination of sparsity settings on **B** and structure of $\mathbf{\Omega}$ (Figure 2). Regarding the Frobenius loss of the estimate of **B** (Figure A6 for $k = 10$ and Figure A7 for $k = 30$), both adaptive and non-adaptive CG-LASSO outperform all other methods. This is particularly true for the circle model for which the Frobenius loss of SRG-LASSO is much higher than any other method.

Note that SRG-LASSO and the Bayesian multiresponse linear regression use the formulation of covariate-adjusted GGMs to fit the marginal model and obtain the conditional regression coefficients by transformation. The simulation results show that SRG-LASSO and the Bayesian multiresponse linear regression models have lower MCCs (Figure A4) and high false positive rates (Figure 3); thus, these models are not very effective to infer **B** when **B** is indeed sparse, which justifies the use of the chain graph model (CG-LASSO). After all, the chain graph model puts a different sparsity assumption and has a different interpretation than the marginal model and the covariate-adjusted GGM.

3.1.2. Performance on the inference of $\mathbf{\Omega}$

In terms of false positive rates in the reconstruction of $\mathbf{\Omega}$, we select the four best methods in terms of the MCC: adaptive CG-LASSO, CG-LASSO, SRG-LASSO, and augmented GLASSO, alongside the Bayesian multiresponse linear regression as a reference (multireg). Figure 4 (for $k = 10$ nodes, $p = 10$ predictors and a sparsity level of 0.8) shows that adaptive CG-LASSO outperforms all other methods in terms of controlled false positive edges, with the Bayesian multiresponse linear regression performing the worst. See Figures A8, A9, and A10 for other numbers of nodes, predictors, and sparsity levels.

Additionally, our simulations show that adaptive CG-LASSO outperforms all other alternatives to accurately reconstruct the structure of $\mathbf{\Omega}$ in most settings (evaluated by MCC in Figures A11 for $k = 10$, and in Figure A12 for $k = 30$). Non-adaptive CG-LASSO and SRG-LASSO show comparable performance, as well as augmented GLASSO. The star model proves to be equally difficult for all methods. In the star model, the off diagonal signal is weak compared to other models (off diagonal entries are close to 0.1 while diagonal entries are close to 1). Moreover, it is the sparsest setting, with 80% of the off diagonal entries to be zero when $k = 10$ and 93% when $k = 30$. This setup might cause penalized methods to over penalize the off diagonal entries. Additionally, the AR2 model for $k = 30$ nodes (Figure A12) proves to be difficult for all methods, except for the adaptive CG-LASSO.

It is worth noting that (augmented) graphical LASSO shows good performance in most cases (Figure A11). However, this method assumes that the joint distribution of responses and predictors is Normal. This is the case in our simulation setup, and hence, the good performance of the method. However, our CG-LASSO model is more flexible because it does not make any assumptions on the

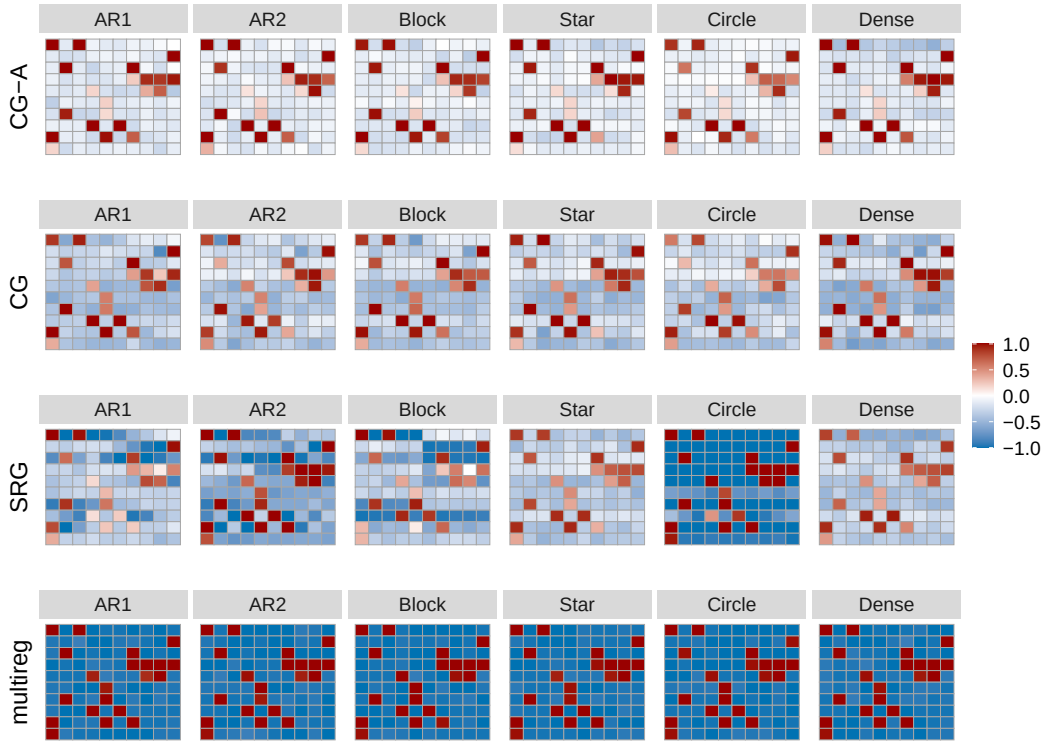

Figure 3. Reconstruction accuracy of the graph between responses and predictors (**B**) for $k = 10$ nodes, $p = 10$ predictors and sparsity of 0.8. Red entries correspond to true positive edges and blue entries correspond to false positive edges. Our proposed method Adaptive CG-LASSO (CG-A) outperforms the other methods by displaying the lowest false positive rate (lighter blue) across all precision matrices (columns).

design matrix.

In terms of the Stein's loss of the estimate of $\mathbf{\Omega}$, (adaptive) CG-LASSO, SRG-LASSO, and augmented GLASSO all have comparable performances with losses close to zero under most settings (Figure A13 for $k = 10$ and Figure A14 for $k = 30$).

3.1.3. Computational speed and scaling

We test the scalability of our estimation procedure by simulating 500 and 1000 samples with 5, 10, 25, 50, and 100 nodes. We sample 1000 generations with 100 burn-in on a machine with a Core-i7 4790 CPU and the Windows 7 operating system. We recorded the CPU seconds in R. While our models are slower than Graphical LASSO or multiresponse regression, running time is not severely impacted by sample size (Figure A15). Instead, speed is mostly influenced by the number of nodes and the number of predictors. However, even the case of 100 nodes and 10 predictors is successfully completed in less than 10 minutes.

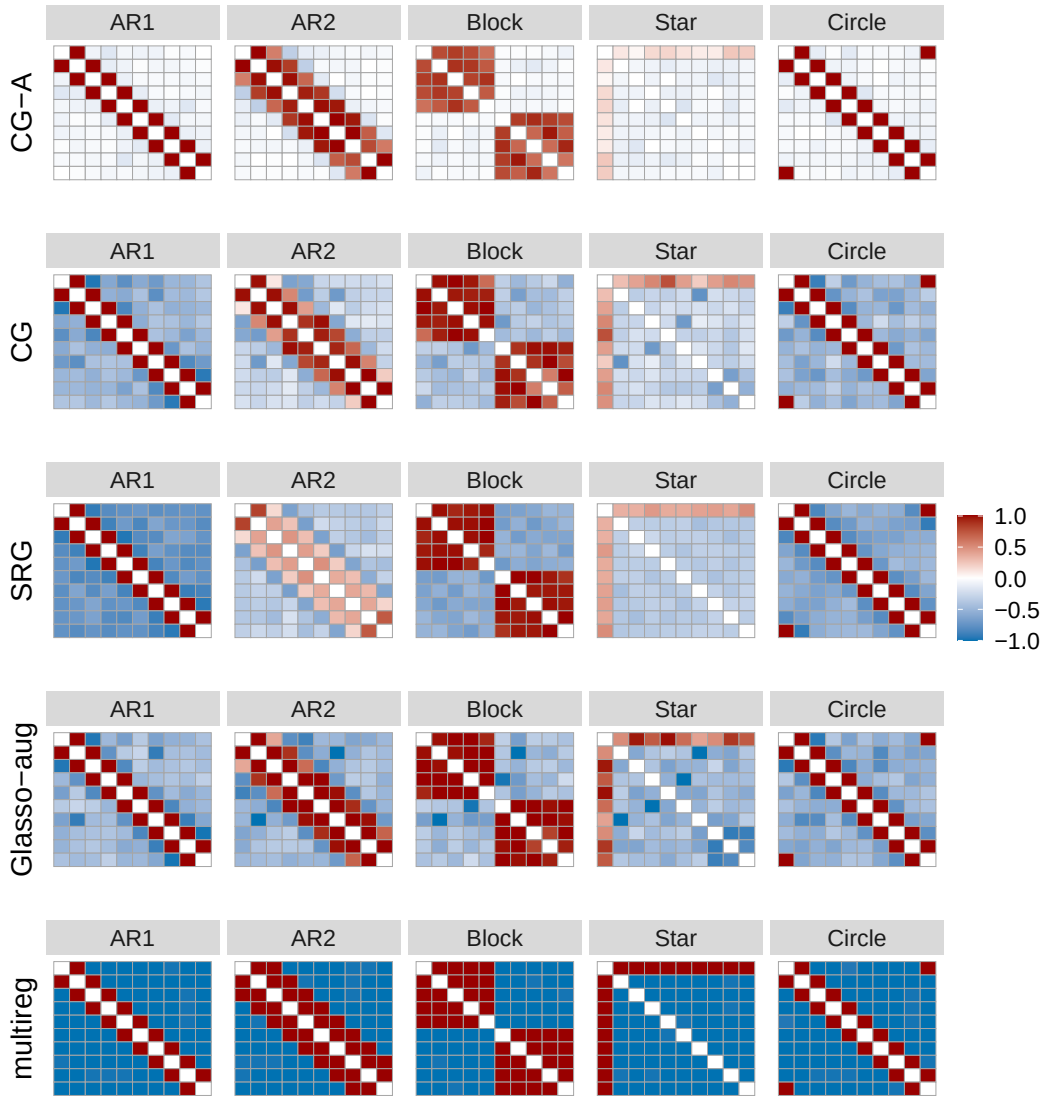

Figure 4. Reconstruction accuracy of the graph among responses ($\mathbf{\Omega}$) for $k = 10$ nodes, $p = 10$ predictors and a sparsity of 0.8. Red entries correspond to true positive edges and blue entries correspond to false positive edges. Our proposed method Adaptive CG-LASSO (CG-A) outperforms the other methods by displaying the lowest false positive rate (lighter blue). We omit the dense model because it has no false positive or true negatives.

3.2. *Application to real microbiome data*

3.2.1. Human gut microbiome compositional data

Figure 5 shows the estimated human gut microbiome network under the adaptive CG-LASSO model where the edges with the most weight correspond to connections between genus nodes, not so much with predictors. The most important predictor is whether the patient's residence was a long-term residential care, which positively affected genus *Caloramator*. This result agrees with the original analysis that also separated elderly subjects based upon where they live in the community. Another important predictor is diet group 4, which corresponds to the high fat/low fiber group. This diet positively affected genus *Caloramator* as well.

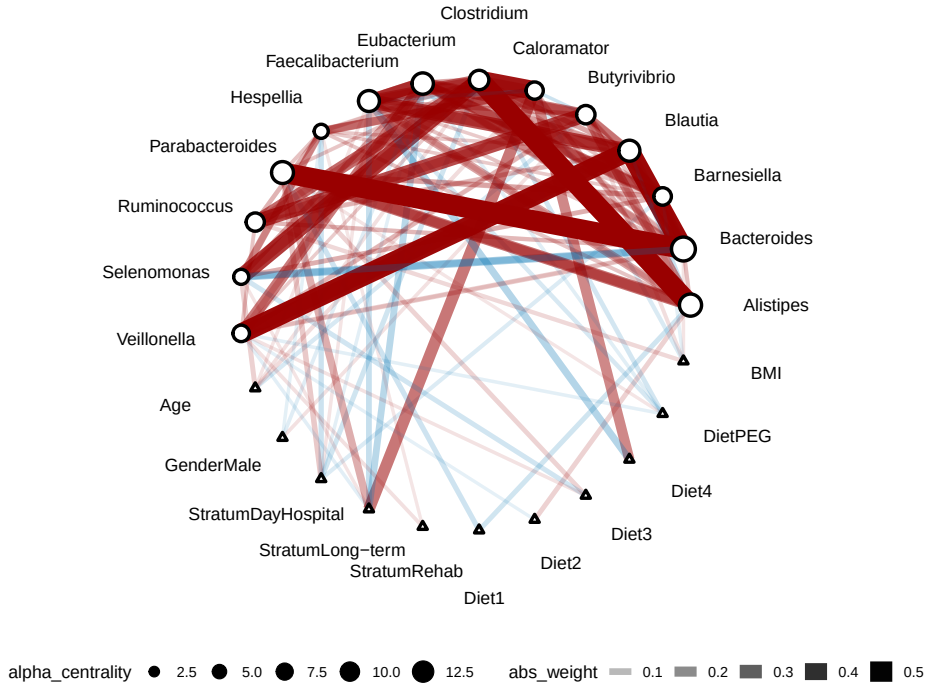

Figure 5. Reconstructed genus conditional network for human gut microbiome using the adaptive CG-LASSO model. Triangle nodes correspond to predictors and circle nodes correspond to relative abundances of the genera. The node size on the circle nodes correspond to the $\alpha-$centrality values [27]. The width of the edges correspond to the absolute weight, and the color to the type of interaction (red positive, blue negative).

Comparison with a marginal network. As a comparison, in Figure 6, we estimate the marginal network to observe the differences with the conditional network in Figure 5. To obtain the marginal network, we start with the conditional network and use the equivalence $\tilde{\mathbf{B}} = \mathbf{B}\boldsymbol{\Omega}^{-1}$ in Section A.1. Edges between predictors and responses are marginal regression coefficients, while edges between responses are covariances. Edges between responses and predictors generally agree within two cases, given that the partial correlation between responses are mostly positive (Figure 6). However,

marginally, the connection between responses and predictors is very dense. This is because all responses (genus) are connected by their partial correlations so that as long as the predictor can conditionally influence one of the responses, it should be able to marginally affect all the responses. We observe that some edges flip color when comparing the conditional and marginal networks. For example, the link between diet group 2, which corresponds to both complex (wholegrain breakfast cereals and breads, boiled potatoes) and simple carbohydrates (white bread), and *Veillonella* is blue (negative) in the conditional network (Figure 5) and red (positive) in the marginal network (Figure 6). *Veillonella* is well known for its lactate fermenting abilities; thus, a negative link with a carbohydrate diet is reasonable. Edges that flip color can be explained by interactions with other genera. Another example is that, marginally, all links are blue (negative) from the diet group percutaneous endoscopic gastrostomy (PEG)-fed subjects to the genus, whereas, conditionally there is a positive (red) link with *Parabacteroides*. These observations further reiterate that we should distinguish marginal and conditional effects of predictors in these research scenarios, especially when sparsity is assumed, since it is usually impossible to have both sparse marginal and conditional effects.

Last, we also compare our conditional network (Figure 5) with the conditional network obtained by fitting a multiresponse regression, and then transform the parameters into the conditional chain graph representation using the equivalence $\tilde{\mathbf{B}} = \mathbf{B}\boldsymbol{\Omega}$ in Section A.1 (Figure A16). This last network (transformed conditional network) is extremely dense, even though biologically, we expect the network to be sparse (hence the sparsity imposed by the chain graph model). This result shows the importance of having a proper sparsity prior, which is what our adaptive CG-LASSO has.

Soil microbiome compositional data. Figure 7 shows the soil microbiome estimated network using the adaptive CG-LASSO model. In this network, the most important link is between *Candidatus Solibacter* and *Candidatus Koribacter*. In this case, there are not important connections with predictors, which seems to suggest that the soil microbial community is robust to environmental perturbations. These results agree with the original research [28, 29] that indicated that core microbial communities within soil aggregates are likely driven by stable and long-term factors such as clay content rather than relative short time scaled land management, such as the ones considered as predictors in this study. We note that the original research concentrated on the diversity of the community, while our analysis focuses on the structure and correlations within the community.

Comparison with a marginal network. Again, we estimate the marginal network (Figure 8), starting with the conditional network, thereby using the equivalence $\tilde{\mathbf{B}} = \mathbf{B}\boldsymbol{\Omega}^{-1}$ in Section A.1. Edges between predictors and responses are marginal regression coefficients while edges between responses are covariances. In the estimated network (Figure 8), we see a much stronger marginal effect of environmental predictors than the one we saw with conditional effects (Figure 7). The LASSO penalty might contribute to this behavior, but it might also be because the observed strong dependence between microbes and the environment is due to the strong partial correlation among microbes. In addition, marginally, crop agriculture and total nitrogen (total N) have strong correlations to multiple genera. However, this pattern is not obvious in the conditional network (nor on the original research). It might suggest that the influence of crop agriculture and the total N might be enhanced by a strong partial correlation among microbes. The largest partial regression coefficient for the total N is to genus *Nitrososphaera*, which is indeed a N-fixing genus that has the highest $\alpha-$centrality (Figure 9).

