Supporting Information
Computationally efficient Bayesian estimation of graphical networks for omics data

Daniel W. Adrian[1], Erik D. VonKaenel[2], Moses Y. Obiri[2], David J. Degnan[2], Amy C. Sims[2], Kristie Oxford[2], Lisa M. Bramer[2]

[1]Grand Valley State University, Allendale, MI, USA, [2]Pacific Northwest National Laboratory, Richland, WA, USA

Table of Contents

Table of Contents

Section S1. Statistical methods.....S1

Section S2. More details regarding simulation experiments S4

Section S3. Additional edge detection properties of estimated graphs.....S4

Figure S1. Sensitivity, sensitivity, Matthews correlation coefficient, and F1 scores based on simulated data.....S5

S1. Statistical methods

S1.1. Bayesian model framework

We assume the data consists of n samples of p biomolecule measurements denoted by the $n \times p$ matrix $\mathbf{Y}$. The rows of this matrix are the samples, the p-vectors (*column* vectors) $\mathbf{Y}_s$, $s = 1,\ldots,n$, and the columns are the measurements for a single biomolecule, the n-vectors $\mathbf{Y}_j$, $j = 1,\ldots,p$. In practice, we may have samples from multiple groups (and therefore multiple covariance matrices), but we suppress group-level notation for simplicity.

BPlane adopts the BAGUS algorithm's (Gan, Narisetty, and Liang 2019) model framework for the data $\mathbf{Y}$ and the precision matrix $\mathbf{\Theta}$, which can be summarized as follows:

$$\{\mathbf{Y}_s\}_{s=1}^{n} | \mathbf{\Theta} \sim iid\ \mathcal{N}_p\left(\mathbf{0}_p, \mathbf{\Theta}^{-1}\right) \qquad (1)$$

$$\theta_{ij} | z_{ij} \sim z_{ij} LP\left(\theta_{ij}; v_1\right) + \left(1 - z_{ij}\right) LP\left(\theta_{ij}; v_0\right) \qquad (2)$$

$$\theta_{ii} \sim Exp(\tau) \qquad (3)$$

$$z_{ij} \sim Bernoulli\left(p_{ij}\right) \qquad (4)$$

In the above, Equation 1 is the multivariate normal distribution of the Gaussian graphical model, and Equation 2 represents the "spike-and-slab" prior (Ročková and George 2018) on the off-diagonal entries of $\boldsymbol{\Theta}$. Specifically, for $1 \leq i < j \leq p$, z_{ij} is a binary edge indicator – *i.e.*, equal to 1 if there is an edge between biomolecules i and j and equal to 0 otherwise. The $LP(\cdot\,;\cdot)$ denote Laplace (aka double exponential) distributions with the probability function (PDF) $LP(\theta;v) = 1/(2v)\exp(-|\theta|/v)$ for $-\infty < \theta < \infty$ and $v > 0$. With $v_0 < v_1$, $LP(\theta;v_0)$ is the "spike" prior and $LP(\theta;v_1)$ is the "slab" prior. Continuing, Equation 3 represents a weakly informative prior on the diagonal entries of $\boldsymbol{\Theta}$, which uses the exponential distribution with PDF $f(\theta|\tau) = \tau\exp(-\tau\theta)$ for $\theta > 0$ and $\tau > 0$. Last, in Equation 4, the binary edge indicators z_{ij} follow Bernoulli distributions with probability p_{ij}. The specification of p_{ij} allows the incorporation of edge-specific prior information.

S1.2. EM algorithm

The BAGUS algorithm utilizes the EM algorithm, with the z_{ij} playing the role of "missing data", to calculate the maximum *a posteriori* estimate of $\boldsymbol{\Theta}$. We maximize the log posterior distribution function, which may derived from Equation 1 through Equation 4. On each iteration t, the E-step substitutes the z_{ij}s with their expectation and the M-step maximizes the resulting function with respect to $\boldsymbol{\Theta}$. The expectation of the binary edge indicators is

$$\omega_{ij}^{(t)} = \mathbb{E}\left(z_{ij}|\mathbf{Y};\boldsymbol{\Theta}^{(t)}\right) = \frac{p_{ij}LP\left(\theta_{ij}^{(t)};v_1\right)}{p_{ij}LP\left(\theta_{ij}^{(t)};v_1\right) + \left(1-p_{ij}\right)LP\left(\theta_{ij}^{(t)};v_0\right)} \qquad (5)$$

for $1 \leq i < j \leq p$. Note that the expectation in Equation 5 is also the edge probability $\mathbb{P}\left(z_{ij} = 1|\mathbf{Y};\boldsymbol{\Theta}^{(t)}\right)$, so this formula serves to calculate posterior edge probabilities when the algorithm has converged. The M-step then seeks to maximize

