

# Supplement to "Enhancing the Power of Gaussian Graphical Model Inference by Modeling the Graph Structure"

Valentin Kilian, Tabea Rebafka and Fanny Villers\*

LPSM, Sorbonne Université, Paris, France

## Abstract

This document provides additional information and details on the article "Enhancing the Power of GGM Inference by Modeling the Graph Structure". First, all mathematical details on the new greedy ICL algorithm for the NSBM are provided. Then another version of the NSBM where variances of the observations are supposed to be unknown is introduced and the associated inference algorithm is detailed. Moreover, further numerical results are presented to support the conclusions on the performance of the NSBM approach for GGM inference.

## S1 Greedy algorithm

### S1.1 Proof of Proposition 1

The expression of  $\text{ICL}^{\text{SBM}}$  in Proposition 1 of the main paper is proven in Côme and Latouche (2015). So we only prove the expression of  $\text{ICL}^{\text{noise}}(X, A, Z)$ , which corresponds to the term of the ICL related to the noise layer of the NSBM. With the notations of the main paper,

$$\begin{aligned} p(X|A, Z, \mu) &= \prod_{(i,j) \in \mathcal{A}} p(X_{i,j}|A_{i,j}, Z_i, Z_j, \mu) \\ &= \prod_{(i,j) \in \mathcal{A}} \left\{ [f_{\mathcal{N}(0,1)}(X_{i,j})]^{1-A_{i,j}} \prod_{q \leq l} [f_{\mathcal{N}(\mu_{q,l}, \sigma^2)}(X_{i,j})]^{A_{i,j} Z_{q,l}^{i,j}} \right\}. \end{aligned}$$

---

\*The authors gratefully acknowledge *please list all relevant funding sources in the unblinded version*

By integration,

$$\begin{aligned}
\text{ICL}^{\text{noise}}(X, A, Z) &= \log \left( \int p(X|A, Z, \mu) \boldsymbol{\pi}_\mu(\mu) d\mu \right) \\
&= \log \left( \prod_{(i,j) \in \mathcal{A}} [f_{\mathcal{N}(0,1)}(X_{i,j})]^{1-A_{i,j}} \times \prod_{q \leq l} \int \left\{ \prod_{(i,j) \in \mathcal{A}} [f_{\mathcal{N}(\mu_{q,l}, \sigma^2)}(X_{i,j})]^{A_{i,j} Z_{q,l}^{i,j}} \right\} f_{\mathcal{N}(\rho_0, \tau_0^2)}(\mu_{q,l}) d\mu_{q,l} \right) \\
&= \sum_{(i,j) \in \mathcal{A}} (1 - A_{i,j}) \log(f_{\mathcal{N}(0,1)}(X_{i,j})) + \sum_{q \leq l} \log(\mathcal{J}_{q,l}),
\end{aligned}$$

where

$$\sum_{(i,j) \in \mathcal{A}} (1 - A_{i,j}) \log(f_{\mathcal{N}(0,1)}(X_{i,j})) = -\frac{1}{2} \sum_{(i,j) \in \mathcal{A}} (1 - A_{i,j}) X_{i,j}^2 - \frac{N - M^A}{2} \log(2\pi),$$

and, for  $q \leq l$ ,

$$\begin{aligned}
\mathcal{J}_{q,l} &= \int \prod_{(i,j) \in I_{q,l}} f_{\mathcal{N}(\mu_{q,l}, \sigma^2)}(X_{i,j}) \times f_{\mathcal{N}(\rho_0, \tau_0^2)}(\mu_{q,l}) d\mu_{q,l} \\
&= \int \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^{n_{q,l}} \exp \left\{ - \sum_{(i,j) \in I_{q,l}} \frac{(X_{i,j} - \mu_{q,l})^2}{2\sigma^2} \right\} \times \frac{1}{\tau_0 \sqrt{2\pi}} \exp \left\{ - \frac{(\mu_{q,l} - \rho_0)^2}{2\tau_0^2} \right\} d\mu_{q,l} \\
&= \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^{n_{q,l}} \frac{1}{\tau_0 \sqrt{2\pi}} \int \exp \left\{ - \frac{1}{2} \left[ \frac{1}{\sigma^2} \sum_{(i,j) \in I_{q,l}} (X_{i,j} - \mu_{q,l})^2 + \frac{(\mu_{q,l} - \rho_0)^2}{\tau_0^2} \right] \right\} d\mu_{q,l}.
\end{aligned}$$

Developing the term in brackets, it turns out that it is equal to

$$\begin{aligned}
&\underbrace{\mu_{q,l}^2 \left( \frac{1}{\tau_0^2} + \frac{n_{q,l}}{\sigma^2} \right)}_{=a} - 2\mu_{q,l} \underbrace{\left( \frac{\rho_0}{\tau_0^2} + \frac{1}{\sigma^2} \sum_{(i,j) \in I_{q,l}} X_{i,j} \right)}_{=b} + \underbrace{\left( \frac{\rho_0^2}{\tau_0^2} + \frac{1}{\sigma^2} \sum_{(i,j) \in I_{q,l}} X_{i,j}^2 \right)}_{=c} \\
&= a \left( \mu - \frac{b}{a} \right)^2 - \frac{1}{a} (b^2 - ac).
\end{aligned}$$

Consequently, by integrating we obtain for  $\mathcal{J}_{q,l}$ ,

$$\mathcal{J}_{q,l} = \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^{n_{q,l}} \frac{1}{\tau_0 \sqrt{a}} \exp \left\{ \frac{1}{2a} (b^2 - ac) \right\}.$$

One can show that

$$b^2 - ac = -\frac{1}{\tau_0^2 \sigma^2} \sum_{(i,j) \in I_{q,l}} (X_{i,j} - \rho_0)^2 - \frac{n_{q,l}}{\sigma^4} S_{q,l}.$$

Hence,

$$\mathcal{J}_{q,l} = \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^{n_{q,l}} \left( \frac{\sigma^2 + \tau_0^2 n_{q,l}}{\sigma^2} \right)^{-\frac{1}{2}} \\ \times \exp \left\{ -\frac{1}{2(\sigma^2 + \tau_0^2 n_{q,l})} \sum_{(i,j) \in I_{q,l}} (X_{i,j} - \rho_0)^2 - \frac{\tau_0^2 n_{q,l}^2}{2(\sigma^4 + \tau_0^2 \sigma^2 n_{q,l})} S_{q,l} \right\}.$$

