8 SUPPLEMENTARY FILE

8.1 Proof of Proposition 1

The claim quickly follows from two well known linear algebra theorems: Woodbury identity and matrix determinant lemma. We recall the both results below.

Theorem 8.1 *Theorem 1. Suppose that A and C are invertible n by n matrices and U, V are n by p matrices. Then*

$$\begin{aligned} \det\left(A+UV^{\mathsf{T}}\right) &= \det\left(\mathrm{I}_p + V^{\mathsf{T}}A^{-1}U\right)\det\left(A\right) \quad \textit{and} \\ \left(A+UCV\right)^{-1} &= A^{-1} - A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1}. \end{aligned}$$

We will start with the proof of the first claim. Thanks to Theorem 1.,

$$\begin{aligned} \det\left(\overset{[\mathrm{k}]}{V}_\lambda\right) &= \det\left(\overset{[\mathrm{k}]}{W} + \overset{[\mathrm{k}]}{Z}\widetilde{Q}_\lambda^{-1}\overset{[\mathrm{k}]\mathsf{T}}{Z}\right) = \det\left(\mathrm{I}_p + \overset{[\mathrm{k}]\mathsf{T}}{Z}\overset{[\mathrm{k}]}{W}{}^{-1}\overset{[\mathrm{k}]}{Z}\widetilde{Q}_\lambda^{-1}\right)\det\left(\overset{[\mathrm{k}]}{W}\right) \\ &= \det\left(\widetilde{Q}_\lambda + \overset{[\mathrm{k}]\mathsf{T}}{Z}\overset{[\mathrm{k}]}{W}{}^{-1}\overset{[\mathrm{k}]}{Z}\right)\det\left(\widetilde{Q}_\lambda\right)^{-1}\det\left(\overset{[\mathrm{k}]}{W}\right). \end{aligned}$$

Now

$$\ln\det(\overset{[\mathrm{k}]}{V}_\lambda) = \ln\det\left(\lambda_Q Q + \lambda_R\,\mathrm{I}_p + \overset{[\mathrm{k}]}{\Omega}\right) - \ln\det\left(\lambda_Q Q + \lambda_R\,\mathrm{I}_p\right) + \ln\det\left(\overset{[\mathrm{k}]}{W}\right),$$

which finishes the proof. To show the second claim, we will rewrite $\overset{[\mathrm{k}]}{V}{}_\lambda^{-1}$ as

$$\overset{[\mathrm{k}]}{V}{}_\lambda^{-1} = \overset{[\mathrm{k}]}{W}{}^{-1} - \overset{[\mathrm{k}]}{W}{}^{-1}\overset{[\mathrm{k}]}{Z}\left(\lambda_Q Q + \lambda_R\,\mathrm{I}_p + \overset{[\mathrm{k}]}{\Omega}\right)^{-1}\overset{[\mathrm{k}]\mathsf{T}}{Z}\overset{[\mathrm{k}]}{W}{}^{-1},$$

thanks to (8.1). Therefore

$$\overset{[\mathrm{k}]}{\widetilde{y}}{}^{\mathsf{T}}\overset{[\mathrm{k}]}{V}{}_\lambda^{-1}\overset{[\mathrm{k}]}{\widetilde{y}} = -\overset{[\mathrm{k}]\mathsf{T}}{q}\left(\lambda_Q Q + \lambda_R\,\mathrm{I}_p + \overset{[\mathrm{k}]}{\Omega}\right)^{-1}\overset{[\mathrm{k}]}{q} + \overset{[\mathrm{k}]}{\widetilde{y}}{}^{\mathsf{T}}\overset{[\mathrm{k}]}{W}{}^{-1}\overset{[\mathrm{k}]}{\widetilde{y}}.$$

8.2 Gradient and Hessian for the objective in (18)

Denote by $h(\lambda_Q, \lambda_R)$ the objective function of interest, i.e.

$$h(\lambda_Q,\lambda_R) := \ln\det\left(\lambda_Q Q + \lambda_R\,\mathrm{I}_p + \Omega\right) - \ln\det\left(\lambda_Q Q + \lambda_R\,\mathrm{I}_p\right) - q^{\mathsf{T}}\left(\lambda_Q Q + \lambda_R\,\mathrm{I}_p + \Omega\right)^{-1} q,$$

where Ω and q were defined in the statement of proposition 3 ("[k]"s symbols were omitted for clarity). After using notations $D_\lambda := \big(\lambda_Q Q + \lambda_R \mathrm{I}_p + \Omega\big)^{-1}$ and $\widetilde{Q}_\lambda := \lambda_Q Q + \lambda_R \mathrm{I}_p$, this function takes the short form

$$h(\lambda_Q, \lambda_R) = \ln\det D_\lambda^{-1} - \ln\det \widetilde{Q}_\lambda - q^\mathsf{T} D_\lambda q.$$