$$\sum_{s=1}^{n} \log f(\mathbf{Y}_s \Big| \boldsymbol{\Theta}) + \sum_{i=1}^{p} \tau\, \theta_{ii} + \sum_{i<j} \lambda_{ij}^{(t)} \Big| \theta_{ij}|, \qquad (6)$$

with respect to $\boldsymbol{\Theta}$, where $\lambda_{ij}^{(t)} = \omega_{ij}^{(t)}/v_1 + \left(1-\omega_{ij}^{(t)}\right)/v_0$. It should be noted that Equation 6 consists of the Gaussian loglikelihood (first term) with entry-wise ℓ_1 penalties on the entries of $\boldsymbol{\Theta}$ (last terms). This penalized loglikelihood is the basis of the graphical LASSO. (Friedman, Hastie, and Tibshirani 2008)

S1.3. The BPlane algorithm

The difference between the BAGUS and BPlane algorithms is the expression used for $f(\mathbf{y}|\boldsymbol{\Theta})$ in Equation 6. Where BAGUS utilizes the multivariate normal loglikelihood $(1/2)\,\mathrm{logdet}(\boldsymbol{\Theta}) - (1/2)\mathrm{tr}(\mathbf{S}\boldsymbol{\Theta})$, with $\mathbf{S} = (1/n)\mathbf{Y}'\mathbf{Y}$ representing the empirical covariance matrix, BPlane substitutes a modified pseudolikelihood. Focusing on a single sample $\mathbf{y} = \left(y_1,\ldots,y_p\right)$, the pseudolikelihood is defined as the product of the full conditional

distributions $\prod_{i=1}^{p} f(y_i|\mathbf{y}_{-i};\boldsymbol{\Theta})$,(Besag 1975) where $\mathbf{y}_{-i}$ is the $(p-1)$-vector consisting of the entries of $\mathbf{y}$ excepting y_i.

For $\mathbf{y} \sim \mathcal{N}_p(\mathbf{0},\boldsymbol{\Theta}^{-1})$, it is well-known that $y_i|\mathbf{y}_{-i}$ follows a normal distribution with mean $\sum_{j\neq i}\left(-\theta_{ij}/\theta_{ii}\right)y_j$ and variance $1/\theta_{ii}$. To be clear, the previous sum is over $\{j{:}1 \leq j \leq p, j \neq i\}$. Thus, the log-pseudolikelihood (up to a constant) is $\tilde{L}(\boldsymbol{\Theta}) = (1/2)\sum_{i=1}^{p} \tilde{L}_i(\boldsymbol{\Theta})$, where $\tilde{L}_i(\boldsymbol{\Theta})$ is the unit log-pseudolikelihood function

$$\tilde{L}_i(\boldsymbol{\Theta}) = \log\theta_{ii} - \theta_{ii}\left(y_i + \sum_{j\neq i}\frac{\theta_{ij}}{\theta_{ii}}y_j\right)^2. \qquad (7)$$

Note that the second term in Equation 7 resembles the objective function for least-squares regression with response variable y_i, explanatory variables $\mathbf{y}_{-i}$, and regression coefficients $\beta_j = -\theta_{ij}/\theta_{ii}$.

The objective function for the CONCORD (Khare, Oh, and Rajaratnam 2015) algorithm results from multiplying the second term in Equation 7 by an additional factor of θ_{ii}, making the penalized modified log-pseudolikelihood convex and providing attractive convergence properties. Summing over all samples and substituting into Equation 6, the function that is maximized in the M-step of BPlane is

$$n\sum_{i=1}^{p}\log\theta_{ii} - \frac{1}{2}\sum_{i=1}^{p}\sum_{s=1}^{n}\left(\theta_{ii}y_{si} + \sum_{j\neq i}\theta_{ij}\,y_{sj}\right)^2 + \tau\sum_{i=1}^{p}\theta_{ii} + \sum_{i<j}\lambda_{ij}^{(t)}\,|\theta_{ij}| \qquad (8)$$