Then,

$$\sum_{q \leq l} \log \mathcal{J}_{q,l} = -\log \sigma \left( \sum_{q \leq l} n_{q,l} - N_Q \right) - \frac{1}{2} \log(2\pi) \sum_{q \leq l} n_{q,l} - \frac{1}{2} \sum_{q \leq l} \log(\sigma^2 + \tau_0^2 n_{q,l}) \\ - \frac{1}{2} \sum_{q \leq l} \frac{1}{\sigma^2 + \tau_0^2 n_{q,l}} \sum_{(i,j) \in I_{q,l}} (X_{i,j} - \rho_0)^2 - \frac{1}{2} \sum_{q \leq l} \frac{\tau_0^2 n_{q,l}^2}{\sigma^4 + \tau_0^2 \sigma^2 n_{q,l}} S_{q,l}.$$

By gathering all terms and noting that  $M^A = \sum_{q \leq l} n_{q,l}$ , we obtain the desired result.

## S1.2 Efficient computation of $\Delta_{g \rightarrow h}$

In the greedy algorithm the impact on the ICL of swapping a node  $i^*$  from its current block, say  $g$ , to another block  $h$  must be evaluated. That is, one has to compute

$$\Delta_{g \rightarrow h} = \text{ICL}^{\text{NSBM}}(X, Z^{\text{swap}}, A^{\text{swap}}) - \text{ICL}^{\text{NSBM}}(X, Z, A).$$

The numerical evaluation of  $\text{ICL}^{\text{NSBM}}$  has some computational cost, however the difference  $\Delta_{g \rightarrow h}$  can be computed more efficiently, as changing the block assignment of a single node has an impact only on a small number of the count statistics that appear in the expression of the  $\text{ICL}^{\text{NSBM}}$  in Proposition 1 of the main paper. For instance, the number of nodes associated with block  $q$ , that is  $n_q = \sum_{i \in [n]} Z_{i,q}$ , only changes for  $q = g$  and  $q = h$ . Let us add superscript  $\text{swap}$  to denote the quantities after the swap. As such we have  $n_g^{\text{swap}} = n_g - 1$ ,  $n_h^{\text{swap}} = n_h + 1$  and  $n_q^{\text{swap}} = n_q$  for all  $q \notin \{g, h\}$ .

For the count statistics  $n_{q,l}$ , it is clear that their value changes only if  $q \in \{g, h\}$  or  $l \in \{g, h\}$ . And as  $A$  and  $A^{\text{swap}}$  are identical except on the  $i^*$ -th row and  $i^*$ -th column, changes only come from edges involving node  $i^*$ . We introduce the sets  $I_l^A = \{j : Z_j = l, A_{i^*,j} = 1\}$  indicating the set of nodes assigned to block  $l$  to which  $i^*$  is connected in graph  $A$ . Then

one can show that  $n_{g,l}^{\text{swap}} = n_{l,g}^{\text{swap}} = n_{g,l} - |I_l^A|$  for all  $l \neq h$ ,  $n_{h,l}^{\text{swap}} = n_{l,h}^{\text{swap}} = n_{h,l} + |I_l^{A^{\text{swap}}}|$  for all  $l \neq g$ , and  $n_{g,h}^{\text{swap}} = n_{h,g}^{\text{swap}} = n_{g,h} - |I_h^A| + |I_g^{A^{\text{swap}}}|$ . If both  $q \notin \{g, h\}$  and  $l \notin \{g, h\}$ , then  $n_{q,l}^{\text{swap}} = n_{q,l}$ .

Concerning the changes in  $\bar{n}_{q,l}$ , first note that  $n_{q,l} + \bar{n}_{q,l}$  is the maximal number of possible edges between nodes in  $q$  and  $l$ , that we may denote by  $m_{q,l}$  and which can also be written as  $m_{q,l} = n_q n_l$  if  $q \neq l$  and  $m_{q,q} = n_q(n_q - 1)/2$  for  $q \in \llbracket Q \rrbracket$ . We have, for all  $l \notin \{g, h\}$ ,  $m_{g,l}^{\text{swap}} = m_{l,g}^{\text{swap}} = m_{g,l} - n_l$  and  $m_{h,l}^{\text{swap}} = m_{l,h}^{\text{swap}} = m_{h,l} + n_l$ . Moreover,  $m_{g,g}^{\text{swap}} = m_{g,g} - n_g + 1$ ,  $m_{h,h}^{\text{swap}} = m_{h,h} + n_h$  and  $m_{g,h}^{\text{swap}} = m_{h,g}^{\text{swap}} = m_{g,h} - n_h + n_g - 1$ . Then,  $\bar{n}_{q,l}^{\text{swap}} = m_{q,l}^{\text{swap}} - n_{q,l}^{\text{swap}}$ .

Likewise, only the sets  $I_{q,l}$  with  $q \in \{g, h\}$  or  $l \in \{g, h\}$  are changed by the swap. More precisely,

$$I_{q,l}^{\text{swap}} = \{I_{q,l} \setminus \{\mathbb{1}_{\{q=g\}} \odot I_l^A \cup \mathbb{1}_{\{l=g\}} \odot I_q^A\}\} \cup \{\mathbb{1}_{\{q=h\}} \odot I_l^{A^{\text{swap}}} \cup \mathbb{1}_{\{l=h\}} \odot I_q^{A^{\text{swap}}}\},$$

where  $1 \odot I = I$  and  $0 \odot I = \emptyset$  for any set  $I$ .

With these observations on the count statistics, it is clear that the computation of  $\Delta_{g \rightarrow h}$  can be simplified, since the difference of the sums taken over all  $(q, l)$  such that  $q \leq l$  in  $\text{ICL}^{\text{NSBM}}$  reduces to sums involving only quantities that change with the swap of  $i^*$  from  $g$  to  $h$ . More precisely, only the terms with pairs of indices  $(q, l)$  in the set  $\mathcal{S}_{g,h}$  are relevant, where  $\mathcal{S}_{g,h}$  is defined as

$$\mathcal{S}_{g,h} = \{(g, l) : l \in \llbracket Q \rrbracket\} \cup \{(h, l) : l \in \llbracket Q \rrbracket \setminus \{g\}\}.$$

Now, to state  $\text{ICL}^{\text{NSBM}}$  explicitly, two cases must be distinguished depending on whether the swap empties block  $g$  (that is  $i^*$  is the last node assigned to block  $g$ ) or not.