To find the gradient and Hessian of h we will use the following well known formulas

Proposition 8.2 *Suppose that A and B are p by p, symmetric, positive semi-definite matrices, ν is p is p-dimensional vector and $tA+sB$ is positive definite. Then it holds*

- $\displaystyle \frac{\partial}{\partial t}\Big\{ \ln\det\big(tA+sB\big)\Big\} = \mathrm{tr}\Big[\big(tA+sB\big)^{-1} A\Big],$
- $\displaystyle \frac{\partial^2}{\partial t \partial s}\Big\{ \ln\det\big(tA+sB\big)\Big\} = -\,\mathrm{tr}\Big[\big(tA+sB\big)^{-1} A\,\big(tA+sB\big)^{-1} B\Big],$
- $\displaystyle \frac{\partial^2}{\partial t^2}\Big\{ \ln\det\big(tA+sB\big)\Big\} = -\,\mathrm{tr}\Big[\Big(\big(tA+sB\big)^{-1} A\Big)^2\Big],$
- $\displaystyle \frac{\partial}{\partial t}\Big\{ -\nu^\mathsf{T}\big(tA+sB\big)^{-1}\nu\Big\} = \nu^\mathsf{T}\big(tA+sB\big)^{-1} A\,\big(tA+sB\big)^{-1}\nu,$
- $$\begin{aligned}\frac{\partial^2}{\partial t\,\partial s}\Big\{ -\nu^\mathsf{T}\big(tA+sB\big)^{-1}\nu\Big\} = &-\nu^\mathsf{T}\big(tA+sB\big)^{-1} A\,\big(tA+sB\big)^{-1} B\,\big(tA+sB\big)^{-1}\nu \\ &-\nu^\mathsf{T}\big(tA+sB\big)^{-1} B\,\big(tA+sB\big)^{-1} A\,\big(tA+sB\big)^{-1}\nu,\end{aligned}$$
- $\displaystyle \frac{\partial^2}{\partial t^2}\Big\{ -\nu^\mathsf{T}\big(tA+sB\big)^{-1}\nu\Big\} = -\,2\nu^\mathsf{T}\big(tA+sB\big)^{-1} A\,\big(tA+sB\big)^{-1} A\,\big(tA+sB\big)^{-1}\nu.$

Thanks to the above, we quickly get

$$\nabla h\big|_{\lambda=\lambda_0} = \begin{bmatrix} \mathrm{tr}\big[(D_{\lambda_0} - \widetilde{Q}_{\lambda_0}^{-1})Q\big] + q^\mathsf{T} D_{\lambda_0} Q D_{\lambda_0} q \\ \mathrm{tr}\big[D_{\lambda_0} - \widetilde{Q}_{\lambda_0}^{-1}\big] + q^\mathsf{T} D_{\lambda_0}^2 q \end{bmatrix}.$$

and

$$\mathbf{H}(h)\big|_{\lambda=\lambda_0} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix},$$

where

$$\begin{aligned}\mathbf{H}_{11} &:= -\,\mathrm{tr}\big[D_{\lambda_0} Q D_{\lambda_0} Q - \widetilde{Q}_{\lambda_0}^{-1} Q \widetilde{Q}_{\lambda_0}^{-1} Q\big] - 2q^\mathsf{T} D_{\lambda_0} Q D_{\lambda_0} Q D_{\lambda_0} q,\\ \mathbf{H}_{22} &:= -\,\mathrm{tr}\big[D_{\lambda_0}^2 - \widetilde{Q}_{\lambda_0}^{-2}\big] - 2q^\mathsf{T} D_{\lambda_0}^3 q,\\ \mathbf{H}_{12} = \mathbf{H}_{21} &:= -\,\mathrm{tr}\big[(D_{\lambda_0}^2 - \widetilde{Q}_{\lambda_0}^{-2})Q\big] - q^\mathsf{T} D_{\lambda_0} Q D_{\lambda_0}^2 q - q^\mathsf{T} D_{\lambda_0}^2 Q D_{\lambda_0} q.\end{aligned}$$