Based on the CONCORD algorithm, the M-step updates the elements of $\boldsymbol{\Theta}$ one-at-time (updating denoted by asterisk below), fixing all other elements at their current values. First, the off-diagonal elements are updated as

$$\theta_{ij}^{*} = \frac{S_{\lambda_{ij}^{(t)}/n}\left\{-\left(\sum_{k\neq j}\theta_{ik}\,s_{jk} + \sum_{k\neq i}\theta_{jk}\,s_{ik}\right)\right\}}{s_{ii}+s_{jj}}, \qquad (9)$$

for $1 \leq i < j \leq p$, where $S_\lambda(x)$ is the soft-threshold function $S_\lambda(x) \equiv \mathrm{sgn}(x)(|x|-\lambda)_+$, with $(x)_+ = \max(x,0)$, and s_{ij} is the (i,j)th entry of $S = (1/n)\mathbf{Y}'\mathbf{Y}$. Then, the diagonal elements of $\boldsymbol{\Theta}$ are updated as

$$\theta_{ii}^{*} = \frac{-\sum_{j\neq i}\theta_{ij}\,s_{ij} + \left\{\left(\sum_{j\neq i}\theta_{ij}\,s_{ij} - \tau\right)^2 + 4s_{ii}\right\}^{1/2}}{2s_{ii}}, \qquad (10)$$

for $1 \leq i \leq p$. Since each element of $\boldsymbol{\Theta}$ is only updated once, the M-step is technically a conditional maximization (or CM-) step. (Meng and Rubin 1993)

The fact that BPlane reduces the computational complexity of the M-step from $O(p^3)$ to $O(np^2)$ when $n < p$ is based on an identity (Khare, Oh, and Rajaratnam 2015) that allows the sum $\sum_{k\neq j}\theta_{ik}s_{jk}$ over $(p-1)$ terms in Equation 9 and Equation 10 to be calculated as an inner product of n-vectors. The identity is

$$\sum_{k \neq j} \theta_{ik}\, s_{jk} = -\,\theta_{ij} s_{jj} + (1/n)\theta_{ii}\mathbf{Y}_j{}'\mathbf{r}_i, \qquad (11)$$

where the residual vector $\mathbf{r}_i = \mathbf{Y}_i + \sum_{k \neq i}(\theta_{ik}/\theta_{ii})\mathbf{Y}_k$. It should be noted that, after using Equation 11 to update θ_{ij}^* via Equation 9, the corresponding residual vectors must also be updated with the equations $\mathbf{r}_i^* = \mathbf{r}_i + \left(\theta_{ij}^* - \theta_{ij}\right)/\theta_{ii}\mathbf{Y}_j$ and $\mathbf{r}_j^* = \mathbf{r}_j + \left(\theta_{ij}^* - \theta_{ij}\right)\theta_{jj}\mathbf{Y}_i$. Further, after updating θ_{ii}^*, one must also update $\mathbf{r}_i^* = (\mathbf{r}_i - \mathbf{Y}_i)\left(\theta_{ii}/\theta_{ii}^*\right) + \mathbf{Y}_i$.

S1.4. Practical considerations

To implement BPlane in practice, we must select the hyperparameters v_0, v_1, and τ and specify initial estimates of $\boldsymbol{\Theta}$. For the former task, we recommend searching for the values that minimize the BIC-type measure (Khare, Oh, and Rajaratnam 2015)

$$\sum_{i=1}^{p}\sum_{s=1}^{n}\left(y_{si} + \sum_{j \neq i}\frac{\hat{\theta}_{ij}}{\hat{\theta}_{ii}}y_{sj}\right)^2 + \log(n)\left|\{(i,j){:}1 \leq i < j \leq p, \hat{\theta}_{ij} \neq 0\}\right| \qquad (12)$$

over a grid of (v_0,v_1) values such from $v_0 = (0.25,0.5,0.75,1.0)/\sqrt{n\log(p)}$ and $v_1 = (2.5,5.0,7.5,10.0)/\sqrt{n\log(p)}$ (Yang et al. 2021). It should be noted that the necessity of fitting the model repeatedly over candidate hyperparameter values is another compelling reason for the computational efficiency of BPlane. Also, we set $\tau = 0.5/\sqrt{n\log(p)}$.

For initial estimates, we set $\boldsymbol{\Theta}^{(0)} = \mathbf{I}_p$, the identity matrix of order p, and run a preliminary M-step with $\omega_{ij}^{(0)} = p_{ij}$ and $\mathbf{r}_i = \mathbf{Y}_i$ before proceeding with full EM iterations.

S2. More details regarding simulation experiments

For reproducibility of our experiments, we provide the following information. First, prior information was assumed to be entirely noninformative – *i.e.*, $p_{ij} = 0.5$ for all i,j. Second, the grid of (v_0,v_1) candidate values for simulated data with sample size n and p biomolecules was constructed from $v_0 = (0.25,0.5,0.75,1.0)/\sqrt{n\log(p)}$ and $v_1 = (2.5,5.0,7.5,10.0)/\sqrt{n\log(p)}$ (Yang et al. 2021). We then found the (v_0,v_1) values that minimized that BIC for the BAGUS algorithm and the BIC-type measure given by Equation 12 for BPlane.