**Proposition S1.** *For  $A^{\text{swap}}$  defined in (4) in the main paper,*

$$\Delta_{g \rightarrow h} = \Delta_{g \rightarrow h}^{\text{SBM}} + \Delta_{g \rightarrow h}^{\text{noise}},$$

where  $\Delta_{g \rightarrow h}^{\text{SBM}} = \text{ICL}^{\text{SBM}}(A^{\text{swap}}, Z^{\text{swap}}) - \text{ICL}^{\text{SBM}}(A, Z)$  refers to the change of  $\text{ICL}^{\text{SBM}}$  induced by the swap and  $\Delta_{g \rightarrow h}^{\text{noise}} = \text{ICL}^{\text{noise}}(X, A^{\text{swap}}, Z^{\text{swap}}) - \text{ICL}^{\text{noise}}(X, A, Z)$  the variation of the

term related to the noise layer in the NSBM. For the form of  $\Delta_{g \rightarrow h}^{\text{SBM}}$  and  $\Delta_{g \rightarrow h}^{\text{noise}}$  two cases are to be distinguished.

**Case 1** If  $i^*$  is not the only node assigned to block  $g$  (that is  $n_g > 1$ ), then moving  $i^*$  to another block does not change the number of blocks  $Q$ . Then

$$\begin{aligned} \Delta_{g \rightarrow h}^{\text{SBM}} &= \log \frac{n_0 + n_h}{n_0 + n_g - 1} + \sum_{(q,l) \in \mathcal{S}_{g,h}} \log \frac{B(\eta_0 + n_{q,l}^{\text{swap}}, \xi_0 + \bar{n}_{q,l}^{\text{swap}})}{B(\eta_0 + n_{q,l}, \xi_0 + \bar{n}_{q,l})}, \\ \Delta_{g \rightarrow h}^{\text{noise}} &= -\frac{1}{2} \sum_{j \in [p]} X_{i^*,j}^2 (A_{i^*,j} - A_{i^*,j}^{\text{swap}}) - \sum_{(q,l) \in \mathcal{S}_{g,h}} (n_{q,l}^{\text{swap}} - n_{q,l}) \log \sigma \\ &\quad - \frac{1}{2} \sum_{(q,l) \in \mathcal{S}_{g,h}} \log \frac{\sigma^2 + \tau_0^2 n_{q,l}^{\text{swap}}}{\sigma^2 + \tau_0^2 n_{q,l}} - \frac{\tau_0^2}{2\sigma^2} \sum_{(q,l) \in \mathcal{S}_{g,h}} \left\{ \frac{(n_{q,l}^{\text{swap}})^2 S_{q,l}^{\text{swap}}}{\tau_0^2 n_{q,l}^{\text{swap}} + \sigma^2} - \frac{n_{q,l}^2 S_{q,l}}{\tau_0^2 n_{q,l} + \sigma^2} \right\} \\ &\quad - \frac{1}{2} \sum_{(q,l) \in \mathcal{S}_{g,h}} \left\{ \frac{1}{\sigma^2 + \tau_0^2 n_{q,l}^{\text{swap}}} \sum_{(i,j) \in I_{q,l}^{\text{swap}}} (X_{i,j} - \rho_0)^2 - \frac{1}{\sigma^2 + \tau_0^2 n_{q,l}} \sum_{(i,j) \in I_{q,l}} (X_{i,j} - \rho_0)^2 \right\}. \end{aligned}$$

**Case 2** If  $i^*$  is the only node assigned to block  $g$  (that is  $n_g = 1$ ), then removing  $i^*$  from  $g$ , block  $g$  disappears from the model so that  $Q^{\text{swap}} = Q - 1$  and

$$\begin{aligned} \Delta_{g \rightarrow h}^{\text{SBM}} &= \log \frac{n_0 + n_h}{n_0} + \log \left( \frac{\Gamma(Qn_0 + p)\Gamma((Q-1)n_0)}{\Gamma((Q-1)n_0 + p)\Gamma(Qn_0)} \right) + QB(\eta_0, \xi_0) \\ &\quad + \sum_{l \in [Q] \setminus \{g\}} \log \frac{B(\eta_0 + n_{h,l}^{\text{swap}}, \xi_0 + \bar{n}_{h,l}^{\text{swap}})}{B(\eta_0 + n_{h,l}, \xi_0 + \bar{n}_{h,l})} - \sum_{l \in [Q]} \log B(\eta_0 + n_{g,l}, \xi_0 + \bar{n}_{g,l}), \\ \Delta_{g \rightarrow h}^{\text{noise}} &= -\frac{1}{2} \sum_{j \in [p]} X_{i^*,j}^2 (A_{i^*,j} - A_{i^*,j}^{\text{swap}}) - \left( Q + \sum_{l \in [Q] \setminus \{g\}} (n_{h,l}^{\text{swap}} - n_{h,l}) - \sum_{l \in [Q]} n_{g,l} \right) \log \sigma \\ &\quad - \frac{1}{2} \sum_{l \in [Q] \setminus \{g\}} \log \frac{\sigma^2 + \tau_0^2 n_{h,l}^{\text{swap}}}{\sigma^2 + \tau_0^2 n_{h,l}} + \frac{1}{2} \sum_{l \in [Q]} \log(\sigma^2 + \tau_0^2 n_{g,l}) \\ &\quad - \frac{\tau_0^2}{2\sigma^2} \sum_{l \in [Q] \setminus \{g\}} \left\{ \frac{(n_{h,l}^{\text{swap}})^2 S_{h,l}^{\text{swap}}}{\tau_0^2 n_{h,l}^{\text{swap}} + \sigma^2} - \frac{n_{h,l}^2 S_{h,l}}{\tau_0^2 n_{h,l} + \sigma^2} \right\} + \frac{\tau_0^2}{2\sigma^2} \sum_{l \in [Q]} \frac{n_{g,l}^2 S_{g,l}}{\tau_0^2 n_{g,l} + \sigma^2} \\ &\quad - \frac{1}{2} \sum_{l \in [Q] \setminus \{g\}} \left\{ \frac{1}{\sigma^2 + \tau_0^2 n_{h,l}^{\text{swap}}} \sum_{(i,j) \in I_{h,l}^{\text{swap}}} (X_{i,j} - \rho_0)^2 - \frac{1}{\sigma^2 + \tau_0^2 n_{h,l}} \sum_{(i,j) \in I_{h,l}} (X_{i,j} - \rho_0)^2 \right\} \\ &\quad + \frac{1}{2} \sum_{l \in [Q]} \frac{1}{\sigma^2 + \tau_0^2 n_{g,l}} \sum_{(i,j) \in I_{g,l}} (X_{i,j} - \rho_0)^2. \end{aligned}$$