8.3 Asymptotic confidence interval

We start from the optimization problem

$$\underset{B\in\mathbb{R}^{p+m}}{\operatorname{argmin}} \left\{ \underbrace{\sum\nolimits_i \psi(\mathcal{X}_i B) \,-\, y^\mathsf{T}\mathcal{X}B \,+\, \frac{1}{2}B^\mathsf{T}\mathcal{Q}B}_{\ell(B)} \right\}. \tag{21}$$

Calculating the derivatives of ℓ yields

$$\frac{\partial \ell}{\partial B}(B) = \mathcal{X}^\mathsf{T}\psi'(\mathcal{X}B) - \mathcal{X}^\mathsf{T}y + \mathcal{Q}B \quad \text{and} \quad \frac{\partial^2 \ell}{\partial B^2}(B) = \mathcal{X}^\mathsf{T}\Psi_{\mathcal{X}B}\mathcal{X} + \mathcal{Q}, \tag{22}$$

where

$$\begin{cases} \psi'(\mathcal{X}B) := \left[\, \psi'(\mathcal{X}_1 B), \ldots, \psi'(\mathcal{X}_n B) \,\right]^\mathsf{T} \\ \Psi_{\mathcal{X}B} := \operatorname{diag}\left\{\psi''(\mathcal{X}_1 B), \ldots, \psi''(\mathcal{X}_n B)\right\} \end{cases}.$$

Denote by B_T the true signal and consider the Taylor series expansion of $\frac{\partial \ell}{\partial B}$ about B_T. If we consider the value of Taylor polynomial in the solution of (21), $\hat{B}$, this yields the following expression

$$\frac{\partial \ell}{\partial B}\big(\hat{B}\big) = \frac{\partial \ell}{\partial B}\big(B_\text{T}\big) \,+\, (\hat{B} - B_\text{T})^\mathsf{T}\frac{\partial^2 \ell}{\partial B^2}\big(B_\text{T}\big) \,+\, o\big(\|\hat{B} - B_\text{T}\|_2^2\big)$$

Since the left-hand side of the above equals zero, using (22) we get the first-order approximation of $\hat{B}$

$$\begin{aligned} \hat{B} = \;& B_\text{T} - \left[\tfrac{\partial^2 \ell}{\partial B^2}\big(B_\text{T}\big)\right]^{-1} \tfrac{\partial \ell}{\partial B}\big(B_\text{T}\big) = \\ & B_\text{T} - \left[\mathcal{X}^\mathsf{T}\Psi_{\mathcal{X}B_\text{T}}\mathcal{X} + \mathcal{Q}\right]^{-1}\left[\mathcal{X}^\mathsf{T}\psi'(\mathcal{X}B_\text{T}) - \mathcal{X}^\mathsf{T}y + \mathcal{Q}B\right] = \\ & \left[\mathcal{X}^\mathsf{T}\Psi_{\mathcal{X}B_\text{T}}\mathcal{X} + \mathcal{Q}\right]^{-1}\left[\big(\mathcal{X}^\mathsf{T}\Psi_{\mathcal{X}B_\text{T}}\mathcal{X} + \mathcal{Q}\big)B_\text{T} - \mathcal{X}^\mathsf{T}\psi'(\mathcal{X}B_\text{T}) + \mathcal{X}^\mathsf{T}y - \mathcal{Q}B_\text{T}\right] = \\ & \left[\mathcal{X}^\mathsf{T}\Psi_{\mathcal{X}B_\text{T}}\mathcal{X} + \mathcal{Q}\right]^{-1}\left[\mathcal{X}^\mathsf{T}\Psi_{\mathcal{X}B_\text{T}}\mathcal{X}B_\text{T} - \mathcal{X}^\mathsf{T}\psi'(\mathcal{X}B_\text{T}) + \mathcal{X}^\mathsf{T}y\right] = \\ & \left[\mathcal{X}^\mathsf{T}\Psi_{\mathcal{X}B_\text{T}}\mathcal{X} + \mathcal{Q}\right]^{-1}\mathcal{X}^\mathsf{T}\Psi_{\mathcal{X}B_\text{T}}\mathcal{X}\hat{B}^0, \end{aligned}$$