Last, we also compare our conditional network (Figure 7) with the conditional network obtained by fitting a multiresponse regression, and then transform the parameters into the conditional chain

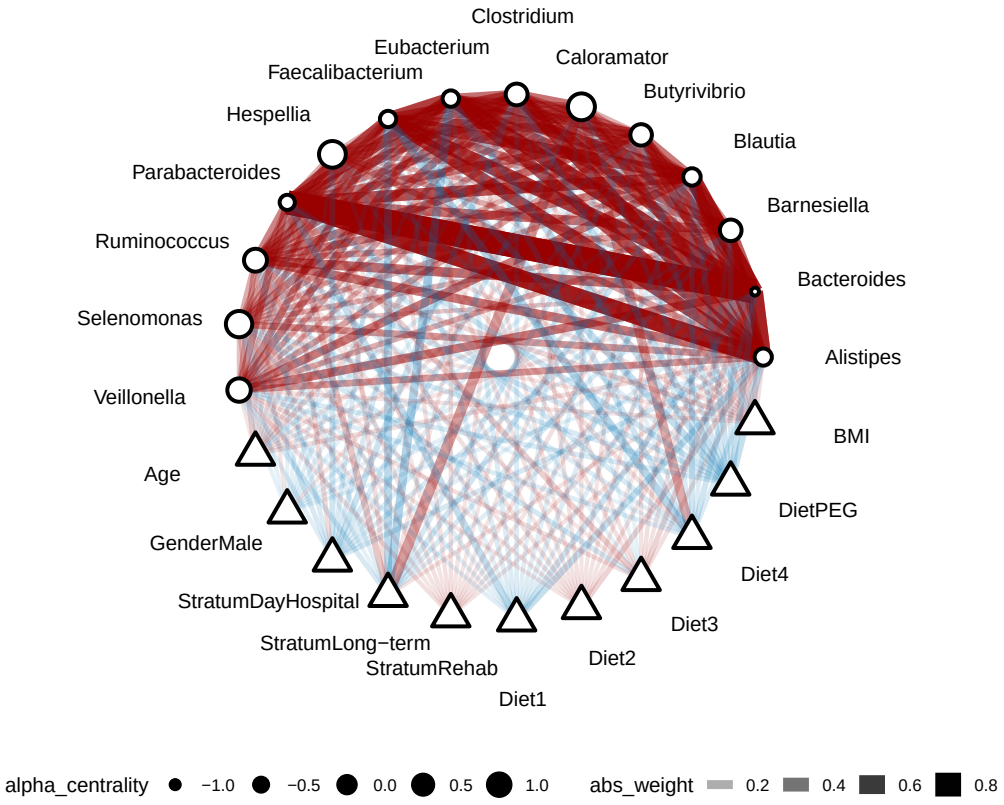

Figure 6. Reconstructed genus marginal network for human gut microbiome using multiresponse regression. Triangle nodes correspond to predictors and circle nodes correspond to relative abundances of genus. The node size on the circle nodes correspond to the $\alpha-$centrality values [27]. The width of the edges correspond to the absolute weight, and the color to the type of interaction (red positive, blue negative). Edges are marginal rather than conditional. Note that the larger size of the triangles compared to the triangles in the conditional network (Figure 5) are an effect of the shrinkage of the circle nodes in this network.

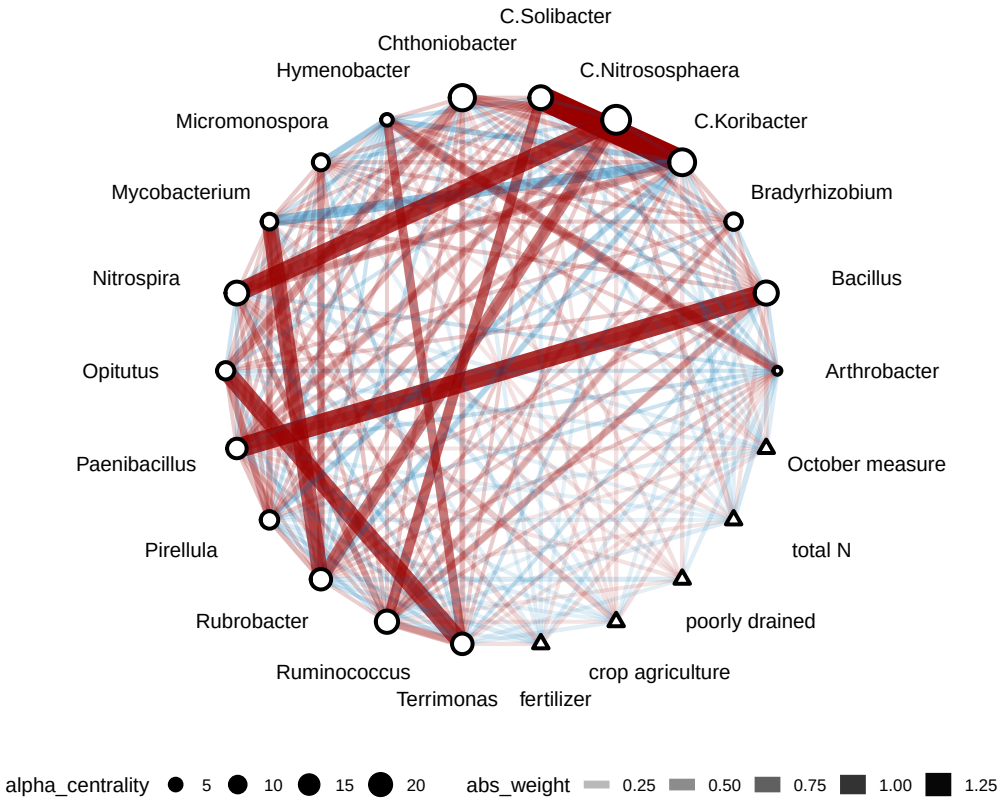

Figure 7. Reconstructed genus conditional network for soil microbiome using adaptive CG-LASSO. Triangle nodes correspond to predictors and circle nodes correspond to relative abundances of genus. The node size on the circle nodes corresponds to the $\alpha-$centrality values [27]. The width of the edges corresponds to the absolute weight and the color corresponds to the type of interaction (red positive, blue negative). Weak links with the environment (triangle nodes) agree with the original research [28, 29] that showed stable microbial community to environmental perturbations.

graph representation using the equivalence $\tilde{\mathbf{B}} = \mathbf{B}\boldsymbol{\Omega}$ in Section A.1 (Figure A17). This last network (transformed conditional network) is extremely dense, even though biologically, we expect the network to be sparse (hence the sparsity imposed by the chain graph model). We expect microbiome networks to be sparse because each microbial unit typically directly interacts with only a small subset of the community, and many taxa are rare or undetected in any given sample. Similar to the human gut data, this result shows the importance of having a proper sparsity prior.

To conclude, both the marginal network (Figure 8) and the conditional network transformed from marginal coefficients (Figure A17) disagree with the original research that core microbial communities within soil aggregates are likely driven by stable and long-term factors instead of the predictors measured in the data. Unlike marginal effects, conditional effects of the environment (Figure 7) can be more informative to biologists who would like to conduct research in understanding the environment's effect on certain microbes, for instance, the effect of environmental antibiotics.

It is worth highlighting that our model can produce biologically relevant results from relatively small sample sizes: 120 samples for the soil microbiome study and 178 samples for the human gut microbiome study.

3.3. Comparison of $\alpha-$centrality and abundances in human and soil communities

We evaluate the $\alpha-$centrality based on the estimated network for both datasets (human gut and soil) to identify keystone genus. Within Figure 9, the y-axis represents the ranking of genera based on the point estimation of the grand mean (μ) (i.e., the log relative abundances), that is, the genus on the top corresponds to the most abundant microbes (*Clostridium* for human gut and *Terrimonas* for soil); the x-axis shows the estimated $\alpha-$centrality for each genus. For the human gut data, the genus with the highest $\alpha-$centrality is *Bacteroides*, which is an abundant genus and known to have the ability to moderate the host's immune response [30] and transfer antibiotic genes to other members of the community [31]. For the soil data, the genus with highest $\alpha-$centrality is *Nitrososphaera*. Members of this genus have the ability to perform ammonia oxidizing, which might play a major role in nitrification [32], which is crucial in the soil microbial community. We observe a general trend that abundant genus have a higher $\alpha-$centrality, but this trend is not definite. For instance, neither *Bacteroides* nor *Nitrososphaera* have the highest abundance. The Spearman correlation coefficients between the estimated relative abundance and $\alpha-$centrality in the human gut dataset is 0.68 while it is 0.58 in the soil dataset. Moreover, the variation of the $\alpha-$centrality measure is larger in the soil (standard deviation is 0.67 times the mean in the soil data, while the standard deviation is 0.38 times the mean in the human gut data). The difference might due to the more variate environmental conditions in soil, which makes it difficult to have just one genus taking the central role.

4. Discussion

4.1. Flexibility of the Bayesian model

Compared with the frequentist method, the Bayesian method allows for an easier extension of the core Normal model to different types of responses via hierarchical structures. As long as one can sample from the full conditional distribution of the (latent) normal variable, the posterior sampling is a straight-forward extension of the proposed Gibbs sampler. Though not shown here, other commonly

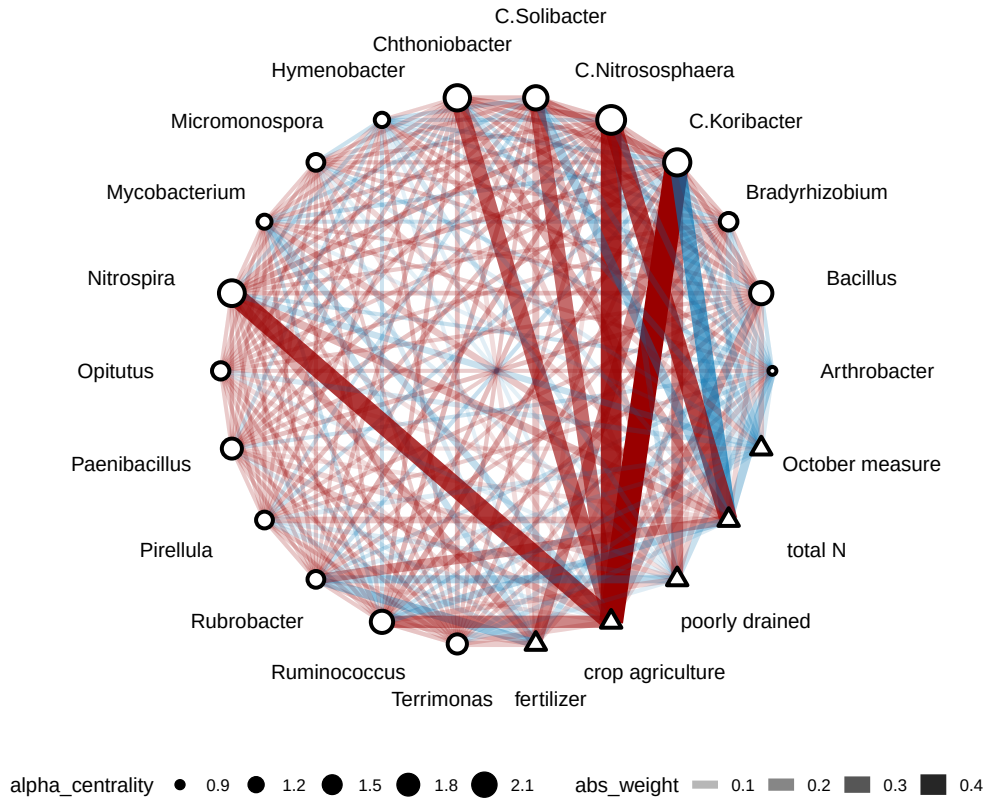

Figure 8. Reconstructed genus marginal network for soil microbiome using multiresponse regression. Triangle nodes correspond to predictors and circle nodes correspond to relative abundances of genus. The node size on the circle nodes correspond to the $\alpha-$centrality values [27]. The width of the edges corresponds to the absolute weight and the color corresponds to the type of interaction (red positive, blue negative). Edges are marginal rather than conditional. Strong links with environment (triangle nodes) disagree with the original research [28, 29] which showed that the microbial community should be stable to environmental perturbations.

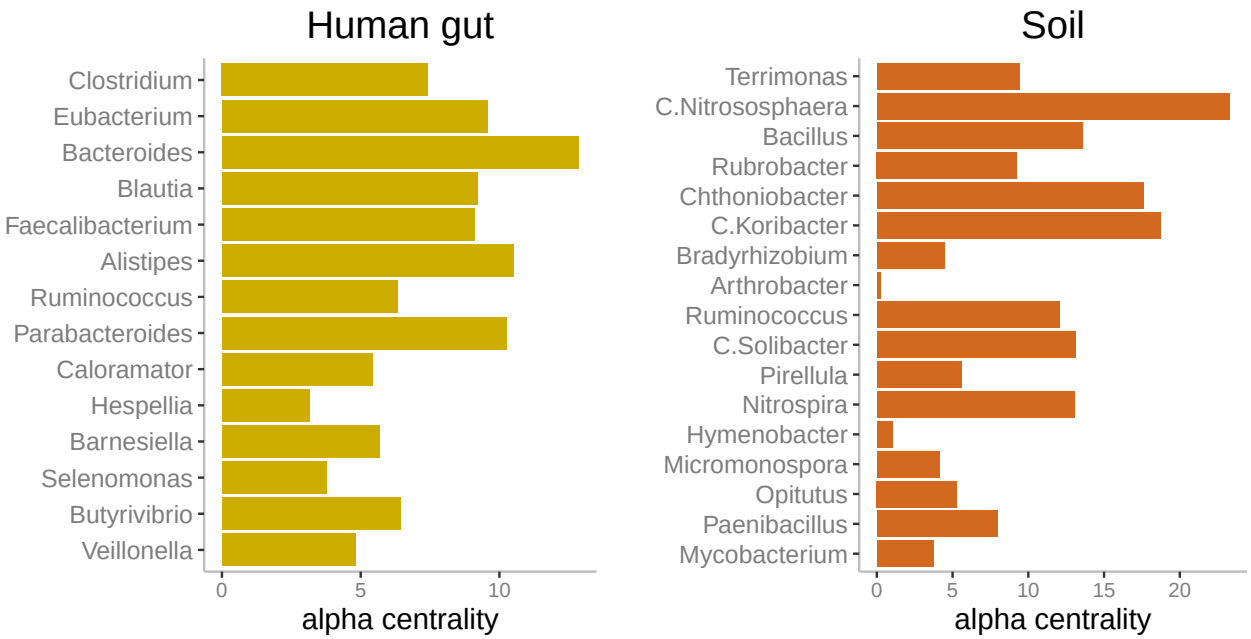

Figure 9. $\alpha-$centrality of genera ranked by abundances. We rank the genera based on estimated relative abundance (upper = higher abundance), where the bars correspond to the estimated $\alpha-$centrality. We see a general trend of abundant genera having a higher $\alpha-$centrality, but it is not definite.

encountered models in biology are also simple extensions of our model (e.g., zero-inflated Poisson and multinomial [33]). By using the Normal distribution as the core model, we can automatically take the over-dispersion into account because the model explicitly considers the variance parameters. Note that one common complaint on the LASSO prior is that it does not put any mass on 0 for any edge. Though a spike-and-slab prior is possible, an efficient posterior sampling algorithm such as the block Gibbs sampler presented in this work (also in [12]) would be hard to derive due to the intractable normalizing constant.

4.2. Challenges in learning the graphical structure

Graphical selection can be difficult because of the confounding nature of its own structure. For example, recall Figure 1(A),(B). These two graphs can produce a similar correlation between Y_1 and Y_2. One extreme example is when all links in Figure 1(A),(B) have no noise (e.g., $Y_1 = X$, $Y_2 = -Y_1$ versus $Y_1 = X$, $Y_2 = -X$). In this extreme example, it is impossible to distinguish graph A from B. Additionally, cases such as Figure 1E are particularly difficult, where all partial correlations are positive (or negative). Additionally, when $\mathbf{\Omega}$ has bad condition numbers, then $\mathbf{B}$ might have a large error in the estimation, since the marginal mean response and $\mathbf{\Omega}$ inform the estimation of $\mathbf{B}$, and a small change in the marginal mean response can have a large influence in $\mathbf{B}$. Future work could focus on how to use experiments to decouple the confounding of $\mathbf{\Omega}$ and $\mathbf{B}$ to address some of these challenges [34]. Other challenges to learn the graphical structure involve different non-Gaussian settings such as zero-inflated or compositional models. In the Appendix A.4, we cover the extension of our Gaussian model to binary, count, and multinomial (compositional) data. However, future work is needed to identify the

data-specific inference limitations such as identifiability under compositional constraints, disentangling structural from technical zeros, and the need for tailored regularization and benchmarking.

4.3. Agreement with experimenter's intuition on mean behavior

Intuitively, an experimenter should be able to make inferences about the interactions among responses from the behavior of the mean structures under treatment. For example, in Figure 1(D), an experimenter might knock out a gene as the treatment ($X = 1$ for knock out and $X = 0$ for not) and compare the gene expression levels of another gene (Y_2) via a t test. The result of this t test will provide information regarding the interaction between Y_1 and Y_2 because there are no other factors that affects Y_1, and Y_2 is conditionally independent with X. Thus, this experiment is specific to Y_1 and provides information on partial correlation between Y_1 and Y_2 by only affecting Y_1, that is, any change in Y_2 is due to the partial correlation with Y_1 rather than a reaction to X. It is precisely the fact that the mean of Y_2 in this experiment depends on the correlation between Y_1 and Y_2 that allows experimenters to test differences in means of Y_2 under the effect of the treatment (X) through standard t tests. However, this intuition is violated under the standard linear regression setting. The vector (Y_1, Y_2) is Normally distributed with mean $\mu = (X\beta_1, 0)$ and covariance $\mathbf{\Sigma}$ under the network in Figure 1(D); thus, the mean of Y_2 is always 0 regardless of the value of X. In contrast, in the CG parametrization, the mean vector is $\mathbf{\Sigma}\mu$ whose second entry is given by $\rho\beta_1 X$ (i.e., the mean value of Y_2 depends on β_1 (the reaction of Y_1 to the treatment) and ρ (the correlation between Y_1 and Y_2)). Given that the experimenter's intuition on specificity is based on the notion of *conditional (in)dependence* between X and Y_1, Y_2, we conclude that it is desirable that the mean vector contains information on the correlation structure among responses; this is a characteristic of the CG model that we propose, unlike other models that tend to treat the correlation structure ($\mathbf{\Omega}$) as a nuisance parameter.