*Proof of Proposition S1.* The expressions of  $\Delta_{g \rightarrow h}^{\text{SBM}}$  are given in Côme and Latouche (2015). The formulae for  $\Delta_{g \rightarrow h}^{\text{noise}}$  are obtained by direct computation and by taking into account two particular facts. First, quantities related to pairs of blocks  $(q, l)$  are unchanged under the swap if both indices  $q$  and  $l$  are different from  $g$  and  $h$ , that is  $q, l \notin \mathcal{S}_{g,h}$ . Second, by definition,  $A^{\text{swap}}$  coincides with  $A$  for all entries not involving  $i^*$ , so that corresponding terms disappear when taking the difference. To obtain the expression in the second case, we use that  $n_{g,l}^{\text{swap}} = n_{l,g}^{\text{swap}} = 0$  for all  $l \in \llbracket Q \rrbracket$ . Then the result is straightforward.  $\square$

### S1.3 Merging entire blocks

In our numerical experiments, we observed that the solution obtained by Algorithm 1 of the main paper can be improved, by checking whether merging entire blocks increases the ICL criterion. Denote by  $\Delta_{g \cup h}$  the variation of  $\text{ICL}^{\text{NSBM}}$  when blocks  $g$  and  $h$  are merged, where without loss of generality we assume that all nodes in  $g$  are assigned to  $h$ . Then,  $\Delta_{g \cup h}$  is defined as

$$\Delta_{g \cup h} = \text{ICL}^{\text{NSBM}}(X, A^{\text{merge}}, Z^{\text{merge}}) - \text{ICL}^{\text{NSBM}}(X, A, Z),$$

where  $Z^{\text{merge}}$  coincides with  $Z$  except for all  $i$  such that  $Z_i = g$ , for which we set  $Z_i^{\text{merge}} = h$ , and  $A^{\text{merge}}$  is obtained from  $A$  by modifying only the entries  $(i, j)$  involving nodes that belong to blocks  $g$  or  $h$ . More precisely, with  $I_l = \{i : Z_i = l\}$  and  $V^{\text{merge}} = \{(i, j) : i \in I_g \cup I_h \text{ or } j \in I_g \cup I_h\} = \{(i, j) : Z_i^{\text{merge}} = h \text{ or } Z_j^{\text{merge}} = h\}$ ,

$$A_{i,j}^{\text{merge}} = \mathbb{1}\{\rho_{i,j}^{\text{merge}} > 0.5\}, \quad \forall (i, j) \in V^{\text{merge}} \quad \text{and} \quad A_{i,j}^{\text{merge}} = A_{i,j}, \quad \forall (i, j) \notin V^{\text{merge}},$$

with  $\rho_{i,j}^{\text{merge}} = \mathbb{P}_{\hat{\theta}^{\text{merge}}}(A_{i,j} = 1 \mid X, Z^{\text{merge}})$ , where  $\hat{\theta}^{\text{merge}}$  is the updated parameter estimate obtained with  $Z^{\text{merge}}$ . The expression of the variation  $\Delta_{g \rightarrow h}$  is given in Proposition S2.

The variation  $\Delta_{g \cup h}$  is evaluated for each pair of blocks  $(g, h)$ . Then the blocks  $g$  and  $h$ , for which the variation is the largest, are definitely merged. Merging blocks is repeated until no further merge increases the ICL criterion anymore. A similar merge algorithm has

---

**Algorithm 1:** Merge ICL algorithm for NSBM inference

---

**Input** : Observation  $X$ , clustering  $Z$  into  $Q$  clusters, parameter estimate  $\theta$ , edge probabilities  $\rho$

**Output:** Clustering  $Z$ , estimate  $\theta$ , edge probabilities  $\rho$

Compute  $A$  with entries  $A_{i,j} = \mathbf{1}\{\rho_{i,j} > 0.5\}$ ;

**while** *not converged* **do**

**for** *every pair of blocks*  $(g, h)$  *with*  $g < h$  **do**

        Calculate  $\Delta_{g \cup h}$  from  $X, Z, A$  and  $A^{\text{merge}}$  according to Proposition S2;

**end**

    Merge the blocks  $g$  and  $h$  that achieve the largest increase  $\Delta_{g \cup h}$ ;

    Update  $Z, Q, \theta, \rho, A$  accordingly.

**end**

---

also been proposed in Côme and Latouche (2015) for the inference in the classical SBM.

The algorithm is summarized in Algorithm S1.