where $\hat{B}^0 := B_\text{T} + \left[\mathcal{X}^\mathsf{T}\Psi_{\mathcal{X}B_\text{T}}\mathcal{X}\right]^{-1}\Big(\mathcal{X}^\mathsf{T}y - \mathcal{X}^\mathsf{T}\psi'(\mathcal{X}B_\text{T})\Big)$ is the first-order approximation of the generalized linear model estimate, i.e. for $\mathcal{Q} = 0$. It was shown that, under some regularity conditions, this estimate is unbiased and asymptotic normal (Fahrmeir & Kaufmann, 1985). The corresponding asymptotic variance is $\left[\mathcal{X}^{\mathsf{T}}\Psi_{\mathcal{X}B_{\mathrm{T}}}\mathcal{X}\right]^{-1}$. Consequently, the asymptotic variance , var_a, of $\hat{B}$ is given by

$$var_a(\hat{B}) = \left[\mathcal{X}^{\mathsf{T}}\Psi_{\mathcal{X}B_{\mathrm{T}}}\mathcal{X} + \mathcal{Q}\right]^{-1}\mathcal{X}^{\mathsf{T}}\Psi_{\mathcal{X}B_{\mathrm{T}}}\mathcal{X}\left[\mathcal{X}^{\mathsf{T}}\Psi_{\mathcal{X}B_{\mathrm{T}}}\mathcal{X} + \mathcal{Q}\right]^{-1}.$$

8.4 Signal patterns

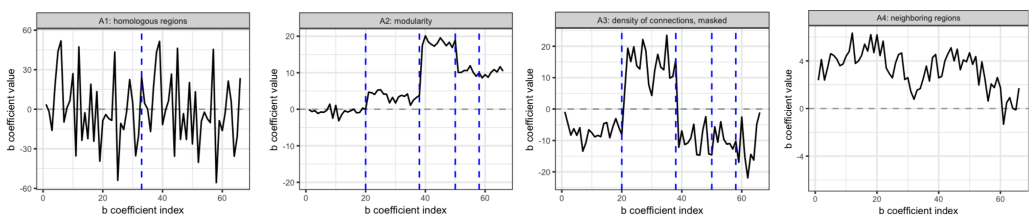

Figure 12: Examplary vectors of b model coefficients generated as $b \sim \mathcal{N}(0, \sigma_b^2(Q^{true})^{-1})$, where Q^{true} is Laplacian matrix of $\mathcal{A}^{true}$ graph adjacency matrix. Clearly, b coefficient values reflect the connectivity structure represented by $\mathcal{A}^{true}$ matrices assumed in the simulation study; left plot: $\mathcal{A}_1$ *"homologous regions"*, middle left plot: $\mathcal{A}_2$ *"modularity"*, middle right plot: $\mathcal{A}_3$ *"density of connections, masked"*, right plot: $\mathcal{A}_4$ *"neighboring regions"* (see: Fig. 2). In the left plot, vertical dashed line marks the separation between coefficients corresponding to left hemisphere brain regions and right remishpere brain regions assumed in $\mathcal{A}_1$ *"homologous regions"* construction. In the middle left plot, vertical dashed lines mark the separation between connectivity modules assumed in $\mathcal{A}_2$ *"homologous regions"* construction. In the middle right plot, vertical dashed lines mark the separation between connectivity modules assumed in $\mathcal{A}_3$ *"homologous regions"* construction.