4.4. Optimal model-based design of experiments

An experimenter should be able to design experiments that decode the links among response nodes when specific experimental interventions towards one node are possible. In practice, when possible, experimenters will always prefer experiments with a better specificity. However, this preference is not evident in the linear regression setting since the Fisher information matrix of the mean vector and the precision matrix is block-diagonal [35]; thus, any information that we have on $\mathbf{B}$ will not affect the estimation of $\mathbf{\Sigma}$. In addition, the information of $\mathbf{\Sigma}$ is not a function of design ($\mathbf{X}$), no matter whether we have prior knowledge about the effect of such an experiment (prior on $\mathbf{B}$). The CG parametrization avoids this disagreement because the Fisher information matrix is no longer block-diagonal and prior information about the treatment can flow into the estimation of $\mathbf{\Sigma}$ via an optimal model-based experimental design [36]. We highlight that due to the confounding between the treatment effect and the interaction among responses, the prior knowledge on the specificity of the treatment is necessary for such an optimal model-based experimental design. Future work could investigate the method of experimental design to best decode such networks and the theoretical properties of such designs [34].

5. Conclusions

- We introduced a novel Bayesian chain graph model to infer a sparse network structure with nodes for responses (microbes) and predictors, where the edges all represent conditional

dependence, which is more biologically interpretable than marginal edges commonly used in microbiome network research.
- We showed the applicability of Bayesian chain graphs in microbiome research via extensive simulations and two real datasets: human and soil microbiomes.
- We implemented our novel theory in an R package `CARlasso`, which is publicly available on GitHub: https://github.com/YunyiShen/CAR-LASSO

Open-source software

We developed our algorithm in R 3.6.3 [37], and all the code and data used is available as an R package `CARlasso` hosted on https://github.com/YunyiShen/CAR-LASSO. All simulations and data analyses code are in the `dev` branch.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the National Institute of Food and Agriculture, United States Department of Agriculture, Hatch projects 1023699 and 7007384. This work was also supported by the Department of Energy [DE-SC0021016 to C.S.L.]. Y.S. would like to thank Xiang Li from the University of Hongkong for discussion on the generalized inverse Gaussian distribution.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. N. Fierer, C. L. Lauber, K. S. Ramirez, J. Zaneveld, M. A. Bradford, R. Knight, Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients, *ISME J.*, **6** (2012), 1007–1017. https://doi.org/10.1038/ismej.2011.159
2. T. Whitman, R. Neurath, A. Perera, I. Chu-Jacoby, D. Ning, J. Zhou, et al., Microbial community assembly differs across minerals in a rhizosphere microcosm, *Environ. Microbiol.*, **20** (2018), 4444–4460. https://doi.org/10.1111/1462-2920.14366
3. A. M. Cates, M. J. Braus, T. L. Whitman, R. D. Jackson, Separate drivers for microbial carbon mineralization and physical protection of carbon, *Soil Biol. Biochem.*, **133** (2019), 72–82. https://doi.org/10.1016/j.soilbio.2019.02.014
4. C. Kranz, T. Whitman, Short communication: Surface charring from prescribed burning has minimal effects on soil bacterial community composition two weeks post-fire in jack pine barrens, *Appl. Soil Ecol.*, **144** (2019), 134–138. https://doi.org/10.1016/j.apsoil.2019.07.004

5. T. Whitman, E. Whitman, J. Woolet, M. D. Flannigan, D. K. Thompson, M. Parisien, Soil bacterial and fungal response to wildfires in the canadian boreal forest across a burn severity gradient, *Soil Biol. Biochem.*, **138** (2019), 107571. https://doi.org/10.1016/j.soilbio.2019.107571
6. C. Allsup, R. Lankau, Migration of soil microbes may promote tree seedling tolerance to drying conditions, *Ecology*, **100** (2019), e02729. https://doi.org/10.1002/ecy.2729
7. R. A. Rioux, C. M. Stephens, J. P. Kerns, Factors affecting pathogenicity of the turfgrass dollar spot pathogen in natural and model hosts, preprint, bioRxiv. https://doi.org/10.1101/630582
8. P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, J. I. Gordon, The human microbiome project, *Nature*, **449** (2007), 804–810. https://doi.org/10.1038/nature06244
9. M. Dave, P. D. Higgins, S. Middha, K. P. Rioux, The human gut microbiome: Current knowledge, challenges, and future directions, *Transl. Res.*, **160** (2012), 246–257. https://doi.org/10.1016/j.trsl.2012.05.003
10. M. Layeghifard, D. M. Hwang, D. S. Guttman, Disentangling interactions in the microbiome: A network perspective, *Trends Microbiol.*, **25** (2017), 217–228. https://doi.org/10.1016/j.tim.2016.11.008
11. J. Friedman, T. Hastie, R. Tibshirani, Sparse inverse covariance estimation with the graphical lasso, *Biostatistics*, **9** (2008), 432–441. https://doi.org/10.1093/biostatistics/kxm045
12. H. Wang, Bayesian graphical lasso models and eficient posterior computation, *Bayesian Anal.*, **7** (2012), 867–886. https://doi.org/10.1214/12-BA729
13. C. Lo, R. Marculescu, Mplasso: Inferring microbial association networks using prior microbial knowledge, *PLoS Comput. Biol.*, **13** (2017), e1005915. https://doi.org/10.1371/journal.pcbi.1005915
14. S. L. Lauritzen, N. Wermuth, Graphical models for associations between variables, some of which are qualitative and some quantitative, *Ann. Statist.*, **17** (1989), 31–57. https://doi.org/10.1214/aos/1176347003
15. S. L. Lauritzen, T. S. Richardson, Chain graph models and their causal interpretations, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **64** (2002), 321–348. https://doi.org/10.1111/1467-9868.00340
16. S. A. Andersson, D. Madigan, M. D. Perlman, Alternative markov properties for chain graphs, *Scand. J. Stat.*, **28** (2001), 33–85. https://doi.org/10.1111/1467-9469.00224
17. M. Frydenberg, The chain graph markov property, *Scand. J. Stat.*, **17** (1990), 333–353.
18. M. Chen, Z. Ren, H. Zhao, H. Zhou, Asymptotically normal and efficient estimation of covariate-adjusted gaussian graphical model, *J. Am. Stat. Assoc.*, **111** (2016), 394–406. https://doi.org/10.1080/01621459.2015.1010039
19. J. Zhang, Y. Li, High-dimensional gaussian graphical regression models with covariates, *J. Am. Stat. Assoc.*, (2022), 2088–2100. https://doi.org/10.1080/01621459.2022.2034632
20. C. Leng, M. N. Tran, D. Nott, Bayesian adaptive Lasso, *Ann. Inst. Stat. Math.*, **66** (2014), 221–244. https://doi.org/10.1007/s10463-013-0429-6
21. J. Fan, Y. Feng, Y. Wu, Network exploration via the adaptive lasso and scad penalties, *Ann. Appl. Stat.*, **3** (2009), 521–541. https://doi.org/10.1214/08-AOAS215

22. S. Xie, D. Zeng, Y. Wang, Integrative network learning for multimodality biomarker data, *Ann. Appl. Stat.*, **15** (2021), 64–87. https://doi.org/10.1214/20-AOAS1382

23. M. J. Claesson, I. B. Jeffery, S. Conde, S. E. Power, E. M. O'connor, S. Cusack, et al., Gut microbiota composition correlates with diet and health in the elderly, *Nature*, **488** (2012), 178–184. https://doi.org/10.1038/nature11319

24. P. W. O'Toole, Gut microbiota in the irish elderly and its links to health and diet (mgp154), 2008. Available: https://www.mg-rast.org/mgmain.html?mgpage=project&project=mgp154.

25. F. G. Blanchet, K. Cazelles, D. Gravel, Co-occurrence is not evidence of ecological interactions, *Ecol. Lett.*, **23** (2020), 1050–1063. https://doi.org/10.1111/ele.13525

26. F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E. M. Glass, M. Kubal, et al., The metagenomics rast server–A public resource for the automatic phylogenetic and functional analysis of metagenomes, *BMC Bioinf.*, **9** (2008), 386. https://doi.org/10.1186/1471-2105-9-386

27. P. Bonacich, P. Lloyd, Eigenvector-like measures of centrality for asymmetric relations, *Soc. Networks*, **23** (2001), 191–201. https://doi.org/10.1016/S0378-8733(01)00038-7

28. K. Hofmockel, Hofmockel soil aggregate cob kbase (mgp2592), 2012. Available from: https://www.mg-rast.org/mgmain.html?mgpage=project&project=mgp2592.

29. E. M. Bach, R. J. Williams, S. K.Hargreaves, F. Yang, K. S. Hofmockel, Greatest soil microbial diversity found in micro-habitats, *Soil Biol. Biochem.*, **118** (2018), 217–226. https://doi.org/10.1016/j.soilbio.2017.12.018

30. J. M. Blander, R. S. Longman, I. D. Iliev, G. F. Sonnenberg, D. Artis, Regulation of inflammation by microbiota interactions with the host. *Nat. Immunol.*, **18** (2017), 851–860. https://doi.org/10.1038/ni.3780

31. N. Shoemaker, H. Vlamakis, K. Hayes, A. Salyers, Evidence for extensive resistance gene transfer amongbacteroides spp, and among bacteroides and other genera in the human colon, *Appl. Environ. Microbiol.*, **67** (2001), 561–568. https://doi.org/10.1128/AEM.67.2.561-568.2001

32. M. Tourna, M. Stieglmeier, A. Spang, M. Könneke, A. Schintlmeister, T. Urich, et al., Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil, *Proc. Natl. Acad. Sci.*, **108** (2011), 8420–8425. https://doi.org/10.1073/pnas.1013488108

33. D. Lambert, Zero-inflated poisson regression, with an application to defects in manufacturing, *Technometrics*, **34** (1992), 1–14. https://doi.org/10.2307/1269547

34. Y. Shen, C. Solís-Lemus, The effect of the prior and the experimental design on the inference of the precision matrix in gaussian chain graph models, *J. Agric. Biol. Environ. Stat.*, **30** (2025), 800–869. https://doi.org/10.1007/s13253-024-00621-1

35. L. Malagò, G. Pistone, Information geometry of the gaussian distribution in view of stochastic optimization, in *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII*, (2015), 150–162. https://doi.org/10.1145/2725494.2725510

36. K. Chaloner, I. Verdinelli, Bayesian experimental design: A review, *Stat. Sci.*, **10** (1995), 273–304.

37. R Core Team, R: A language and environment for statistical computing, 2020. Available from: https://cran.rstudio.com/manuals.html.

38. T. Park, G. Casella, The Bayesian lasso, *J. Am. Stat. Assoc.*, **103** (2008), 681–686. https://doi.org/10.1198/016214508000000337

39. D. F. Andrews, C. L. Mallows, Scale mixtures of normal distributions, *J. R. Stat. Soc Series B*, **36** (1974), 99–102. https://doi.org/10.1111/j.2517-6161.1974.tb00989.x

40. C. M. Carvalho, N. G. Polson, J. G. Scott, The horseshoe estimator for sparse signals, *Biometrika*, **97** (2010), 465–480. https://doi.org/10.1093/biomet/asq017

41. B. Jones, C. Carvalho, A. Dobra, C. Hans, C. Carter, M. West, Experiments in stochastic computation for high-dimensional graphical models, *Stat. Sci.*, **20** (2005), 388–400. https://doi.org/10.1214/088342305000000304

42. J. Aitchison, C. H. Ho, The multivariate Poisson-log normal distribution, *Biometrika*, **76** (1989), 643–653. https://doi.org/10.1093/biomet/76.4.643

43. W. R. Gilks, P. Wild, Adaptive rejection sampling for gibbs sampling, *J. R. Stat. Soc. Ser. C*, **41** (1992), 337–348. https://doi.org/10.2307/2347565

44. F. Xia, J. Chen, W. K. Fung, H. Li, A logistic normal multinomial regression model for microbiome compositional data analysis, *Biometrics*, **69** (2013), 1053–1063. https://doi.org/10.1111/biom.12079

45. W. Hörmann, J. Leydold, Generating generalized inverse gaussian random variates, *Stat. Comput.*, **24** (2014), 547–557. https://doi.org/10.1007/s11222-013-9387-3

46. B. Jorgensen, *Statistical Properties of the Generalized Inverse Gaussian Distribution*, Springer, 1982.

Appendix

A. Full model specification and interpretation

A.1. Model specification

Let $\mathbf{Y}_i \in \mathbb{R}^k$ be a multivariate response with k entries for $i = 1, \ldots, n$ observations. Let $\mathbf{X}_i \in \mathbb{R}^{1\times p}$ be the row vector of predictors for $i = 1, \ldots, n$ (i.e., the i^{th} row of the design matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$). We assume that the design matrix is standardized so that each column has a mean of 0 and the same standard deviation (set to be 1 in the simulations); therefore, the same shrinkage parameter will not have different effects on different predictors.

Let $\mathbf{Y}_i$ follow a Normal distribution with mean vector $\mathbf{\Omega}^{-1}(\mathbf{B}^T\mathbf{X}_i^T + \mu)$ and precision matrix $\mathbf{\Omega} \in \mathbb{R}^{k\times k}$ (positive definite) where $\mathbf{B} \in \mathbb{R}^{p\times k}$ corresponds to the regression coefficients that connects the responses ($\mathbf{Y}_i \in \mathbb{R}^k$) and the predictors ($\mathbf{X}_i \in \mathbb{R}^{1\times p}$) and $\mu \in \mathbb{R}^k$ corresponds to the intercept. We use the transpose $\mathbf{B}^T\mathbf{X}_i^T \in \mathbb{R}^{k\times 1}$ because samples are encoded as row vectors in the design matrix while multivariate Normal samples are column vectors by convention. While the inclusion of $\mathbf{\Omega}$ in the mean function is unusual in statistical settings, we argue that it is necessary to draw the right biological interpretations (see the Discussion).

The likelihood function of the model is as follows:

$$p(\mathbf{Y}_i|\mathbf{X}_i, \mu, \mathbf{B}, \mathbf{\Omega}) \propto \exp[(\mathbf{B}^T\mathbf{X}_i^T + \mu)^T\mathbf{Y}_i - \frac{1}{2}\mathbf{Y}_i^T\mathbf{\Omega}\mathbf{Y}_i]. \tag{A.1}$$

Note that in this parametrization, $\mathbf{B}$ encodes *conditional dependence* between $\mathbf{Y}$ and $\mathbf{X}$ because B_{jq} is the coefficient of product between X_j and Y_q in the kernel of the density. Thus, if $B_{jq} = 0$, then X_j and Y_q are *conditionally independent*. This is analogous to the case of $\mathbf{\Omega}$ whose off-diagonal entries encode the conditional dependence between responses Y_q and $Y_{q'}$. To see the analogy, we provide an interpretation of the parameters in the same manner as in univariate linear regression. Let $\mathbf{Y}_{-q}$ be the vector of responses without the qth component, let $\mathbf{Y}_{-(q',q)}$ be the vector of responses without the qth and q'th components, and let $\mathbf{X}_{-j}$ be the vector of predictors without the jth component.

$$\begin{aligned}\mathbb{E}[Y_q|X_j = x_j + 1, \mathbf{Y}_{-q}, \mathbf{X}_{-j}, \mathbf{B}, \mathbf{\Omega}] - \mathbb{E}[Y_q|X_j = x_j, \mathbf{Y}_{-q}, \mathbf{X}_{-j}, \mathbf{B}, \mathbf{\Omega}] = B_{jq}/(\omega_{qq}),\\ \mathbb{E}[Y_q|Y_{q'} = y_{q'} + 1, \mathbf{Y}_{-(q',q)}, \mathbf{X}, \mathbf{B}, \mathbf{\Omega}] - \mathbb{E}[Y_q||Y_{q'} = y_{q'}, \mathbf{Y}_{-(q,q')}, \mathbf{X}, \mathbf{B}, \mathbf{\Omega}] = \omega_{qq'}/(-\omega_{qq}).\end{aligned} \tag{A.2}$$

By fixing the values of all but one predictor and all of the other responses, an increase of one unit in X_j is associated with $B_{jq}/(\omega_{qq})$ unit increase in the expectation of Y_q, hence conditioned on the values of all other responses $Y_{q'}$.

More details about these equations are explained next. We can observe the kernel of a chain graph model as follows:

$$\begin{aligned}P(\mathbf{Y}|\mathbf{X}, \mathbf{B}, \mathbf{\Omega}) &\propto \exp\left\{-\frac{1}{2}\mathrm{tr}\left[(\mathbf{Y} - \mathbf{X}\mathbf{B}\mathbf{\Omega}^{-1})\mathbf{\Omega}((\mathbf{Y} - \mathbf{X}\mathbf{B}\mathbf{\Omega}^{-1}))^T\right]\right\}\\ &\propto \exp\left\{-\frac{1}{2}\mathrm{tr}\left[-2\mathbf{Y}\mathbf{B}\mathbf{X}^T + \mathbf{Y}\mathbf{\Omega}\mathbf{Y}\right]\right\}\\ &= \exp\left\{-\frac{1}{2}\left[-2\sum_{j,q} B_{jq}Y_qX_j + \sum_{q,q'}\omega_{qq'}Y_qY_{q'}\right]\right\}.\end{aligned}$$

Thus, the conditional distribution of Y_q is given by the following:

$$\begin{aligned}P(Y_q|Y_{-q}, \mathbf{X}, \mathbf{B}, \mathbf{\Omega}) &\propto \exp\left\{-\frac{1}{2}\left[-2\sum_{j,q} B_{jq}Y_qX_j + \sum_{q,q'}\omega_{qq'}Y_qY_{q'}\right]\right\}\\ &\propto \exp\left\{-\frac{1}{2}\left[2Y_q\left[-\sum_j B_{jq}X_j + \sum_{q'\neq q}\omega_{qq'}Y_{q'}\right] + Y_q\omega_{qq}Y_q\right]\right\}.\end{aligned}$$

This is a normal distribution with mean $\frac{1}{\omega_{qq}}(\sum_j B_{jq}X_j - \sum_{q'\neq q}\omega_{qq'}Y_{q'})$ and variance $1/\omega_{qq}$. Observe that the mean is essentially a linear regression towards X_j and $Y_{q'}$ with regression coefficients B_{jq}/ω_{qq} and $-\omega_{qq'}/\omega_{qq}$, thus, we can interpret these parameters as usual linear regression coefficients, i.e., B_{jq}/ω_{qq} is the expected response change of Y_q when X_j changed by 1 unit while holding X_{-j} and Y_{-q} constant, and $-\omega_{qq'}/\omega_{qq}$ is the expected response change of Y_q when $Y_{q'}$ changed by 1 unit while holding $\mathbf{X}$ and other Y's constant.