We introduce notations similar to the previous algorithm with superscript  $\text{merge}$  to denote the quantities after merging blocks  $g$  and  $h$ . For the number of nodes per block, we have  $n_h^{\text{merge}} = n_h + n_g$ ,  $n_g^{\text{merge}} = 0$ , and  $n_q^{\text{merge}} = n_q$  for all  $q \notin \{g, h\}$ . For the set of edges  $(i, j)$  between blocks  $q$  and  $l$ ,  $I_{g,l}^{\text{merge}} = I_{l,g}^{\text{merge}} = \emptyset$  for  $l \in \llbracket Q \rrbracket$ ,  $I_{h,l}^{\text{merge}} = I_{l,h}^{\text{merge}} = \{(i, j) : \mathcal{Z}_{h,l}^{i,j \text{ merge}} A_{i,j}^{\text{merge}}\}$  for  $l \neq g$  and  $I_{q,l}^{\text{merge}} = I_{q,l}$  for  $q, l \notin \{g, h\}$ . For the number of edges between blocks  $q$  and  $l$ , we find  $n_{g,l}^{\text{merge}} = n_{l,g}^{\text{merge}} = 0$  for  $l \in \llbracket Q \rrbracket$ ,  $n_{h,l}^{\text{merge}} = n_{l,h}^{\text{merge}} = |I_{h,l}^{\text{merge}}|$  for  $l \neq g$  and  $n_{q,l}^{\text{merge}} = n_{q,l}$  for all other indices  $(q, l)$ . The number of possible edges between blocks  $q$  and  $l$  satisfies:  $m_{g,l}^{\text{merge}} = m_{l,g}^{\text{merge}} = 0$  for  $l \in \llbracket Q \rrbracket$ ,  $m_{h,l}^{\text{merge}} = m_{l,h}^{\text{merge}} = m_{h,l} + m_{g,l}$  for  $l \neq \{g, h\}$  and  $m_{h,h}^{\text{merge}} = m_{h,h} + m_{g,g} + m_{g,h}$ . This implies that  $\bar{n}_{g,l}^{\text{merge}} = \bar{n}_{l,g}^{\text{merge}} = 0$  for  $l \in \llbracket Q \rrbracket$ ,  $\bar{n}_{h,l}^{\text{merge}} = \bar{n}_{l,h}^{\text{merge}} = m_{h,l}^{\text{merge}} - n_{h,l}^{\text{merge}}$  for  $l \neq g$  and  $\bar{n}_{q,l}^{\text{merge}} = \bar{n}_{q,l}$  for all other indices  $(q, l)$ . Finally, we denote  $S_{h,l}^{\text{merge}}$  the empirical variance associated with  $\{X_{i,j} : (i, j) \in I_{h,l}^{\text{merge}}\}$ .

**Proposition S2.** *The variation of the ICL when blocks  $g$  and  $h$  are merged is equal to*

$$\Delta_{g \cup h} = \Delta_{g \cup h}^{\text{SBM}} + \Delta_{g \cup h}^{\text{noise}},$$

with

$$\begin{aligned} \Delta_{g \cup h}^{\text{SBM}} &= \log \frac{\Gamma(n_0 + n_h + n_g) \Gamma(Qn_0 + p) \Gamma(n_0) \Gamma((Q-1)n_0)}{\Gamma(n_0 + n_g) \Gamma(n_0 + n_h) \Gamma((Q-1)n_0 + p) \Gamma(Qn_0)} \\ &+ \sum_{l \in \llbracket Q \rrbracket \setminus \{g\}} \log \frac{B(\eta_0 + n_{h,l}^{\text{merge}}, \xi_0 + \bar{n}_{h,l}^{\text{merge}})}{B(\eta_0 + n_{h,l}, \xi_0 + \bar{n}_{h,l})} - \sum_{l \in \llbracket Q \rrbracket} \log B(\eta_0 + n_{g,l}, \xi_0 + \bar{n}_{g,l}) + QB(\eta_0, \xi_0), \end{aligned}$$

and

$$\begin{aligned} \Delta_{g \cup h}^{\text{noise}} &= -\frac{1}{2} \sum_{(i,j) \in V^{\text{merge}}} (A_{i,j} - A_{i,j}^{\text{merge}}) X_{i,j}^2 - \left( Q + \sum_{l \in \llbracket Q \rrbracket \setminus \{g\}} (n_{h,l}^{\text{merge}} - n_{h,l}) - \sum_{l \in \llbracket Q \rrbracket} n_{g,l} \right) \log \sigma \\ &- \frac{1}{2} \sum_{l \in \llbracket Q \rrbracket \setminus \{g\}} \log \frac{\sigma^2 + \tau_0^2 n_{h,l}^{\text{merge}}}{\sigma^2 + \tau_0^2 n_{h,l}} + \frac{1}{2} \sum_{l \in \llbracket Q \rrbracket} \log(\sigma^2 + \tau_0^2 n_{g,l}) \\ &- \frac{\tau_0^2}{2\sigma^2} \sum_{l \in \llbracket Q \rrbracket \setminus \{g\}} \left\{ \frac{(n_{h,l}^{\text{merge}})^2 S_{h,l}^{\text{merge}}}{\tau_0^2 n_{h,l}^{\text{merge}} + \sigma^2} - \frac{n_{h,l}^2 S_{h,l}}{\tau_0^2 n_{h,l} + \sigma^2} \right\} + \frac{\tau_0^2}{2\sigma^2} \sum_{l \in \llbracket Q \rrbracket} \frac{n_{g,l}^2 S_{g,l}}{\tau_0^2 n_{g,l} + \sigma^2} \\ &- \frac{1}{2} \sum_{l \in \llbracket Q \rrbracket \setminus \{g\}} \left\{ \frac{1}{\sigma^2 + \tau_0^2 n_{h,l}^{\text{merge}}} \sum_{(i,j) \in I_{h,l}^{\text{merge}}} (X_{i,j} - \rho_0)^2 - \frac{1}{\sigma^2 + \tau_0^2 n_{h,l}} \sum_{(i,j) \in I_{h,l}} (X_{i,j} - \rho_0)^2 \right\} \\ &+ \frac{1}{2} \sum_{l \in \llbracket Q \rrbracket} \frac{1}{\sigma^2 + \tau_0^2 n_{g,l}} \sum_{(i,j) \in I_{g,l}} (X_{i,j} - \rho_0)^2. \end{aligned}$$

## S2 NSBM with unknown variances

The test statistics proposed in Liu (2013) are asymptotically normal, but their limit variances are unknown. To deal with this case, we propose a natural extension of our model and an adaptation of the algorithm, both presented in this section.

### S2.1 Model with unknown variances

The definition of the NSBM is the same as in Section 2.1 of the main paper except the last layer, which describes the blurring mechanism. It is replaced with

$$(X_{i,j})_{(i,j) \in \mathcal{A}} \mid Z, A \sim \bigotimes_{(i,j) \in \mathcal{A}} (1 - A_{i,j}) \mathcal{N}(0, 1) + A_{i,j} \mathcal{N}(\mu_{Z_i, Z_j}, \sigma_{Z_i, Z_j}^2),$$



with additional unknown symmetric parameter matrix  $\sigma^2 = (\sigma_{q,l}^2)_{q,l} \in \mathbb{R}_+^{Q \times Q}$ . In this case the unknown model parameter is given by  $\theta = (\pi, w, \mu, \sigma^2)$ .