8.5 Numerical experiments settings

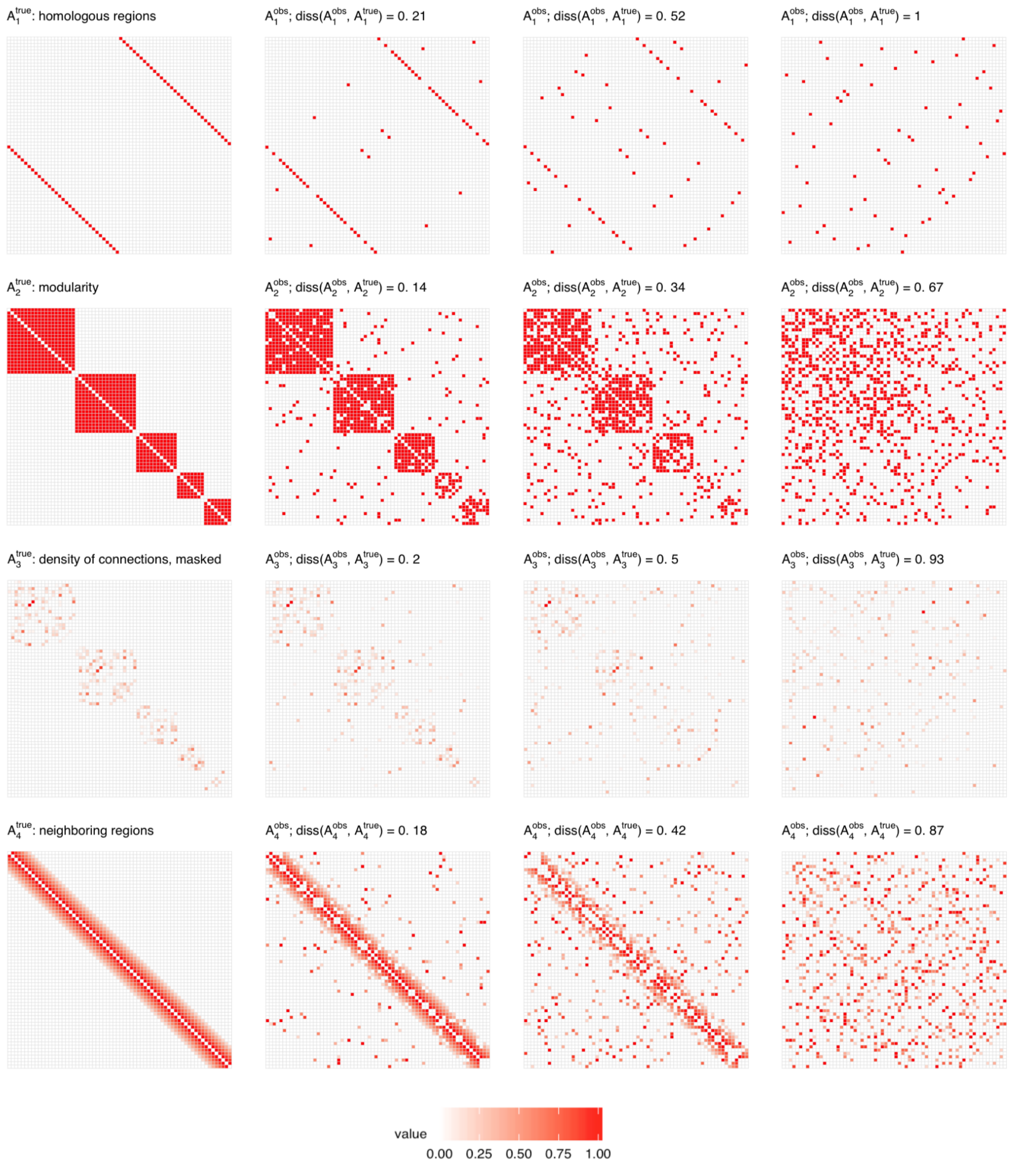

Figure 13: $\mathcal{A}^{true}$ connectivity graph adjacency matrices (1st column panels) and $\mathcal{A}^{obs}$ connectivity graph adjacency matrices (2nd-4th column panels) used in Scenario 1. $\mathcal{A}^{obs}$ matrix is constructed by randomizing $\mathcal{A}^{true}$ until a desired dissimilarity, $diss(\mathcal{A}^{obs}, \mathcal{A}^{true})$, is achieved