On the other hand, the regression coefficients in multiresponse linear regression, denoted as $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{\Omega}$, are marginal effects. That is,

$$\mathbb{E}[Y_q|X_j = x_j + 1, \mathbf{X}_{-j}, \tilde{\mathbf{B}}, \mathbf{\Omega}] - \mathbb{E}[Y_q|X_j = x_j, \mathbf{X}_{-j}, \tilde{\mathbf{B}}, \mathbf{\Omega}] = \tilde{B}_{jq}. \tag{A.3}$$

Note that the crucial difference here is that we do not condition on $\mathbf{Y}_{-q}$. Specifically in microbiome, the presence of $\mathbf{\Omega}^{-1}$ in the mean represents the ecological knowledge that the responses of a species (e.g., relative abundances of microbes) depend on both its reaction to the environment ($\mathbf{B}$) and interactions with other species ($\mathbf{\Omega}$). In our model, the regression coefficients matrix $\mathbf{B}$ encode conditional dependence among the responses (scaled by variance) and the predictors that arguably has a more mechanistic interpretation [14, 16, 17].

In general, it is not possible to simultaneously find sparse marginal predictions of single nodes and sparse graph. That is, the marginal effect $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{\Omega}^{-1}$ typically has different support to the direct effect $\mathbf{B}$. It is possible for both parameters $\mathbf{B}$ and $\tilde{\mathbf{B}}$ to be sparse when $\mathbf{\Omega}^{-1}$ is diagonal. In this case, the responses are independent, and thus, $\mathbf{B}_{jq} = 0$ implies $\tilde{\mathbf{B}}_{jq} = 0$ for any $\mathbf{B}$.

While our model focuses on finding the sparse graph, we also implement a model for the sparse marginal regression coefficient and the sparse precision matrix by combining the Gibbs sampling for the precision matrix in [12] with the Gibbs sampler in [38]. We denote this model Simultaneous Regression and Graphical LASSO (SRG-LASSO) and we use it to compare to the CG-LASSO in the simulation study (Section 2.2).

A.2. Prior specification

We assume a Laplace prior on the entries of $\mathbf{B}$ and graphical LASSO prior on $\mathbf{\Omega}$ [12, 38]. That is,

$$\begin{aligned} \mathbf{Y}_i|\mathbf{X}_i, \mu, \mathbf{B}, \mathbf{\Omega} &\sim N(\mathbf{\Omega}^{-1}(\mathbf{B}^T\mathbf{X}_i^T + \mu), \mathbf{\Omega}^{-1}), \\ B_{jq}|\tau_{jq} &\sim \text{Laplace}(\lambda_\beta^2), \\ \Omega_{q,q'} &\sim \text{Laplace}(\lambda_\Omega^2) \quad \text{for} \quad q < q', \\ \Omega_{q,q} &\sim \text{Exp}(\lambda_\Omega) \end{aligned} \tag{A.4}$$

with the caveat that $\mathbf{\Omega}$ is positive semi-definite. This specification is not easy to sample because of the Laplace prior and the truncation on positive semi-definite cone. We approach the first problem using the Normal scale mixture representation of Laplace distribution [12, 38, 39]. Let η_{ml} be the latent scale parameters for $\mathbf{\Omega}$ for $1 \le q < q' \le k$ since $\mathbf{\Omega}$ is symmetric, and let τ_{jq} $(1 \le j \le p, 1 \le q \le k)$ be the latent scale parameters for $\mathbf{B}$. For the second problem, we follow the method proposed by [12].

Then, the full model specification is as follows:

$$\begin{aligned} \mathbf{Y}_i|\mathbf{X}_i, \mu, \mathbf{B}, \mathbf{\Omega} &\sim N(\mathbf{\Omega}^{-1}(\mathbf{B}^T\mathbf{X}_i^T + \mu), \mathbf{\Omega}^{-1}), \\ B_{jq}|\tau_{jq}, &\sim N(0, \tau_{jq}^2), \tau_{jq} \sim \frac{\lambda_\beta^2}{2} e^{-\lambda_\beta^2 \tau_{jq}}, \\ p(\mathbf{\Omega}|\boldsymbol{\eta}, \lambda_\Omega) &= C_\eta^{-1} \prod_{q<q'} \left[\frac{1}{\sqrt{2\pi\eta_{qq'}}} \exp\left(-\frac{\omega_{qq'}^2}{2\eta_{qq'}} \right) \right] \prod_{q=1}^{q} \left[\frac{\lambda_\Omega}{2} \exp\left(-\frac{\lambda_\Omega \omega_{qq}}{2} \right) \right] I_{\Omega \in M^+}, \\ p(\boldsymbol{\eta}|\lambda_\Omega) &\propto C_\eta \prod_{q<q'} \frac{\lambda_\Omega^2}{2} \exp\left(-\frac{\lambda_\Omega^2 \eta_{qq'}}{2} \right) \end{aligned} \tag{A.5}$$

where $I_{\Omega \in M^+}$ means that $\mathbf{\Omega}$ must be positive definite.

A.3. Implementation

A.3.1. Sampling scheme

We derive an efficient Gibbs sampler for all parameters in this model due to the scale mixture representation of the graphical LASSO prior [12]. Details on derivation of the sampling scheme are summarized in Algorithm 1 with all the extensions (Section A.4).

A.3.2. Choice of hyperparameters

The shrinkage parameters λ_Ω and λ_β (Equation (A.5)) are hyperparameters to be determined. As in [12, 38], we assume that these shrinkage parameters have a hyperprior Gamma distribution with shape parameter r and rate parameter δ which can be set to produce a relatively flat density for a non-informative prior scenario. Note that since the prior on $\mathbf{\Omega}$ is not a Laplace but a graphical LASSO prior [12], the Gamma prior is on λ, not on λ^2 as it would be under a Laplace prior: $\lambda_\beta^2 \sim \text{Gamma}(r_\beta, \delta_\beta)$ and $\lambda_\Omega \sim \text{Gamma}(r_\Omega, \delta_\Omega)$.

The shrinkage parameters λ_Ω and λ_β are included in the Gibbs sampler with full conditional distribution still Gamma with shape parameters $r_\beta + kp, \delta_\beta + \sum \tau_i/2$ and rate parameters $r_\Omega + k(k+1)/2, \delta_\Omega + \|\mathbf{\Omega}\|_1/2$ respectively.

A.3.3. Learning the graphical structure

Given the continuous priors, our model has a zero posterior probability for a parameter to be zero. Yet, we still need to determine the cases when the edges of the graph will be considered non-existent. Here, we infer the graph structure using the horseshoe method in [12, 40] which compares the LASSO estimate for the regression coefficient with the posterior mean of a standard conjugate (non-shrinkage) prior [41].

Let $\pi = \frac{\tilde{\theta}}{E_g(\theta|\mathbf{Y})}$ where $\tilde{\theta}$ represents the estimate of the parameter under the LASSO prior and $E_g(\theta|\mathbf{Y})$ is the posterior mean of that parameter under non-shrinkage prior (e.g., Normal for $\mathbf{B}$ and Wishart for $\mathbf{\Omega}$). The statistics $1-\pi$ characterizes the amount of shrinkage due to the LASSO prior. We use $\pi > 0.5$ as the threshold to decide that $\theta \neq 0$ as in [12].

A.4. Extensions

A.4.1. Adaptive LASSO

One simple extension to LASSO is Adaptive LASSO, in which the shrinkage parameter λ can be different for all elements in $\mathbf{B}$ and $\mathbf{\Omega}$ [12, 20]. This extension is particularly useful when we have prior knowledge of independence among certain nodes. For example, larger shrinkage parameters (λ) on specific entries can be used to indicate prior knowledge of independence.

As suggested in [12, 20], we set the hyperpriors on $\lambda_{jq,\beta}^2$ as Gamma distributions with shape parameters $r_{jq,\beta}$ and rate parameter $\delta_{jq,\beta}$. Additionally, we set the prior suggested in [12] for $\lambda_{qq',\Omega}$ (with $q \neq q'$). While shrinkage on diagonal $\lambda_{qq,\Omega}$ is a hyperparameter in [12], here we set it to 0, that is, we are not shrinking the diagonal entries of $\mathbf{\Omega}$. However, such shrinkage can be included by multiplying the prior of $\mathbf{\Omega}$ by $\prod_{q=1}^{k} \frac{\lambda_{qq}}{2} exp(-\lambda_{qq,\Omega}\omega_{qq})$.

The prior for $\mathbf{\Omega}$ is

Algorithm 1: Implementation of Gibbs sampling

Result: MCMC samples of the posterior distribution of parameters of interest
Initialization;
while *not enough samples* **do**
  if *responses not Normal* **then**
    Update the latent Normal variables using adaptive rejection sampling (ARS) for counting and compositional data and truncated Normal for binary data (Section A.4.2);
  end
  for q^{th} *diagonal entries in* $\mathbf{\Omega}$ **do**
    /* blockwise update for $\mathbf{\Omega}$ */
    Sample the determinant of $\mathbf{\Omega}$: $\gamma|\mathbf{\Omega}_{[q]}, \eta, \lambda_\Omega \sim$ generalize inverse Gaussian (GIG) distribution (Equation (B.3)) where $\mathbf{\Omega}_{[q]}$ corresponds to some partition of $\mathbf{\Omega}$ based on the diagonal entry ω_{qq};
    Update the off diagonal entries in the qth row (column): $\boldsymbol{\omega}_{-qq}|\gamma, \mathbf{\Omega}_{[q]}, \eta, \lambda_\Omega \sim$ Normal distribution (Equation (B.4));
    Compute the updated qth diagonal entry (ω_{qq}) with Equation (B.2) that depends on the determinant (γ), the off diagonal entries ($\boldsymbol{\omega}_{-qq}$) and the partition of $\mathbf{\Omega}$ ($\mathbf{\Omega}_{[q]}$);
  end
  Update $\mathbf{B}|\boldsymbol{\tau}^2, \mathbf{\Omega}, \boldsymbol{\mu}, \mathbf{X}, \mathbf{Y} \sim$ Normal distribution (Equation (B.6));
  Update $\boldsymbol{\mu} \sim$ Normal$((\mathbf{Y\Omega} - \mathbf{XB})^T, \mathbf{\Omega}/n)$;
  Update latent variables in the scale mixture representation of the two LASSO priors: $\boldsymbol{\eta} = \{\eta_{jq}\}, \boldsymbol{\tau} = \{\tau_{jq}\}$ following an iInverse Gaussian distribution (Equation (B.5));
  if *adaptive shrinkage* **then**
    Update the shrinkage parameters on $\mathbf{\Omega}$ ($\lambda_{jq,\Omega}$) and on $\mathbf{B}$ ($\lambda^2_{jq,\beta}$) for individual entries (main text);
  else
    Update the shrinkage parameters on $\mathbf{\Omega}$ (λ_Ω) and on $\mathbf{B}$ (λ^2_β) uniformly for all entries following a Gamma distribution (main text);
  end
end

$$p(\mathbf{\Omega}|\{\lambda_{qq',\Omega}\}_{q<q'}) = C^{-1}_{\{\lambda_{qq',\Omega}\}_{q<q'}} \prod_{q<q'} \frac{\lambda_{qq',\Omega}}{2} \exp(-\lambda_{qq',\Omega}|\omega_{qq'}|) I_{\mathbf{\Omega}\in M_+}$$

$$p(\{\lambda_{qq',\Omega}\}_{q<q'}) \propto C_{\{\lambda_{qq',\Omega}\}_{q<q'}} \prod_{q<q'} \frac{1}{\Gamma(r_{qq',\Omega})} \lambda^{r_{qq',\Omega}-1}_{qq',\Omega} \exp(-\delta_{qq',\Omega}\lambda_{qq',\Omega}).$$

Then, the full conditional distribution of the shrinkage parameters is Gamma (shape and rate parametrization) are $\lambda_{qq',\Omega}|\mathbf{\Omega} \sim \text{Gamma}(r_{qq',\Omega} + 1, \delta_{qq',\Omega} + |\omega_{qq'}|)$ for $q \neq q'$, and $\lambda^2_{qq',\beta}|\tau \sim \text{Gamma}(r_{qq',\beta} + 1, \delta_{qq',\beta} + \tau_{qq'}/2)$. We set the hyperparameters as $r = 10^{-2}$ and $\delta = 10^{-6}$ for both $\mathbf{\Omega}$ and $\mathbf{B}$ [12, 20] with a small value of δ selected to take advantage of the adaptiveness of the shrinkage.

A.4.2. Extension to other types of responses

The model has been defined for continuous responses, yet there are different extensions for the case of binary data, counts and compositional data that we describe below.

- *Probit model for binary data.* For binary responses, we can use a Probit model with CG in the core of the dependence structure. We denote the CG latent variable as $\mathbf{Z}_i \in \mathbb{R}^k$, and let $\Phi(Z_{ij})$ model the probability of observing a 1 where Φ is the cumulative distribution function of a standard normal. Equation (A.6) shows the alternative representation of the model:

$$\begin{aligned} \mathbf{Z}_i &\sim N(\mathbf{\Omega}^{-1}(\mathbf{B}^T\mathbf{X}_i^T + \mu), \mathbf{\Omega}^{-1}), \\ Y^*_{ij} &\sim N(Z_{ij}, 1), \\ Y_{ij} &= \mathbf{1}_{Y^*_{ij}>0}. \end{aligned} \tag{A.6}$$

  Then, the full conditional probability of Y^*_{ij} is a truncated Normal with mean Z_{ij} and variance 1. By denoting $\hat{\mu}_i = (\mathbf{B}^T\mathbf{X}_i^T + \mu)$, we have the full conditional distribution of $\mathbf{Z}_i$:

$$\mathbf{Z}_i|Y^*_i, \hat{\mu}_i, \mathbf{\Omega} \sim N([\mathbf{\Omega} + I]^{-1}(\hat{\mu}_i + Y^*_i), [\mathbf{\Omega} + I]^{-1}).$$

- *Log-normal Poisson model for counts.* To model a response of multivariate counts, we use a Lognormal-Poisson model [42]. Let $\mathbf{Z}_i \in \mathbb{R}^k$ be the latent vector of log expected counts of the i^{th} sample and let $\mathbf{Y}_i \in \mathbb{N}^k$ be the observed counts. We use $\mathbf{Z}_{i,-j} \in \mathbb{R}^{k-1}$ to denote the vector of log expected counts of the i^{th} sample but without response j and Z_{ij} as the log expected counts of the i^{th} sample and j^{th} response.
  The covariance matrix accounts for both over-dispersion and correlation of the counts:

$$\begin{aligned} \mathbf{Z}_i &\sim N(\mathbf{\Omega}^{-1}(\mathbf{B}^T\mathbf{X}_i^T + \mu), \mathbf{\Omega}^{-1}), \\ \lambda_{ij} &= \exp(Z_{ij}), \\ Y_{ij} &\sim \text{Poisson}(\lambda_{ij}). \end{aligned} \tag{A.7}$$

  Then, the density of Y_{ij} is as follows:

$$p(Y_{ij}|Z_{ij}) \propto \exp\{Y_{ij}Z_{ij} - e^{Z_{ij}}\}.$$

Let $Z_{ij}|\mathbf{Z}_{i,-j} \sim N(\tilde{\mu}_{ij}, \tilde{\sigma}^2_{ij})$ be the conditional prior so that the log full conditional is:

$$\log[p(Z_{ij}|\mathbf{Z}_{i,-j}, \hat{\mu}, \Omega, Y)] = Y_{ij}Z_{ij} - \exp(Z_{ij}) - \frac{1}{2\tilde{\sigma}^2_{ij}}(Z_{ij} - \tilde{\mu}_{ij})^2 + C$$

which is concave. This means that we can sample the full conditional distribution of the latent variables using adaptive rejection sampling (ARS) [43], and this can be done in parallel to further speed up the sampling.