## S2.2 ICL criterion

To define the ICL criterion in the model with unknown variances, one has to choose an appropriate prior for  $\theta = (\pi, w, \mu, \sigma^2)$ . As previously, we use a factorized prior, that is,  $\pi_\theta(\pi, w, \mu, \sigma^2) = \pi_\pi(\pi)\pi_w(w)\pi_{\mu,\sigma^2}(\mu, \sigma^2)$ , where  $\pi_\pi$  and  $\pi_w$  are the same priors as before. For the Gaussian parameters, we assume  $\pi_{\mu,\sigma^2}(\mu, \sigma^2) = \otimes_{q \leq l} \pi_{\mu_{q,l}, \sigma_{q,l}^2}(\mu_{q,l}, \sigma_{q,l}^2)$ . In Bayesian statistics, when both the mean and the variance of a Gaussian distribution are unknown, it is common to use a normal-inverse Gamma (NIG) distribution as prior, since it is conjugate. That is,  $\pi_{\mu_{q,l}, \sigma_{q,l}^2}(m, s) = \pi_{\mu_{q,l} | \sigma_{q,l}^2}(m | s)\pi_{\sigma_{q,l}^2}(s)$ , where  $\pi_{\sigma_{q,l}^2}$  is the inverse Gamma distribution  $\text{IG}(c_0, d_0)$  with hyperparameters  $c_0, d_0$ , that is, the prior density is given by

$$\pi_{\sigma_{q,l}^2}(z) = \frac{d_0^{c_0}}{\Gamma(c_0)} z^{-c_0-1} e^{-\frac{d_0}{z}}, \quad z > 0,$$

and  $\pi_{\mu_{q,l} | \sigma_{q,l}^2}(\cdot | s)$  is the Gaussian distribution  $\mathcal{N}\left(a_0, \frac{s}{b_0}\right)$  with hyperparameters  $a_0, b_0$ . We use the standard values for the hyperparameters, which are  $a_0 = 0$  and  $b_0 = c_0 = d_0 = 1$ .

With this choice of the prior, the ICL has an analytical form stated in the following proposition.

**Proposition S3.** *The integrated complete data log-likelihood is given by*

$$\text{ICL}^{\text{NSBM}}(X, A, Z) = \log p(X, A, Z) = \text{ICL}^{\text{SBM}}(A, Z) + \text{ICL}^{\text{noise}}(X, A, Z),$$

with  $\text{ICL}^{\text{SBM}}(A, Z)$  given by (2) in the main paper and

$$\begin{aligned} \text{ICL}^{\text{noise}}(X, A, Z) &= \log p(X|A, Z) \\ &= -\frac{N}{2} \log(2\pi) - N_Q \log\left(\frac{\Gamma(c_0)}{d_0^{c_0} \sqrt{b_0}}\right) - \frac{1}{2} \sum_{(i,j) \in \mathcal{A}} (1 - A_{i,j}) X_{i,j}^2 \\ &\quad + \sum_{q \leq l} \left\{ \log \Gamma\left(c_0 + \frac{n_{q,l}}{2}\right) - \frac{1}{2} \log(b_0 + n_{q,l}) - \left(c_0 + \frac{n_{q,l}}{2}\right) \log(d_{q,l}) \right\}, \end{aligned}$$

with the notations of Proposition 1 in the main paper and

$$d_{q,l} = d_0 + \frac{n_{q,l}S_{q,l}}{2} + \frac{n_{q,l}b_0}{2(b_0 + n_{q,l})}(\bar{X}_{q,l} - a_0)^2,$$

where  $\bar{X}_{q,l}$  denotes the sample mean of the observations  $X_{i,j}$  with  $(i,j) \in I_{q,l}$ .

*Proof of Proposition S3.* Compared to the previous case, only the term  $\text{ICL}^{\text{noise}}$  changes.

We have

$$\begin{aligned} \text{ICL}^{\text{noise}}(X, A, Z) &= \log \int_{\mathbb{R}_+^{N_Q}} \int_{\mathbb{R}^{N_Q}} p(X|A, Z, \mu, \sigma^2) \boldsymbol{\pi}(\mu, \sigma^2) d\sigma^2 d\mu \\ &= \sum_{(i,j) \in \mathcal{A}} (1 - A_{i,j}) \log(f_{\mathcal{N}(0,1)}(X_{i,j})) + \sum_{q \leq l} \log(\mathcal{J}_{q,l}), \end{aligned}$$

with, for all  $q \leq l$ ,

$$\mathcal{J}_{q,l} = \int_{\mathbb{R}_+} \int_{\mathbb{R}} \left[ \prod_{(i,j) \in I_{q,l}} f_{\mathcal{N}(\mu_{q,l}, \sigma_{q,l}^2)}(X_{i,j}) \right] f_{\mathcal{N}(a_0, \sigma_{q,l}^2/b_0)}(\mu_{q,l}) f_{IG(c_0, d_0)}(\sigma_{q,l}^2) d\mu_{q,l} d\sigma_{q,l}^2.$$