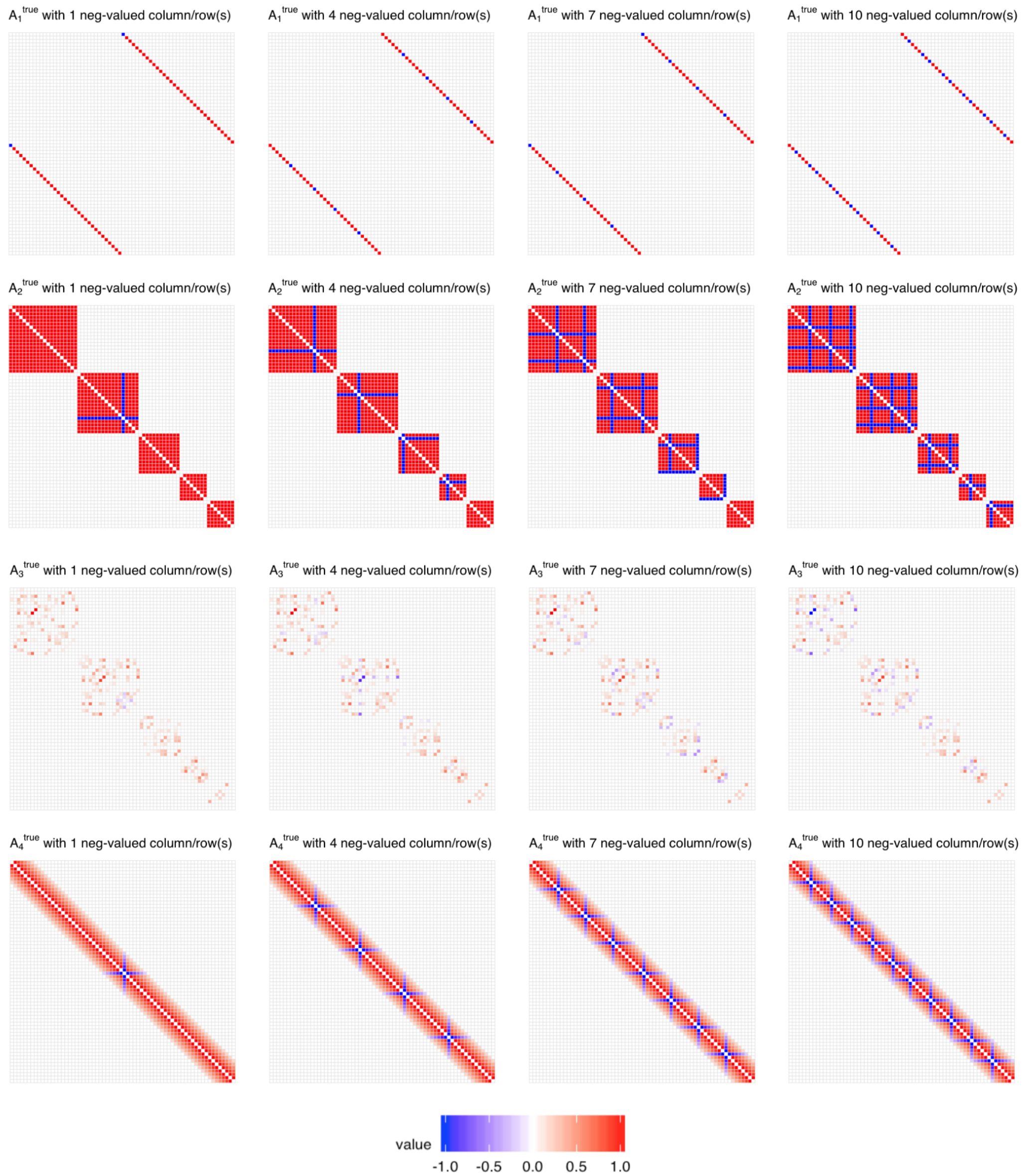

Figure 14: $\mathcal{A}^{true}$ connectivity graph adjacency matrices used in Scenario 2. $\mathcal{A}^{true}$ matrix is constructed from $\mathcal{A}_1, ..., \mathcal{A}_4$ matrices (1st-4th row panels, respectively) by changing of k, $k \in \{1, 4, 7, 10\}$, columns (and corresponding rows) of this matrix into their negative values (such structure of $\mathcal{A}^{true}$ yields the tendency that k coefficients of true signal will be separated from others.)