- *Normal-logistic for multinomial data.* As in [44], we develop a normal-logistic model for multinomial compositional data. This type of data is very common in microbiome and ecology studies.
  Assume that we have $k + 1$ responses in our sample and the last response serves as reference group. Let $\mathbf{Z}_i \in \mathbb{R}^{k+1}$ denote the latent vector of logit transformed relative abundances for i^{th} sample, and let $\mathbf{Y}_i \in \mathbb{N}^k$ be the observed species counts. Denote the known total count as M (e.g., sequence depth in microbiome studies). Similarly we use $\mathbf{Z}_{i,-j}$ to denote the vector logit transformed relative abundance of the i^{th} sample but without response j and Z_{ij} as the log expected counts of the i^{th} sample and j^{th} response.
  The Normal-Logistic model has the following structure:

$$\begin{aligned} \mathbf{Z}_i &\sim N(\mathbf{\Omega}^{-1}(\mathbf{B}^T\mathbf{X}_i^T + \mu), \mathbf{\Omega}^{-1}), \\ p_{ij} &= \frac{\exp(Z_{ij})}{\sum_{i=1}^k \exp(Z_{ij}) + 1}, \\ \mathbf{Y}_i &\sim \text{Multinomial}(p_{i1}, \ldots, p_{ik}, M). \end{aligned} \tag{A.8}$$

  Note that the Normal latent variables take care of the over-dispersion.
  Then, the likelihood of $\mathbf{Y}_i$ is:

$$\begin{aligned} p(\mathbf{Y}_i|\mathbf{Z}_i) &= M! \prod_{j=1}^k \frac{1}{Y_{ij}!} \frac{\exp(Y_{ij}Z_{ij})}{\sum_{j=1}^k \exp(Z_{ij}) + 1} \\ &= \frac{M!}{\prod_{j=1}^k Y_{ij}!} \frac{\exp(\sum_{j=1}^k Y_{ij}Z_{ij})}{(\sum_{j=1}^k \exp(Z_{ij}) + 1)^M} \end{aligned}$$

  Let $Z_{ij}|\mathbf{Z}_{i,-j} \sim N(\tilde{\mu}_{ij}, \tilde{\sigma}^2_{ij})$ be the conditional prior so that the log full conditional is as follows:

$$\log[p(Z_{ij}|\mathbf{Z}_{i,-j}, \hat{\mu}, \Omega, Y)] = Y_{ij}Z_{ij} - M\log\left(\sum_{j=1}^k \exp(Z_{ij}) + 1\right) - \frac{1}{2\tilde{\sigma}^2_{ij}}(Z_{ij} - \tilde{\mu}_{ij})^2 + C$$

  This function is concave because the first term is an affine, the second term is the negative log sum of exponential of an affine function, and the last term is a concave quadratic form. Thus, ARS [43] can be used again during the Gibbs sampling, and this process can be parallelized for extra speed.

B. Derivation of the Gibbs sampling

Let $\mathbf{1}_n$ be the column vector of ones with dimension n, let $\mathbf{S} = \mathbf{Y}^T\mathbf{Y} \in \mathbb{R}^{k\times k}$ (here we have samples as row vectors in $\mathbf{Y}$), let $\hat{\mu} = \mathbf{XB} + \mathbf{1}_n\mu^T$, and let $\mathbf{U} = \hat{\mu}^T\hat{\mu} \in \mathbb{R}^{k\times k}$. Equation (B.1) shows the full conditional distribution of $\mathbf{\Omega}$ and η.

$$\begin{aligned} p(\mathbf{\Omega}, \eta|\mathbf{Y}, \lambda_\Omega, \hat{\mu}) \propto |\mathbf{\Omega}|^{\frac{n}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}(\mathbf{S\Omega}) - \frac{1}{2}\operatorname{tr}(\mathbf{U\Omega}^{-1})\right) \prod_{q<q'}\left[\frac{1}{\sqrt{2\pi\eta_{qq'}}}\exp\left(-\frac{\omega_{qq'}^2}{2\eta_{qq'}}\right)\right] \\ \times \prod_{q=1}^{k}\left[\frac{\lambda_\Omega}{2}\exp\left(-\frac{\lambda_\Omega\omega_{qq}}{2}\right)\right] I_{\mathbf{\Omega}\in M^+} \end{aligned} \tag{B.1}$$

We can update one row (column) at one iteration. Let $\mathbf{H}$ be the symmetric matrix with $\mathbf{H}_{ml} = \mathbf{H}_{lm} = \eta_{ml}$ $(m < l)$ on the off-diagonal entries and on the diagonal $\mathbf{H}_{mm} = 0$. We take one column out and partition $\mathbf{\Omega}$, $\mathbf{S}$, $\mathbf{U}$, and $\mathbf{H}$. Without lose of generality, we show the sampling scheme for the last row (column). Let $\mathbf{\Omega}_{11} \in \mathbb{R}^{(k-1)\times(k-1)}$, $\boldsymbol{\omega}_{12} \in \mathbb{R}^{k-1}$, and $\omega_{22} \in \mathbb{R}$. We partition $\mathbf{S}$, $\mathbf{U}$ and $\mathbf{H}$ in the same manner.

$$\mathbf{\Omega} = \begin{bmatrix}\mathbf{\Omega}_{11} & \boldsymbol{\omega}_{12}\\ \boldsymbol{\omega}_{12}^T & \omega_{22}\end{bmatrix}, \mathbf{S} = \begin{bmatrix}\mathbf{S}_{11} & \boldsymbol{s}_{12}\\ \boldsymbol{s}_{12}^T & s_{22}\end{bmatrix}, \mathbf{U} = \begin{bmatrix}\mathbf{U}_{11} & \boldsymbol{u}_{12}\\ \boldsymbol{u}_{12}^T & u_{22}\end{bmatrix}, \mathbf{H} = \begin{bmatrix}\mathbf{H}_{11} & \boldsymbol{\eta}_{12}\\ \boldsymbol{\eta}_{12}^T & 0\end{bmatrix}.$$

By setting

$$\gamma = \omega_{22} - \boldsymbol{\omega}_{12}^T\mathbf{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12} \in \mathbb{R}, \tag{B.2}$$

$\mathbf{\Omega}^{-1}$ can be written in a block form as follows:

$$\mathbf{\Omega}^{-1} = \begin{bmatrix}\mathbf{\Omega}_{11}^{-1} + \frac{1}{\gamma}\mathbf{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}\boldsymbol{\omega}_{12}^T\mathbf{\Omega}_{11}^{-1} & -\frac{1}{\gamma}\mathbf{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}\\ -\frac{1}{\gamma}\boldsymbol{\omega}_{12}^T\mathbf{\Omega}_{11}^{-1} & \frac{1}{\gamma}\end{bmatrix}.$$

Given

$$\operatorname{tr}(\mathbf{U\Omega}^{-1}) = \operatorname{tr}(\mathbf{U}_{11}\mathbf{\Omega}_{11}^{-1}) + \frac{1}{\gamma}(\boldsymbol{\omega}_{12}^T\mathbf{\Omega}_{11}^{-1}\mathbf{U}_{11}\mathbf{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12} - 2\boldsymbol{u}_{12}^T\mathbf{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12} + u_{22}),$$

we have the full conditional distribution of $\boldsymbol{\omega}_{12}$ and γ as follows:

$$\begin{aligned} p(\boldsymbol{\omega}_{12}, \gamma|\mathbf{\Omega}_{11}, \eta, \lambda_\Omega) \propto & \gamma^{\frac{n}{2}}\exp\left(-\frac{1}{2}(s_{22}+\lambda_\Omega)\gamma - \frac{u_{22}}{2\gamma}\right)\\ & \times \exp\{-[\boldsymbol{s}_{12} - \frac{1}{\gamma}\mathbf{\Omega}_{11}^{-1}\boldsymbol{u}_{12}]^T\boldsymbol{\omega}_{12}\\ & - \frac{1}{2}\boldsymbol{\omega}_{12}^T[D_\eta^{-1} + (s_{22}+\lambda_\Omega)\mathbf{\Omega}_{11}^{-1} + \frac{1}{\gamma}\mathbf{\Omega}_{11}^{-1}\mathbf{U}_{11}\mathbf{\Omega}_{11}^{-1}]\boldsymbol{\omega}_{12}\}. \end{aligned}$$

From the above equation, we get a closed form expression for the conditional distribution of γ as follows:

$$\begin{aligned} & p(\gamma|\boldsymbol{\omega}_{12}, \mathbf{\Omega}_{11}, \eta, \lambda_\Omega) \propto\\ & \gamma^{\frac{n}{2}}\exp\left(-\frac{1}{2}(s_{22}+\lambda_\Omega)\gamma - \frac{u_{22} - 2\boldsymbol{u}_{12}^T\mathbf{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12} + \boldsymbol{\omega}_{12}^T\mathbf{\Omega}_{11}^{-1}\mathbf{U}_{11}\mathbf{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}}{2\gamma}\right) I_{\gamma\geq 0} \end{aligned} \tag{B.3}$$

which is a generalized inverse Gaussian (GIG) distribution [45, 46] with parameters:

$$\begin{aligned}\lambda &= \frac{n}{2} + 1,\\ \psi &= s_{22} + \lambda_\Omega,\\ \chi &= u_{22} - 2\boldsymbol{u}_{12}^T\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12} + \boldsymbol{\omega}_{12}^T\boldsymbol{\Omega}_{11}^{-1}\mathbf{U}_{11}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}.\end{aligned}$$

GIG has a positive support. Thus, the determinant and the k^{th} principal minor of the updated $\boldsymbol{\Omega}$ are positive, while the first $k-1$ principal minors remain unchanged and positive. In this manner, the updated $\boldsymbol{\Omega}$ always remains positive definite.

By denoting $\mathbf{D}_\eta = \text{diag}(\boldsymbol{\eta}_{12}) \in \mathbb{R}^{(k-1)\times(k-1)}$, the full conditional distribution of $\boldsymbol{\omega}_{12}$ is a Normal distribution:

$$\begin{aligned}p(\boldsymbol{\omega}_{12}|\gamma, \boldsymbol{\Omega}_{11}, \eta, \lambda_\Omega) \propto \exp\{&-[\boldsymbol{s}_{12} - \frac{1}{\gamma}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{u}_{12}]^T\boldsymbol{\omega}_{12}\\ &- \frac{1}{2}\boldsymbol{\omega}_{12}^T[\mathbf{D}_\eta^{-1} + (s_{22} + \lambda_\Omega)\boldsymbol{\Omega}_{11}^{-1} + \frac{1}{\gamma}\boldsymbol{\Omega}_{11}^{-1}\mathbf{U}_{11}\boldsymbol{\Omega}_{11}^{-1}]\boldsymbol{\omega}_{12}\}\end{aligned} \tag{B.4}$$

with parameters:

$$\begin{aligned}\boldsymbol{\Sigma}_{\omega_{12}}^{-1} &= \mathbf{D}_\eta^{-1} + (s_{22} + \lambda_\Omega)\boldsymbol{\Omega}_{11}^{-1} + \frac{1}{\gamma}\boldsymbol{\Omega}_{11}^{-1}\mathbf{U}_{11}\boldsymbol{\Omega}_{11}^{-1}\\ \mu_{\omega_{12}} &= -\boldsymbol{\Sigma}_{\omega_{12}}[\boldsymbol{s}_{12} - \frac{1}{\gamma}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{u}_{12}].\end{aligned}$$

As in [12], the $z_{qq'} = 1/\eta_{qq'}$ are independent Inverse Gaussians with the following parameters:

$$\begin{aligned}\mu_{z_{qq'}} &= \sqrt{\lambda_\Omega^2/\omega_{qq'}^2},\\ \lambda_{z_{qq'}} &= \lambda_\Omega^2,\end{aligned}$$

and density:

$$p(z_{qq'}|\boldsymbol{\Omega}, \lambda_\Omega) = \left(\frac{\lambda_{z_{qq'}}}{2\pi z_{qq'}^3}\right)^{1/2} \exp\left(\frac{-\lambda_{z_{qq'}}(z_{qq'} - \mu_{z_{qq'}})^2}{2(\mu_{z_{qq'}})^2 z_{qq'}}\right) I_{z_{qq'}>0}. \tag{B.5}$$

The full conditional distribution of vec($\mathbf{B}$) can be represented using tensor product [20]. Let $\mathbf{D}_{\tau^2} = \text{diag}(\tau^2) \in \mathbb{R}^{kp\times kp}$ for τ the scaling parameters in the prior density of $\mathbf{B}$. Then, the conditional distribution of vec($\mathbf{B}$) has the following form:

$$\begin{aligned}p(\text{vec}(\mathbf{B})|\mathbf{D}_{\tau^2}, \boldsymbol{\Omega}, \mu, \mathbf{X}, \mathbf{Y}) \propto \exp\{&\mathbf{X}^T(\mathbf{Y} - \mathbf{1}_n\mu^T\boldsymbol{\Omega}^{-1})\\ &- \frac{1}{2}\,\text{vec}(\mathbf{B})^T(\boldsymbol{\Omega}^{-1} \otimes \mathbf{X}^T\mathbf{X} + \mathbf{D}_{\tau^2}^{-1})\,\text{vec}(\mathbf{B})\}.\end{aligned} \tag{B.6}$$

Note that the information from data is encoded by $\boldsymbol{\Omega}^{-1} \otimes \mathbf{X}^T\mathbf{X}$ which differs from the canonical parameterization of the multiresponse linear regression model in which the information from data is encoded by $\boldsymbol{\Omega} \otimes \mathbf{X}^T\mathbf{X}$. This is because in the kernel of the likelihood, the term that involves $\mathbf{B}$ is $\mathbf{X}_i\mathbf{B}\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}\mathbf{B}^T\mathbf{X}_i^T = \mathbf{X}_i\mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{B}^T\mathbf{X}_i^T$, instead of $\mathbf{X}_i\tilde{\mathbf{B}}\boldsymbol{\Omega}\tilde{\mathbf{B}}^T\mathbf{X}_i^T$ as in the canonical parametrization.

Finally, we update $1/\tau_{jq}$ using an Inverse Gaussian distribution with parameters $\sqrt{\lambda_\beta^2/B_{jq}^2}$ and λ_β^2, and we update μ using a Normal distribution with mean $(\mathbf{Y}\boldsymbol{\Omega} - \mathbf{X}\mathbf{B})^T$ and variance $\boldsymbol{\Omega}/n$.

C. Simulation settings

Below we provide the details on the $\mathbf{\Omega}$ structure for the six graphical models:

- Model 1 (AR1): An AR(1) model with $\sigma_{qq'} = 0.7^{|q-q'|}$.
- Model 2 (AR2): An AR(2) model with $\omega_{qq} = 1$, $\omega_{q-1,q} = \omega_{q,q-1} = 0.5$, $\omega_{q-2,q} = \omega_{q,q-2} = 0.25$ for $i = 1, \ldots, k$.
- Model 3 (Block): A block model with $\sigma_{qq} = 1$ for $q = 1, \ldots, k$, $\sigma_{qq'} = 0.5$ for $1 \leq q \neq q' \leq k/2$, $\sigma_{qq'} = 0.5$ for $k/2 + 1 \leq q \neq q' \leq 10$ and $\sigma_{qq'} = 0$ otherwise.
- Model 4 (Star): A star model with every node connected to the first node, with $\omega_{qq} = 1$, $\omega_{1,q} = \omega_{q,1} = 0.1$ for $q = 1, \ldots, k$, and $\omega_{qq'} = 0$ otherwise.
- Model 5 (Circle): A circle model with $\omega_{qq} = 2$, $\omega_{q-1,q} = \omega_{q,q-1} = 1$ for $q = 1, \ldots, k$, and $\omega_{1,q'} = \omega_{q',1} = 0.9$ for $q' = 1, \ldots, k$.
- Model 6 (Dense): A full model with $\omega_{qq} = 2$ and $\omega_{qq'} = 1$ for $q \neq q' \in \{1, \ldots, k\}$.

Note that model 1 and model 3 specify the entries of the covariance matrix $\mathbf{\Sigma}$ ($\sigma_{qq'}$) while the other models specify the entries of the precision matrix $\mathbf{\Omega}$ ($\omega_{qq'}$).

D. Log concavity of posterior with fixed λ's

Here, we show the log-concavity of the posterior which makes the Gibbs sampler efficient for any fixed λ.

The log-posterior is equivalent to a penalized likelihood. Let $\hat{\mu} = \mathbf{XB} + \mathbf{1}_n \mu^T$, and let $\mathbf{U} = \hat{\mu}^T \hat{\mu} \in \mathbb{R}^{k \times k}$.

$$\log p(\mathbf{B}, \mathbf{\Omega} | \mathbf{X}, \mathbf{Y}) = C + \frac{n}{2} \log(|\mathbf{\Omega}|) - \frac{1}{2} \operatorname{tr}\left(\mathbf{Y^T Y \Omega}\right) - \frac{1}{2} \operatorname{tr}\left(\mathbf{U \Omega}^{-1}\right) - \lambda_\Omega ||\mathbf{\Omega}||_1 - \lambda_\beta ||\mathbf{B}||_1$$

The last two terms are concave, so it remains to show that the first three terms are concave too. That is, we want to show that the log-likelihood is concave as well.

This can be shown by calculating the Hessian. We observe that the random component $\mathbf{Y}$ is only involved in a linear term of $\mathbf{\Omega}$. Thus, the Hessian has no $\mathbf{Y}$ involved, i.e., the Hessian of the log-likelihood is itself the negative Fisher information of the chain graph model (expectation of constant is constant). Thus, the Hessian must be negative definite, thus we have the posterior being log-concave.