By dropping subscripts  $_{q,l}$  for readability, we obtain

$$\begin{aligned} \mathcal{J}_{q,l} &= (2\pi)^{-\frac{n}{2} - \frac{1}{2}} \frac{\sqrt{b_0} d_0^{c_0}}{\Gamma(c_0)} \\ &\quad \times \int_{\mathbb{R}_+} \int_{\mathbb{R}} (\sigma^2)^{-(c_0 + \frac{n}{2} + \frac{3}{2})} e^{-\frac{d_0}{\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \left(\sum_I (X_{i,j} - \mu)^2 + b_0(\mu - a_0)^2\right)\right) d\mu d\sigma^2 \\ &= (2\pi)^{-\frac{n}{2} - \frac{1}{2}} \frac{\sqrt{b_0} d_0^{c_0}}{\Gamma(c_0)} \int_{\mathbb{R}_+} (\sigma^2)^{-c_0 - \frac{n}{2} - \frac{3}{2}} e^{-\frac{d}{\sigma^2}} \int_{\mathbb{R}} \exp\left\{-\frac{(b_0 + n)(\mu - \bar{a})^2}{2\sigma^2}\right\} d\mu d\sigma^2 \\ &= (2\pi)^{-\frac{n}{2} - \frac{1}{2}} \frac{\sqrt{b_0} d_0^{c_0}}{\Gamma(c_0)} \int_{\mathbb{R}_+} (\sigma^2)^{-c_0 - \frac{n}{2} - \frac{3}{2}} e^{-\frac{d}{\sigma^2}} \sqrt{\frac{2\pi\sigma^2}{b_0 + n}} \int_{\mathbb{R}} f_{\mathcal{N}(\bar{a}, \frac{\sigma^2}{b_0 + n})}(\mu) d\mu d\sigma^2 \\ &= (2\pi)^{-\frac{n}{2}} \sqrt{\frac{b_0}{b_0 + n}} \frac{d_0^{c_0}}{\Gamma(c_0)} \frac{\Gamma(c_0 + \frac{n}{2})}{d^{c_0 + \frac{n}{2}}} \int_{\mathbb{R}_+} f_{IG(c_0 + \frac{n}{2}, d)}(\sigma^2) d\sigma^2 \\ &= (2\pi)^{-\frac{n}{2}} \sqrt{\frac{b_0}{b_0 + n}} \frac{\Gamma(c_0 + \frac{n}{2})}{\Gamma(c_0)} \frac{d_0^{c_0}}{d^{c_0 + \frac{n}{2}}}, \end{aligned}$$

with  $\bar{a} = (n_{q,l}\bar{X}_{q,l} + a_0b_0)/(b_0 + n_{q,l})$ . Moreover,

$$\sum_{(i,j) \in \mathcal{A}} (1 - A_{i,j}) \log(f_{\mathcal{N}(0,1)}(X_{i,j})) = -\frac{1}{2} \log(2\pi) \sum_{(i,j) \in \mathcal{A}} (1 - A_{i,j}) - \frac{1}{2} \sum_{(i,j) \in \mathcal{A}} (1 - A_{i,j}) X_{i,j}^2.$$

Putting all terms together and as  $N_Q = \frac{Q(Q+1)}{2}$  and  $N = \sum_{q \leq l} n_{q,l} + \sum_{(i,j) \in \mathcal{A}} (1 - A_{i,j})$ , the result follows.  $\square$

## S2.3 Greedy ICL inference algorithm

To maximize the ICL criterion in the NSBM with unknown variances, Algorithm 1 from the main paper can still be used, but the expression of the variation  $\Delta_{g \rightarrow h}$  of the ICL when node  $i^*$  is swapped from block  $g$  to block  $h$  is slightly changes. The decomposition

$$\Delta_{g \rightarrow h} = \Delta_{g \rightarrow h}^{\text{SBM}} + \Delta_{g \rightarrow h}^{\text{noise}},$$

still holds, where  $\Delta_{g \rightarrow h}^{\text{SBM}}$  is given by Proposition S1. Also the changes in the count statistics  $n_q, n_{q,l}$  etc. are the same. The new expression of  $\Delta_{g \rightarrow h}^{\text{noise}}$  is given in the following proposition.

**Proposition S4.** *According to two situations, the following expressions hold.*

**Case 1** *If  $n_g > 1$ , that is  $i^*$  is not the only node assigned to block  $g$ , then*

$$\begin{aligned} \Delta_{g \rightarrow h}^{\text{noise}} = & -\frac{1}{2} \sum_{j \in \llbracket p \rrbracket} (A_{i^*,j} - A_{i^*,j}^{\text{swap}}) X_{i^*,j}^2 \\ & + \sum_{(q,l) \in \mathcal{S}_{q,h}} \left\{ \log \frac{\Gamma\left(c_0 + \frac{n_{q,l}^{\text{swap}}}{2}\right)}{\Gamma\left(c_0 + \frac{n_{q,l}}{2}\right)} - \frac{1}{2} \log \frac{b_0 + n_{q,l}^{\text{swap}}}{b_0 + n_{q,l}} - \log \frac{(d_{q,l}^{\text{swap}})^{c_0 + n_{q,l}^{\text{swap}}/2}}{(d_{q,l})^{c_0 + n_{q,l}/2}} \right\}. \end{aligned}$$

**Case 2** *If  $n_g = 1$ , that is, after the swap block  $g$  is empty, so that block  $g$  disappears from the model and the number of blocks diminishes by 1, that is  $Q^{\text{swap}} = Q - 1$ , then*

$$\begin{aligned} \Delta_{g \rightarrow h}^{\text{noise}} = & Q \log \frac{\Gamma(c_0)}{d_0^{c_0} \sqrt{b_0}} - \frac{1}{2} \sum_{j \in \llbracket p \rrbracket} (A_{i^*,j} - A_{i^*,j}^{\text{swap}}) X_{i^*,j}^2 \\ & + \sum_{l \in \llbracket Q \rrbracket \setminus \{g\}} \left\{ \log \frac{\Gamma\left(c_0 + \frac{n_{h,l}^{\text{swap}}}{2}\right)}{\Gamma\left(c_0 + \frac{n_{h,l}}{2}\right)} - \frac{1}{2} \log \frac{b_0 + n_{h,l}^{\text{swap}}}{b_0 + n_{h,l}} - \log \frac{(d_{h,l}^{\text{swap}})^{c_0 + n_{h,l}^{\text{swap}}/2}}{(d_{h,l})^{c_0 + n_{h,l}/2}} \right\} \\ & - \sum_{l \in \llbracket Q \rrbracket} \left\{ \log \Gamma\left(c_0 + \frac{n_{g,l}}{2}\right) - \frac{1}{2} \log(b_0 + n_{g,l}) - \left(c_0 + \frac{n_{g,l}}{2}\right) \log(d_{g,l}) \right\}. \end{aligned}$$

Finally, Algorithm 1 to merge blocks can be applied where only the expression of the term  $\Delta_{g \cup h}^{\text{noise}}$  is changed.