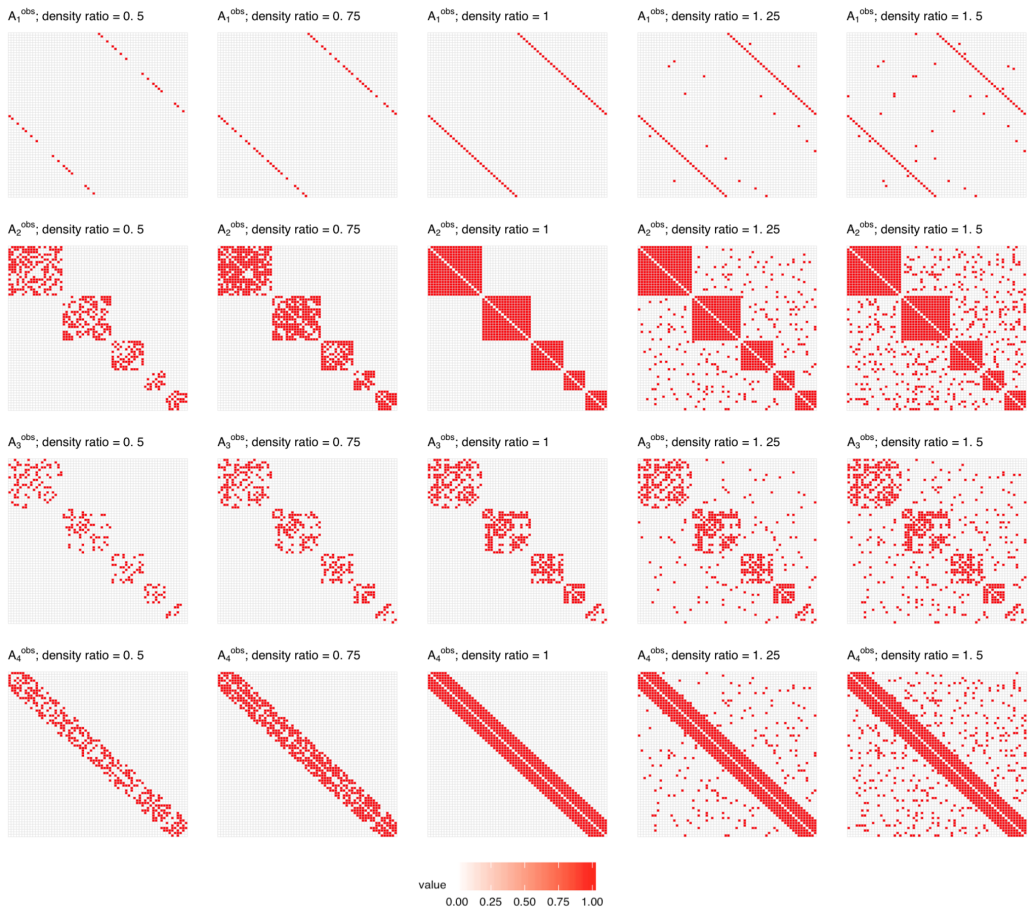

Figure 15: $\mathcal{A}^{true}$ and $\mathcal{A}^{obs}$ connectivity graph adjacency matrices used in Scenario 3. $\mathcal{A}^{true}$ matrix is defined as one of $\mathcal{A}_1, ..., \mathcal{A}_4$ matrices (1st-4th row panels, respectively). Corresponding $\mathcal{A}^{obs}$ is constructed by randomly removing or adding edges to the graph of connections represented by $\mathcal{A}^{true}$ until desired density ratio, $dens(\mathcal{A}^{obs})/dens(\mathcal{A}^{true})$, is obtained (the ratio increases from 0.5 to 1.5 from left to right in each row plot panel).