E. More simulation results

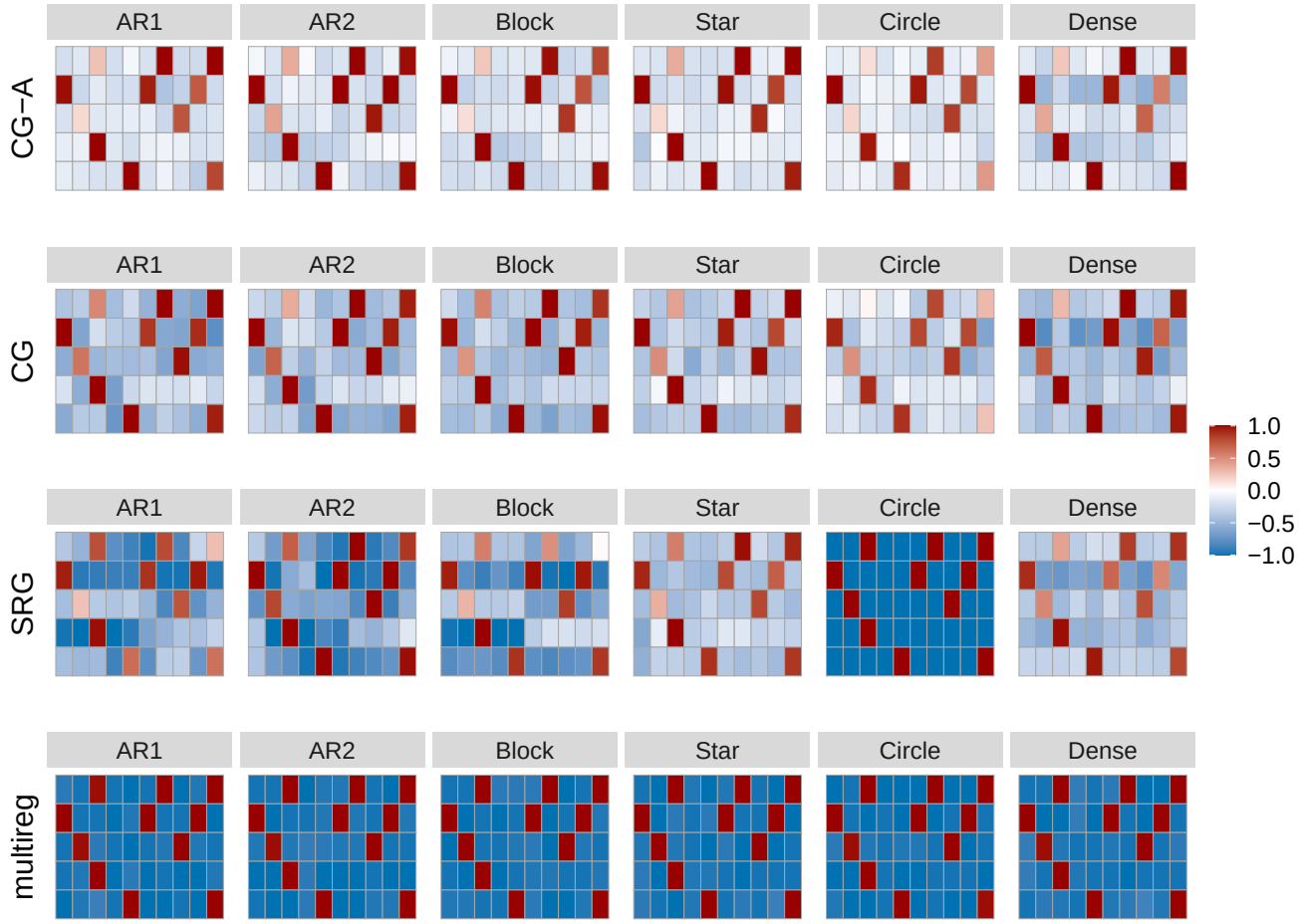

Figure A1. Reconstruction accuracy of the graph between responses and predictors (**B**) for $k = 10$ nodes, $p = 5$ predictors and sparsity of 0.8. Red entries correspond to true positive edges and blue entries correspond to false positive edges. Darker color means higher frequency of being estimated in 50 reconstructions. Our proposed method adaptive CG-LASSO (CG-A) outperforms the other methods by displaying the lowest false positive rate (less blue).

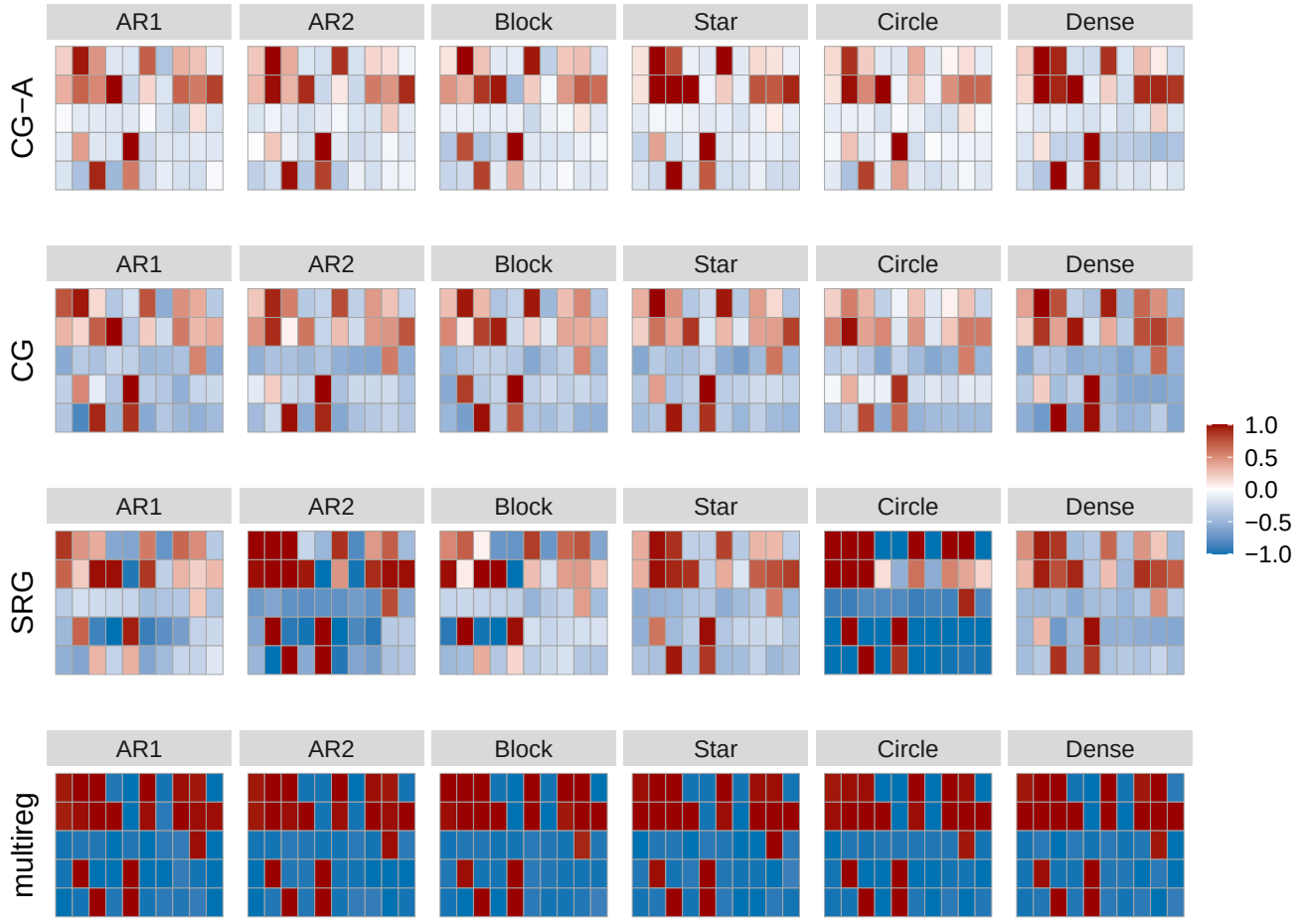

Figure A2. Reconstruction accuracy of the graph between responses and predictors (**B**) for $k = 10$ nodes, $p = 5$ predictors and sparsity of 0.5. Red entries correspond to true positive edges and blue entries correspond to false positive edges. Darker color means higher frequency of being estimated in 50 reconstructions. Our proposed method adaptive CG-LASSO (CG-A) outperforms the other methods by displaying the lowest false positive rate (less blue).

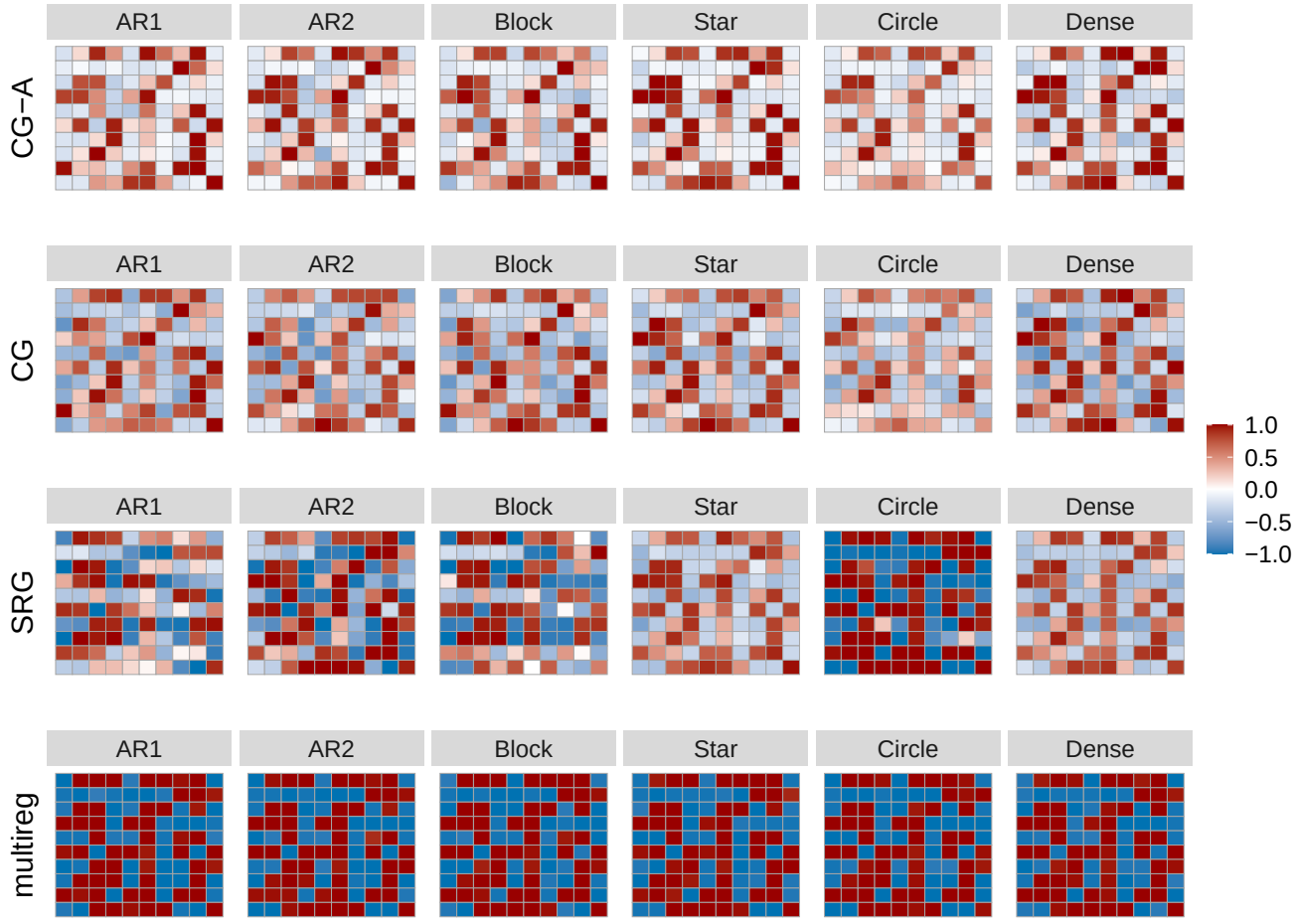

Figure A3. Reconstruction accuracy of the graph between responses and predictors (**B**) for $k = 10$ nodes, $p = 10$ predictors and sparsity of 0.5. Red entries correspond to true positive edges and blue entries correspond to false positive edges. Darker color means higher frequency of being estimated in 50 reconstructions. Our proposed method adaptive CG-LASSO (CG-A) outperforms the other methods by displaying the lowest false positive rate (less blue).

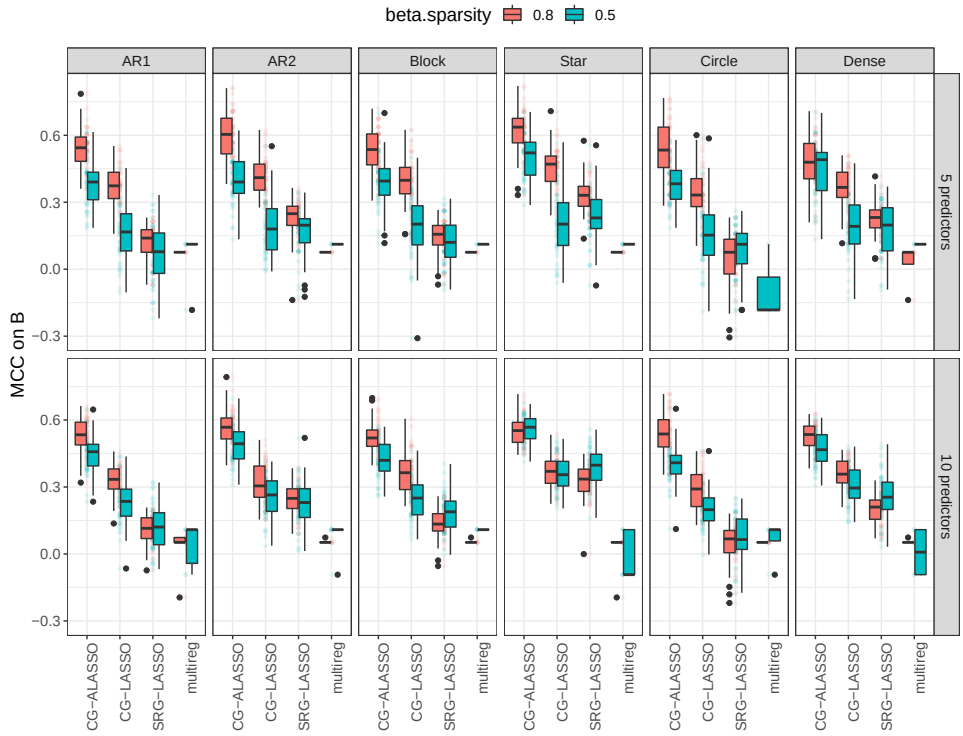

Figure A4. Matthews Correlation Coefficients for **B** for simulated datasets with 10 nodes and 50 samples under two levels of beta sparsity (red 0.8 and blue 0.5), two different number of predictors (10 in bottom row and 5 in top row) and six precision matrix models (columns, fully connected model was omitted from $\mathbf{\Omega}$ result since MCC was not defined). X-axis corresponds to the models compared. MCC = 1 means a perfect reconstruction. Our model adaptive CG-LASSO gets the highest MCC in most cases. We omit the `multireg_mu0` model because it performs poorly across all cases (MCC close to 0).

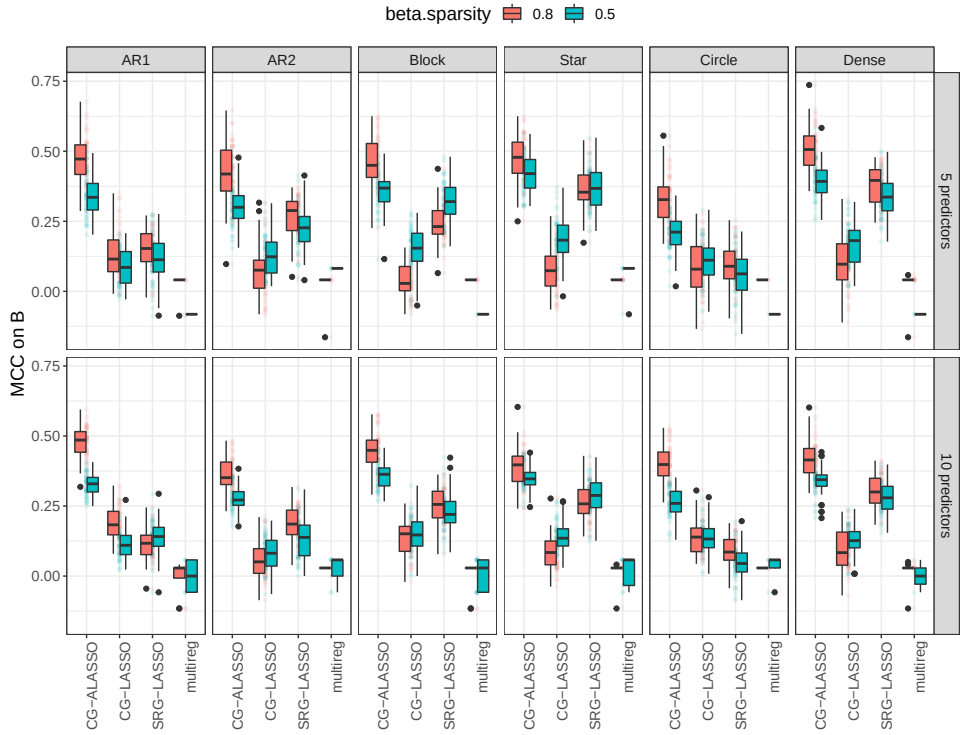

Figure A5. Matthews Correlation Coefficients for **B** for simulated datasets with 30 nodes and 50 samples under two levels of beta sparsity (red 0.8 and blue 0.5), two different number of predictors (10 in bottom row and 5 in top row) and six precision matrix models (columns, fully connected model was omitted from $\mathbf{\Omega}$ result since MCC was not defined). X-axis corresponds to the models compared. MCC = 1 means a perfect reconstruction. Our model adaptive CG-LASSO gets the highest MCC in most cases. We omit the `multireg_mu0` model because it performs poorly across all cases (MCC close to 0).