S3. Additional edge detection properties of estimated graphs

In addition to the results presented in Figure 2, we calculated the specificity, sensitivity, Matthews correlation coefficient (MCC) (Matthews 1975), and F1 score (Taha and Allan 2015) based on thresholding the estimated edge detection probabilities to a cutoff of 0.5. The mean values of these metrics, plus or minus two standard errors, are presented in Figure S1.

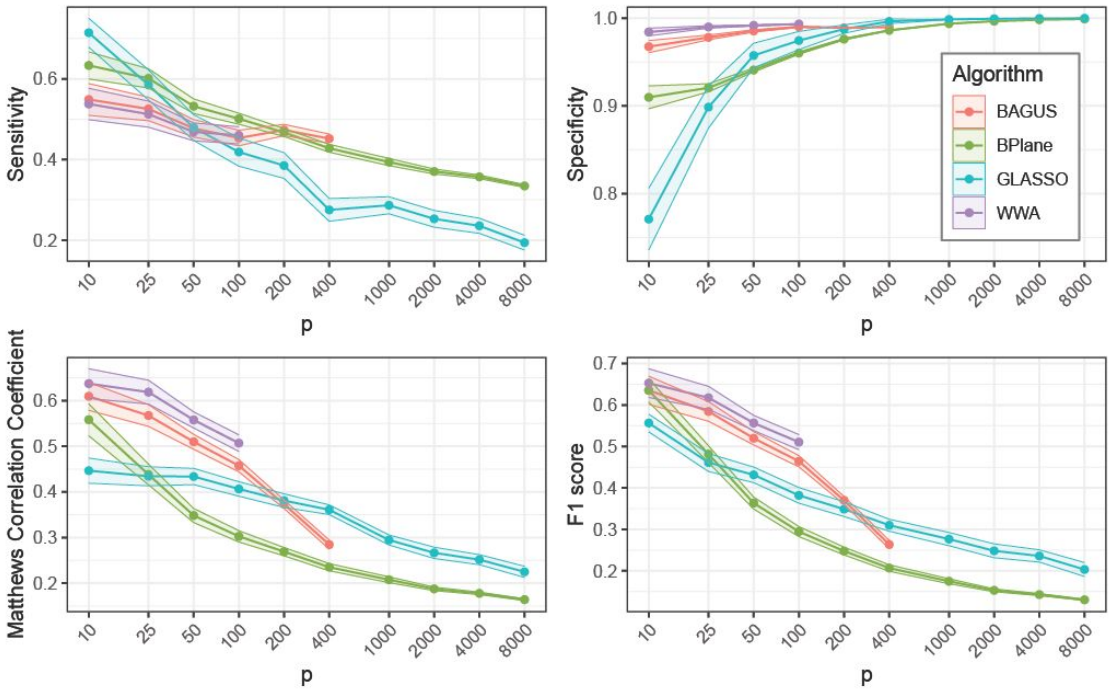

Figure S1: Sensitivity, sensitivity, Matthews correlation coefficient, and F1 scores based on simulated data.

References

Besag, J. Statiatical Analysis of Non-Lattice Data. *Journal of the Royal Statistical Society, Series D* **1975**, *24* (3): 179–95. https://doi.org/10.2307/2987782.

Friedman, J.; Hastie, T.; Tibshirani, R. Sparse Inverse Covariance Estimation with the Graphical Lasso. *Biostatistics* **2008**, *9* (3): 432–41. https://doi.org/10.1093/biostatistics/kxm045.

Gan, L.; Narisetty, N. N.; Liang, F. Regularization for Graphical Models with Unequal Shrinkage. *Journal of the American Statistical Association* **2019**, *114* (527): 1218–31. https://doi.org/10.1080/016214592018.1482755.

Khare, K.; Oh, S.-Y.; Rajaratnam, B. A Convex Pseudolikelihood Framework for High Dimensional Partial Correlation Estimation with Convergence Guarantees. *Journal of the Royal Statistical Society, Series B* **2015**, *77* (4): 803–25.

Matthews, B. W. Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme. *Biochimica Et Biophysica Acta (BBA) Protein Structure* **1975**, *405* (2): 442–51. https://doi.org/10.1016/0005-2795(75)90109-9.

Meng, X.-L.; Rubin, D. B. Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. *Biometrika* **1993**, *80* (2): 267–78. http://www.jstor.org/stable/2337198.

Ročková, V.; George, E. I. The Spike-and-Slab LASSO. *Journal of the American Statistical Association* **2018**, *113*: 431–44.

Taha, A. A.; Allan, H. Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool. *BMC Medical Imaging* **2015,** *15* (29).

Yang, X.; Gan, L.; Narisetty, N. N.; Liang, F. GemBag: Group Estimation of Multiple Bayesian Graphical Models. *Journal of Machine Learning Research* **2021**, *22*: 1–48.