8.6 Numerical experiments results

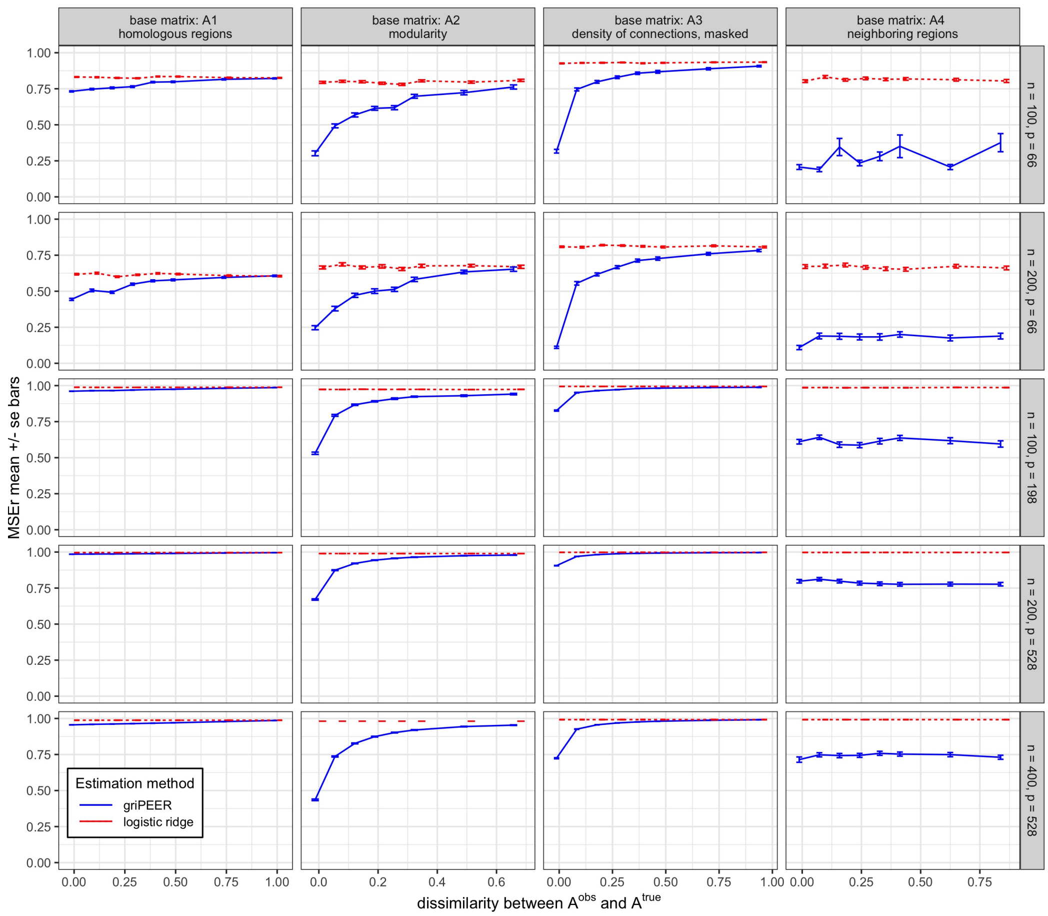

Figure 16: MSEr for estimation of b as a function of dissimilarity between $\mathcal{A}^{obs}$ and $\mathcal{A}^{true}$ as measured by $diss(\mathcal{A}^{obs}, \mathcal{A}^{true})$ (Scenario 1). Results for griPEER (blue line) and logistic ridge (red line). Presented are the average values of MSEr from 100 runs for combinations of: $n \in \{100, 200, 400\}$, $p \in \{66, 198, 528\}$. Standard errors of the mean are shown.

Table 3: Regularization parameters and execution time of numerical experiments. Presented are median values out of 100 repetitions in each of the experiment's setting for all scenario index-specific parameter values ($Param$; scenario 1: dissimilarity, scenario 2: number of connectivity matrix columns with signs switched, scenario 3: density ratio). Experiment settings are summarized by experiment scenario index (SC; ranging 1-3), base connectivity matrix ($\mathcal{A}$), number of observations (n), and number of predictors (p).

	SC	$\mathcal{A}$	n	p	griP λ_Q	griP λ_R	rid λ	Exec [s]
1	1	A1	100	66	52.9	0.2	0.3	22.8
2	1	A1	100	198	271.7	0.9	2.0	60.4
3	1	A1	200	66	0.1	0.1	0.1	12.8
4	1	A1	200	528	737.7	1.3	3.1	707.0
5	1	A1	400	528	434.7	0.5	1.2	958.5
6	1	A2	100	66	0.4	0.0	0.2	12.0
7	1	A2	100	198	2.7	0.0	1.6	25.4
8	1	A2	200	66	0.1	0.0	0.1	11.2
9	1	A2	200	528	4.3	0.1	2.4	1277.0
10	1	A2	400	528	1.8	0.0	1.0	288.0
11	1	A3	100	66	0.6	0.0	0.4	8.0
12	1	A3	100	198	2.1	0.1	1.8	21.6
13	1	A3	200	66	0.4	0.0	0.2	6.7
14	1	A3	200	528	3.1	0.1	2.5	258.5
15	1	A3	400	528	1.4	0.0	1.1	120.0
16	1	A4	100	66	2.2	0.0	0.2	24.5
17	1	A4	100	198	2.5	0.0	1.3	34.1
18	1	A4	200	66	0.9	0.0	0.1	10.3
19	1	A4	200	528	3.9	0.0	2.0	541.0
20	1	A4	400	528	1.4	0.0	0.9	354.5
21	2	A1	100	66	0.7	0.2	0.3	20.5
22	2	A2	100	66	0.4	0.0	0.2	9.7
23	2	A3	100	66	0.4	0.0	0.3	6.0
24	2	A4	100	66	1.1	0.0	0.2	3.8
25	3	A1	100	66	0.8	0.1	0.3	13.0
26	3	A2	100	66	0.8	0.0	0.3	12.0
27	3	A3	100	66	1.4	0.0	0.4	10.2
28	3	A4	100	66	0.5	0.0	0.2	13.8