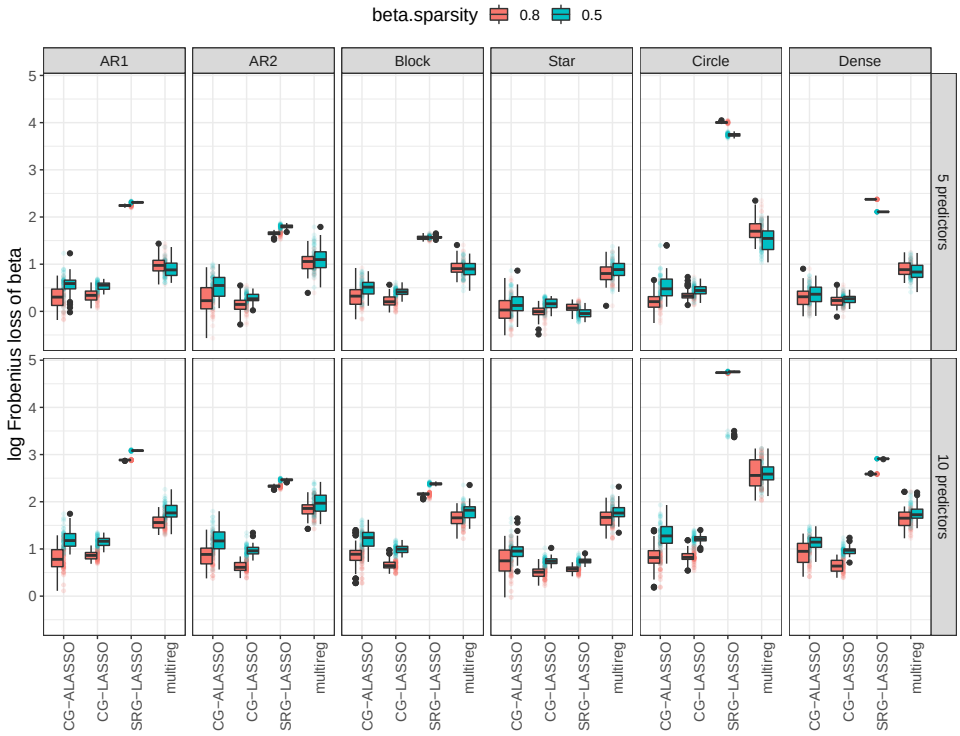

Figure A6. Frobenius Loss of **B** (Y-axis in logarithmic scale) for simulated datasets with 10 nodes and 50 samples under two levels of beta sparsity (red 0.8 and blue 0.5), two different number of predictors (10 in bottom row and 5 in top row) and six precision matrix models (columns). X-axis corresponds to the models compared. Our models (adaptive) CG-LASSO get the lowest loss in most cases.

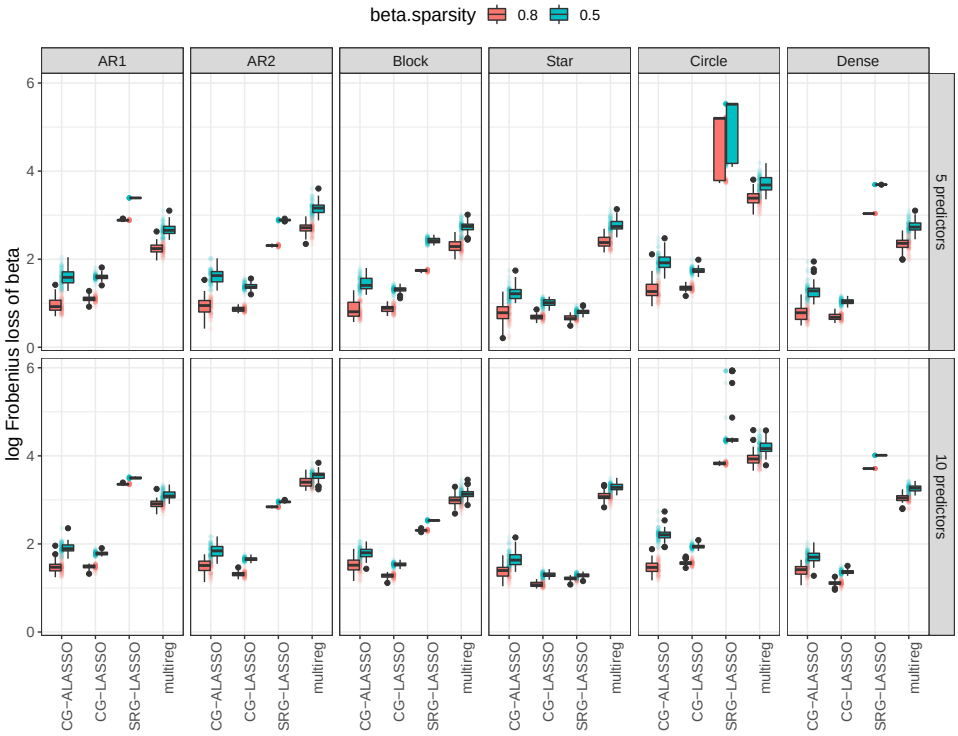

Figure A7. Frobenius Loss of **B** (Y-axis in logarithmic scale) for simulated datasets with 30 nodes and 50 samples under two levels of beta sparsity (red 0.8 and blue 0.5), two different number of predictors (10 in bottom row and 5 in top row) and six precision matrix models (columns). X-axis corresponds to the models compared. Our models (adaptive) CG-LASSO get the lowest loss in most cases.

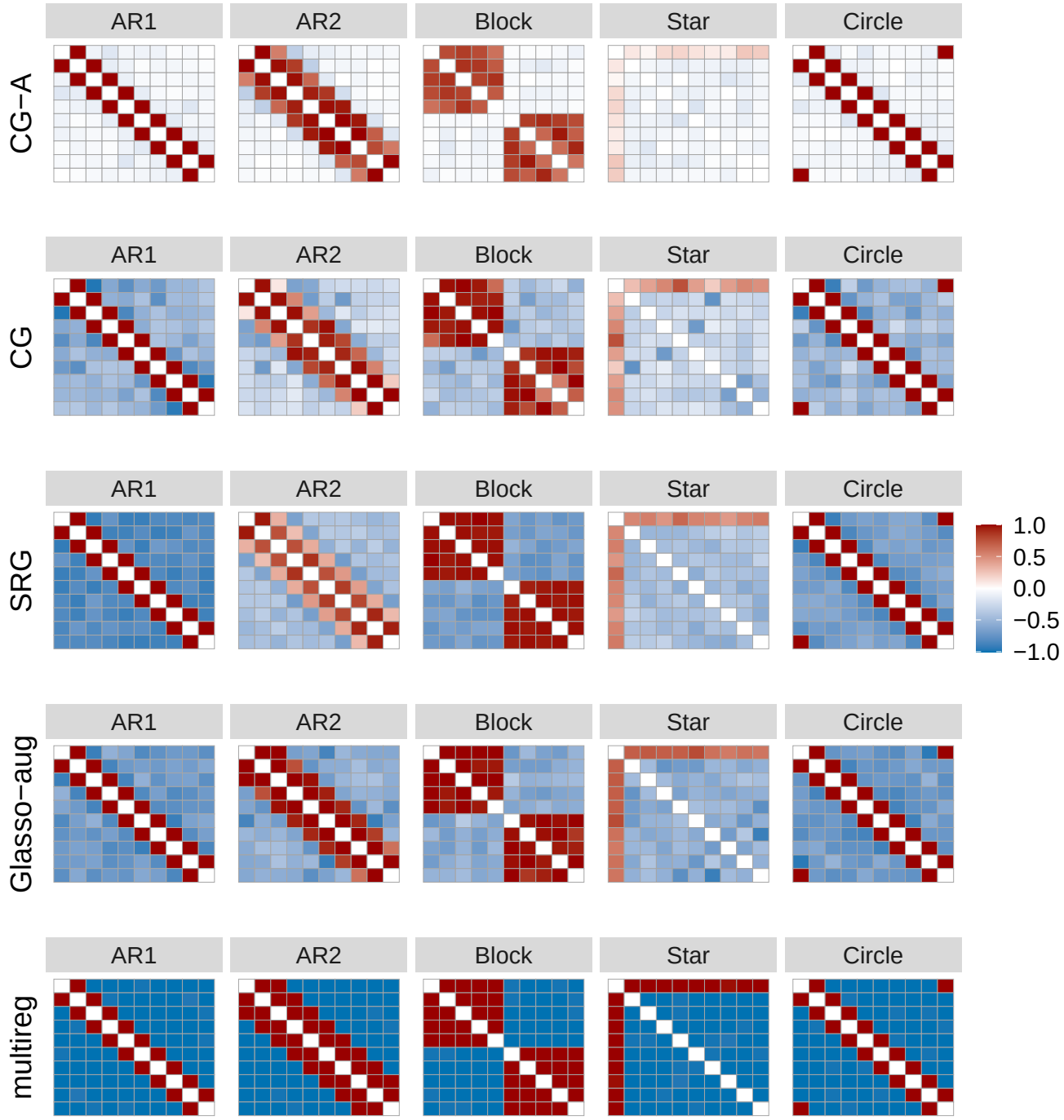

Figure A8. Reconstruction accuracy of the graph among responses ($\mathbf{\Omega}$) for $k = 10$ nodes, $p = 5$ predictors and sparsity of 0.8. Red entries correspond to true positive edges and blue entries correspond to false positive edges. Darker color means higher frequency of being estimated in 50 reconstructions. Our proposed method adaptive CG-LASSO (CG-A) outperforms the other methods by displaying the lowest false positive rate (less blue). We omit the dense model because it has no false positive or true negatives.

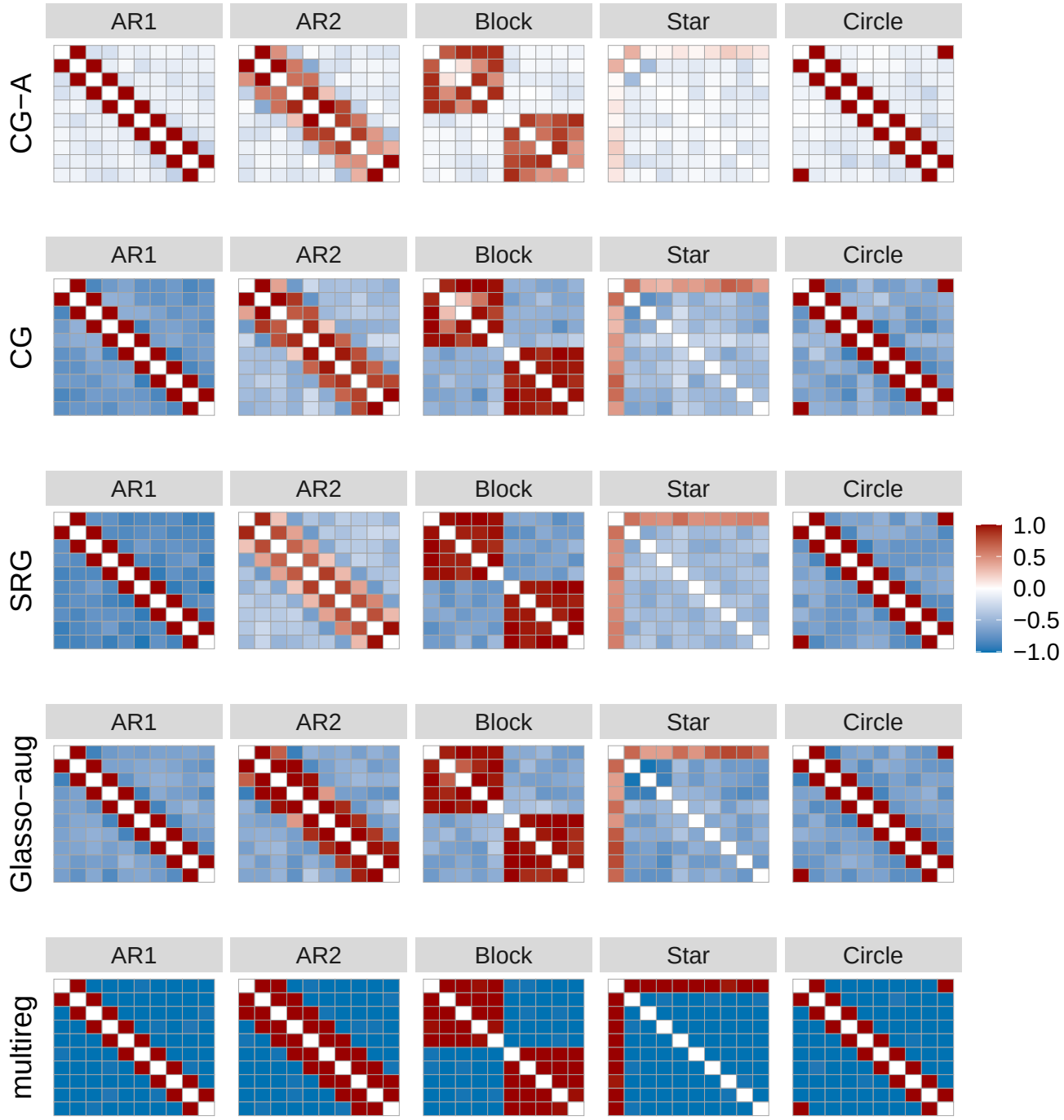

Figure A9. Reconstruction accuracy of the graph among responses ($\mathbf{\Omega}$) for $k = 10$ nodes, $p = 5$ predictors and sparsity of 0.5. Red entries correspond to true positive edges and blue entries correspond to false positive edges. Darker color means higher frequency of being estimated in 50 reconstructions. Our proposed method adaptive CG-LASSO (CG-A) outperforms the other methods by displaying the lowest false positive rate (less blue). We omit the dense model because it has no false positive or true negatives.

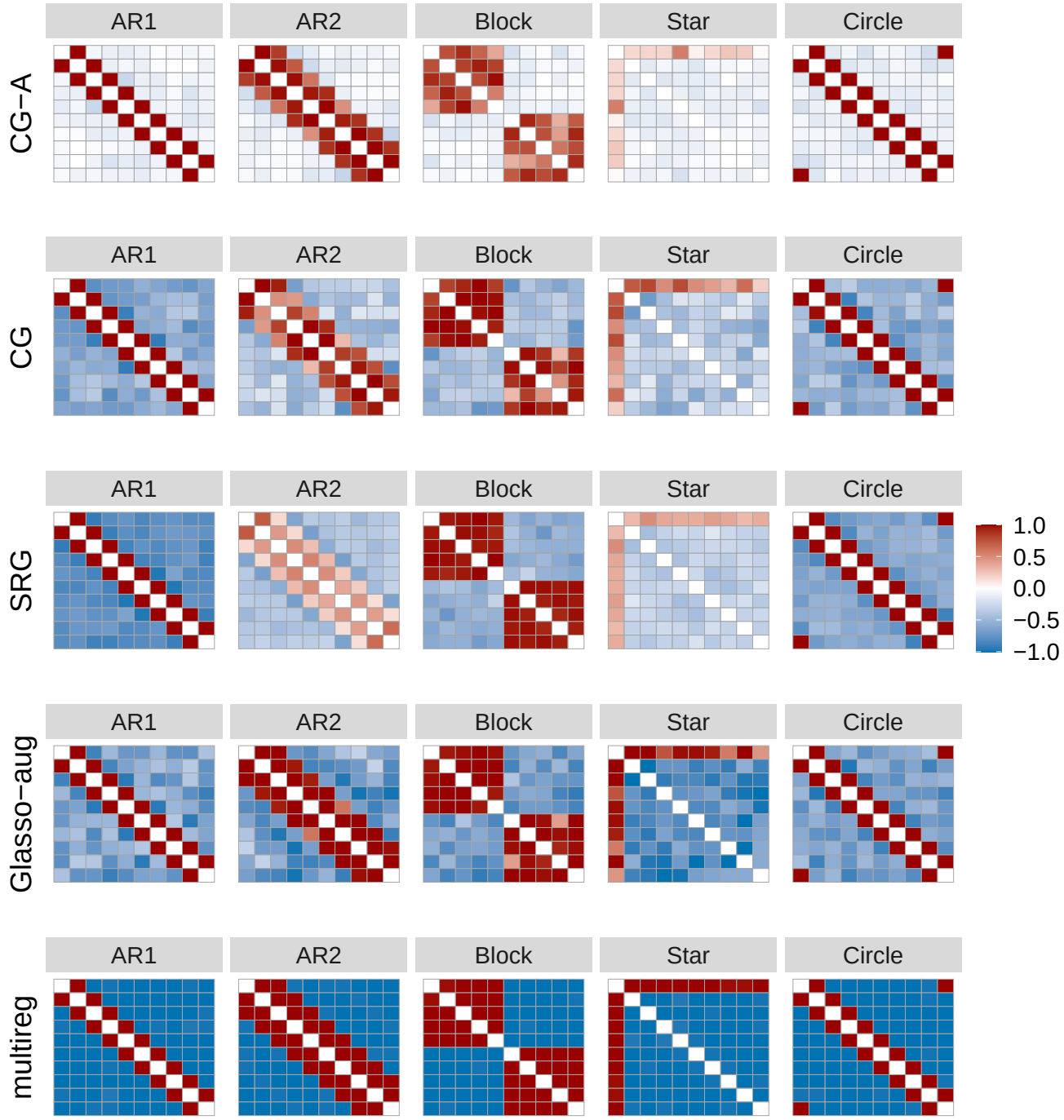

Figure A10. Reconstruction accuracy of the graph among responses ($\mathbf{\Omega}$) for $k = 10$ nodes, $p = 10$ predictors and sparsity of 0.5. Red entries correspond to true positive edges and blue entries correspond to false positive edges. Darker color means higher frequency of being estimated in 50 reconstructions. Our proposed method adaptive CG-LASSO (CG-A) outperforms the other methods by displaying the lowest false positive rate (less blue). We omit the dense model because it has no false positive or true negatives.

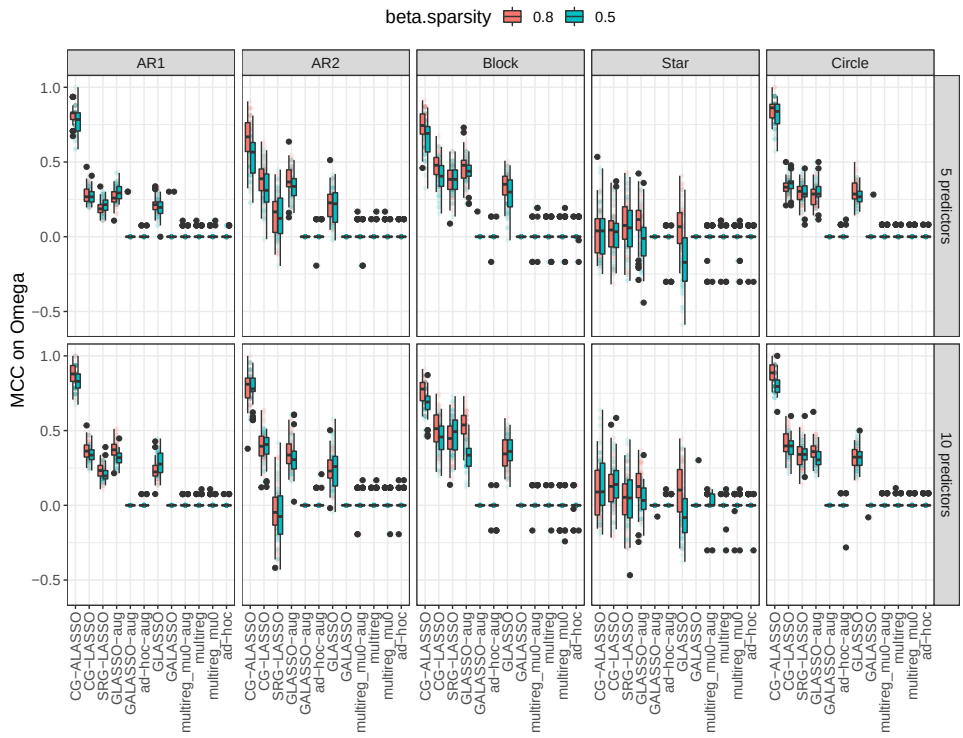

Figure A11. Matthews Correlation Coefficients for $\mathbf{\Omega}$ for simulated datasets with 10 nodes and 50 samples under two levels of beta sparsity (red 0.8 and blue 0.5), two different number of predictors (10 in bottom row and 5 in top row) and six precision matrix models (columns, fully connected model was omitted from $\mathbf{\Omega}$ result since MCC was not defined). X-axis corresponds to the models compared. MCC = 1 means a perfect reconstruction. Our model adaptive CG-LASSO gets the highest MCC in most cases. We omit the dense model because MCC is not defined.

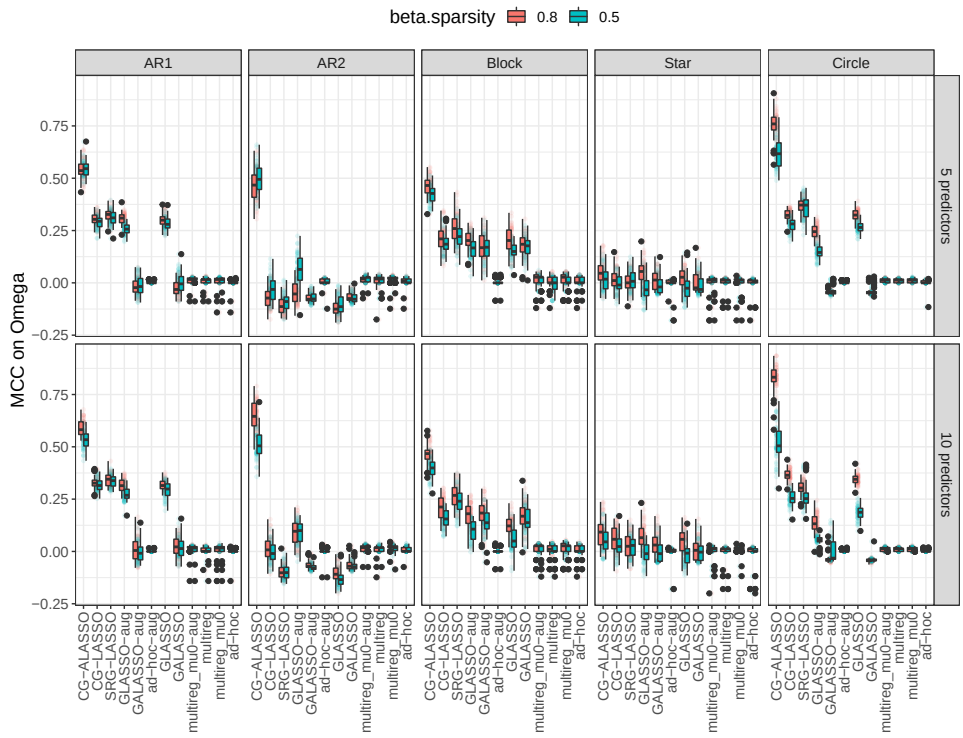

Figure A12. Matthews Correlation Coefficients for $\mathbf{\Omega}$ for simulated datasets with 30 nodes and 50 samples under two levels of beta sparsity (red 0.8 and blue 0.5), two different number of predictors (10 in bottom row and 5 in top row) and six precision matrix models (columns, fully connected model was omitted from $\mathbf{\Omega}$ result since MCC was not defined). X-axis corresponds to the models compared. MCC = 1 means a perfect reconstruction. Our model adaptive CG-LASSO gets the highest MCC in most cases. We omit the dense model because MCC is not defined.

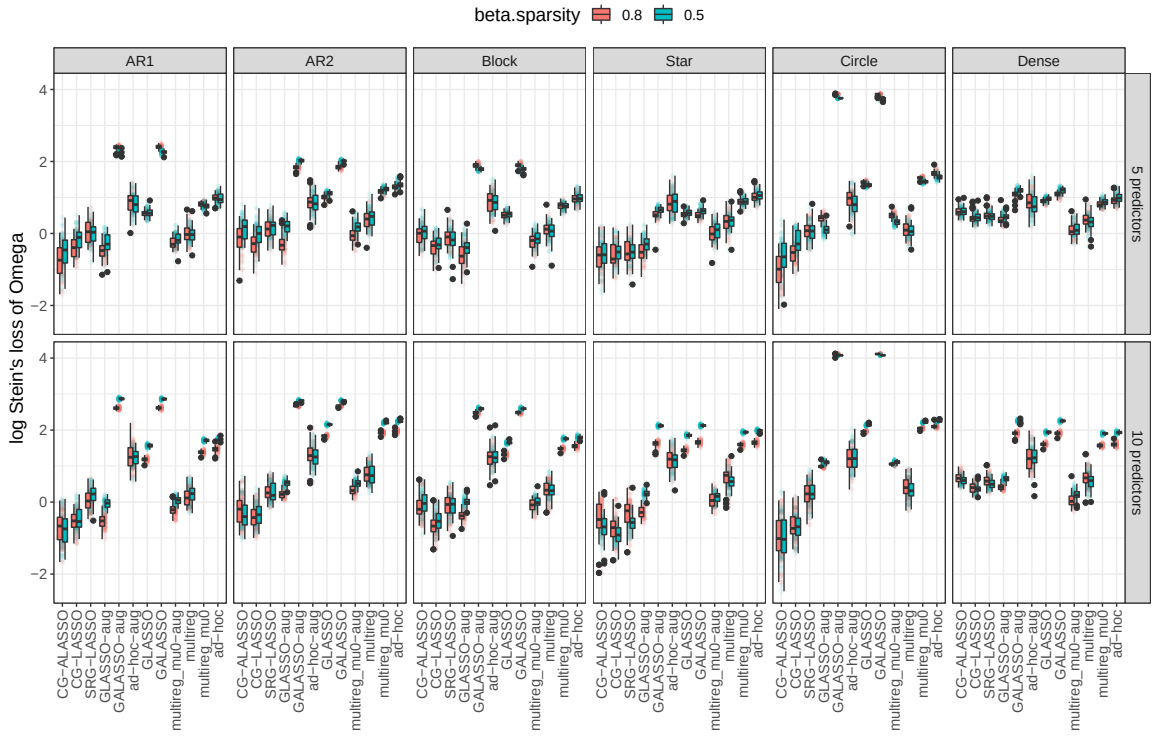

Figure A13. Stein's Loss of $\boldsymbol{\Omega}$, (Y-axis in logarithmic scale) for simulated datasets with 10 nodes and 50 samples under two levels of beta sparsity (red 0.8 and blue 0.5), two different number of predictors (10 in bottom row and 5 in top row) and six precision matrix models (columns). X-axis corresponds to the models compared. Our models (adaptive) CG-LASSO get the lowest loss in most cases.

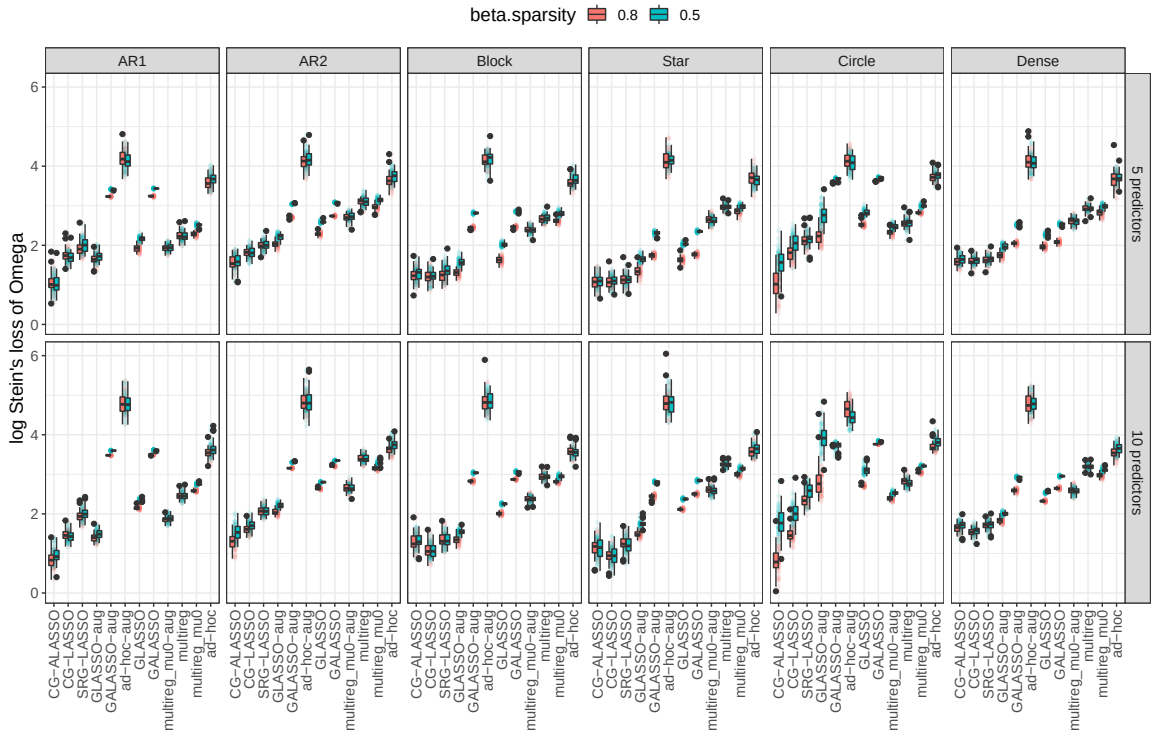

Figure A14. Stein's Loss of $\boldsymbol{\Omega}$, (Y-axis in logarithmic scale) for simulated datasets with 30 nodes and 50 samples under two levels of beta sparsity (red 0.8 and blue 0.5), two different number of predictors (10 in bottom row and 5 in top row) and six precision matrix models (columns). X-axis corresponds to the models compared. Our models (adaptive) CG-LASSO get the lowest loss in most cases.

F. Computational speed and scaling

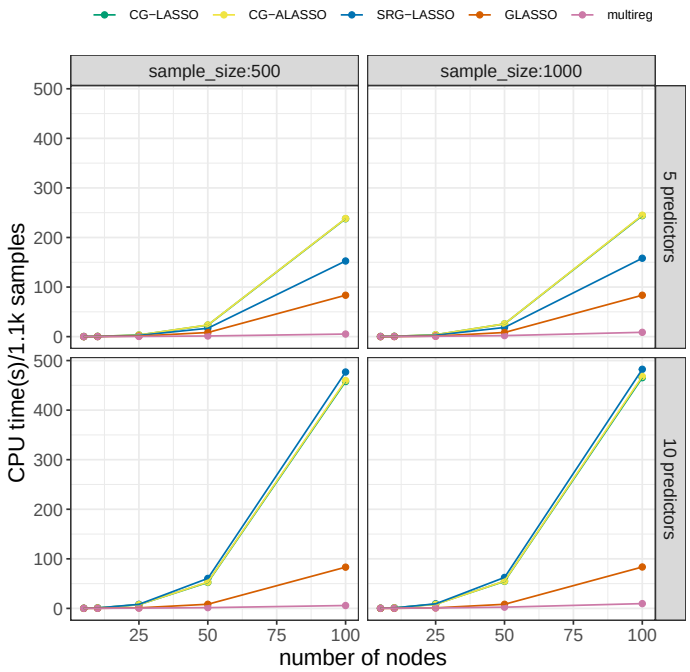

Figure A15. Computational time for each algorithm in CPU seconds as a function of the number of nodes, the number of predictors, and sample size. Speed depends on the number of nodes and number of predictors, but not on the sample size. Our proposed method is efficient, yet slower than graphical LASSO.

G. More on real data analyses

G.1. Transformation of marginal effects into conditional effect in real data

We fit a hierarchical model with multiresponse linear regression as the core and the same logit-multinomial sampling distribution (Equation (A.8) with $\mathbf{Z}_i$ as multiresponse regression model). The density of the networks show that the sparsity assumption on the chain graph model has a strong impact on the estimated network (Figures A16 and A17).

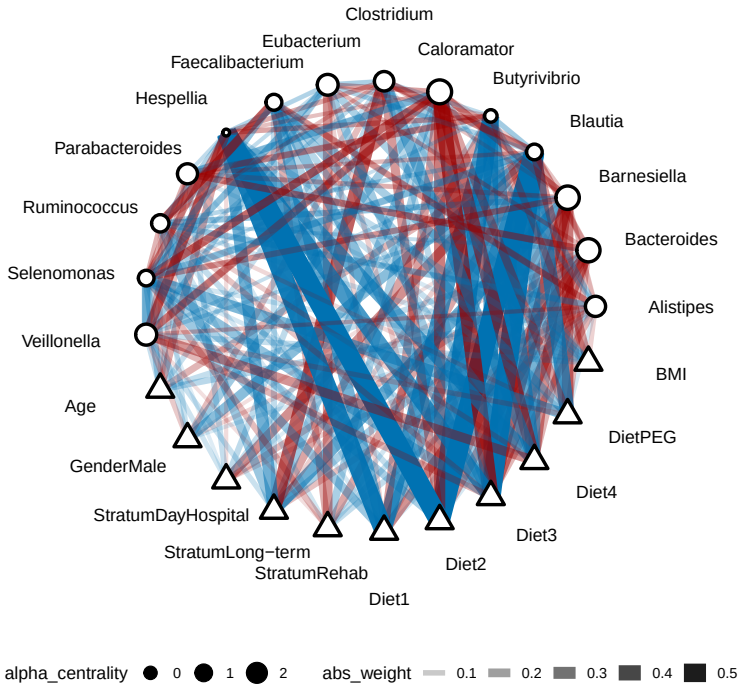

Figure A16. Conditional network of human gut network via multiresponse regression and transformation into chain graph parameterization. Triangle nodes correspond to predictors and circle nodes correspond to relative abundances of genera. The node size on the circle nodes correspond to the $\alpha-$centrality values [27]. The width of the edges corresponds to the absolute weight and the color corresponds to the type of interaction (red positive, blue negative).

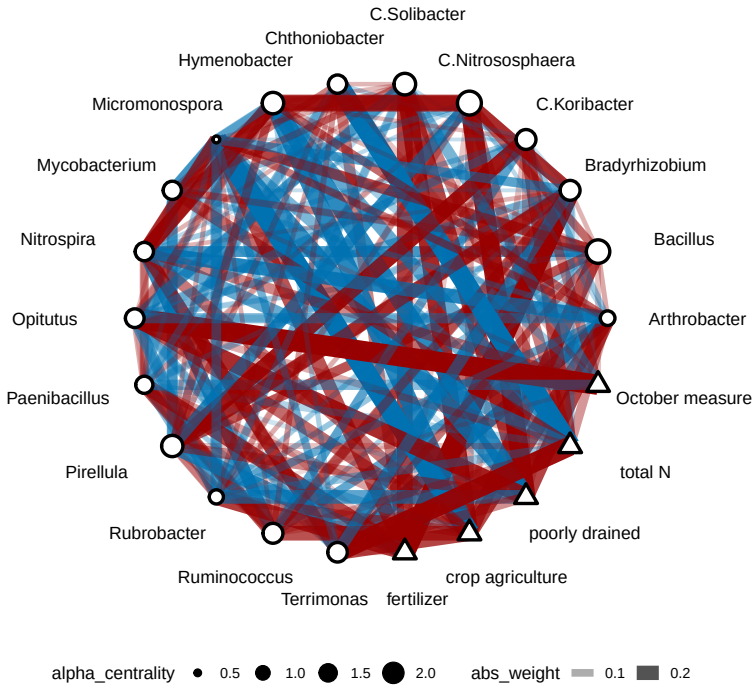

Figure A17. Conditional network of soil network via multiresponse regression and transformation into chain graph parameterization. Triangle nodes correspond to predictors and circle nodes correspond to relative abundances of genera. The node size on the circle nodes correspond to the $\alpha-$centrality values bonacich2001eigenvector. The width of the edges corresponds to the absolute weight and the color corresponds to the type of interaction (red positive, blue negative).