**Proposition S5.** *When all nodes from block  $g$  are moved to block  $h$ , then it holds*

$$\begin{aligned} \Delta_{g \cup h}^{\text{noise}} = & Q \log \frac{\Gamma(c_0)}{d_0^{c_0} \sqrt{b_0}} - \frac{1}{2} \sum_{(i,j) \in V^{\text{merge}}} (A_{i,j} - A_{i,j}^{\text{merge}}) X_{i,j}^2 \\ & + \sum_{l \in [Q] \setminus \{g\}} \left\{ \log \frac{\Gamma\left(c_0 + \frac{n_{h,l}^{\text{merge}}}{2}\right)}{\Gamma\left(c_0 + \frac{n_{h,l}}{2}\right)} - \frac{1}{2} \log \frac{b_0 + n_{h,l}^{\text{merge}}}{b_0 + n_{h,l}} - \log \frac{(d_{h,l}^{\text{merge}})^{c_0 + n_{h,l}^{\text{merge}}/2}}{(d_{h,l})^{c_0 + n_{h,l}/2}} \right\} \\ & - \sum_{l \in [Q]} \left\{ \log \Gamma\left(c_0 + \frac{n_{g,l}}{2}\right) - \frac{1}{2} \log(b_0 + n_{g,l}) - \left(c_0 + \frac{n_{g,l}}{2}\right) \log(d_{g,l}) \right\}. \end{aligned}$$

### S3 Numerical results

To complete the numerical study of Section 5.1 in the main paper, this section provides results for settings with  $p < n$  (Figure S1) and  $p = n$  (Figure S2). The `ztransform-NSBM` procedure corresponds to the NSBM approach applied to test statistics computed from the inverse of the sample covariance matrix (Anderson, 2003) and is only applicable when  $n$  is larger than  $p$ . The conclusions are roughly the same as in the more difficult setting with  $p > n$ . Namely, in general we observe a significant improvement by applying the NSBM approach over the classical multiple testing approach for all test statistics.

## References

- Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley.
- Côme, E. and P. Latouche (2015). Model selection and clustering in stochastic block models based on the exact integrated complete data likelihood. *Statistical Modelling* 15(6), 564–589.
- Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics* 41(6), 2948 – 2978.

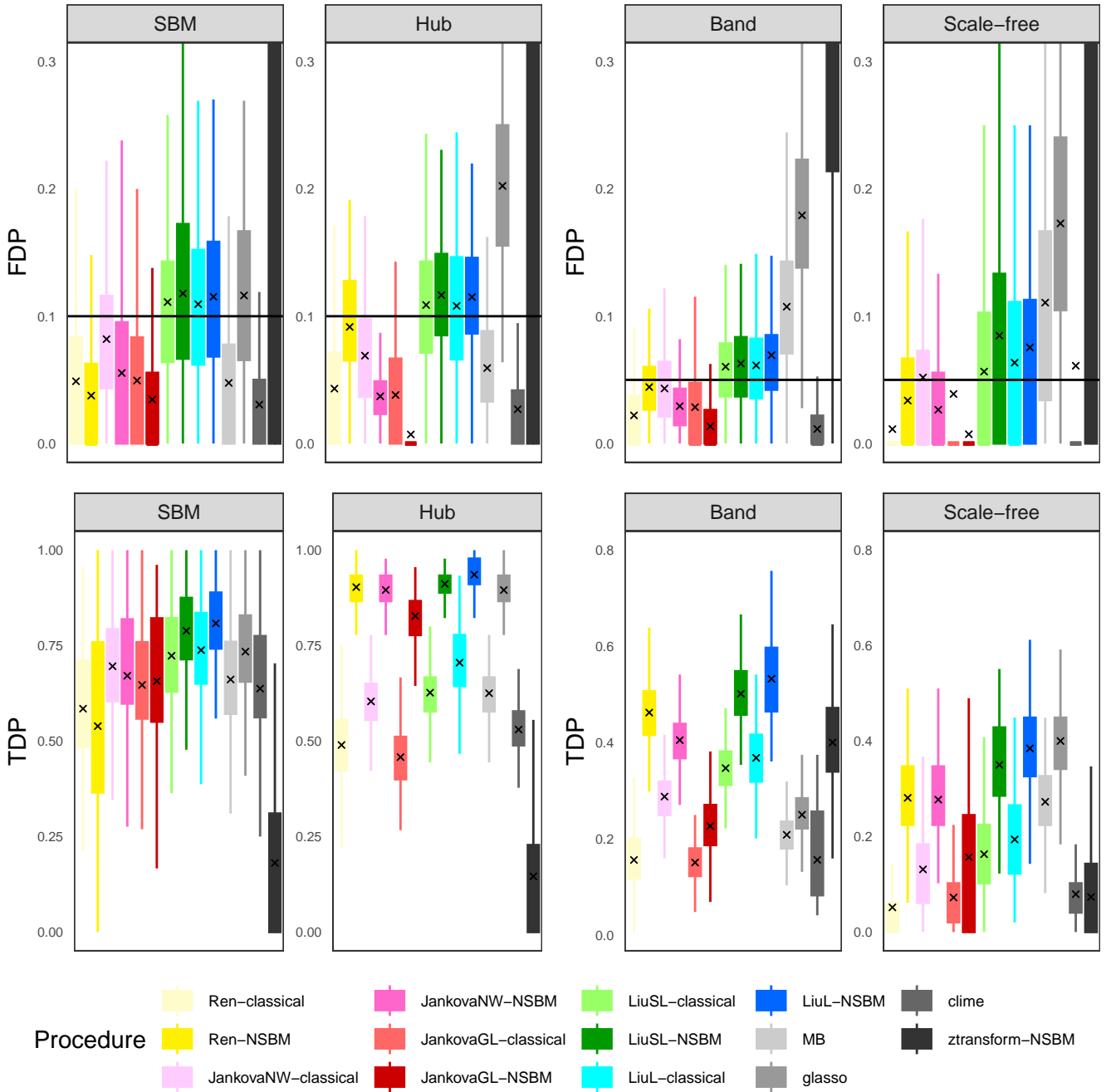


Figure S1: Boxplots of the FDP and TDP for all procedures described in Section 4.2 of the main paper for the settings SBM, hub, band and scale-free for 200 simulated datasets, where  $p = 50$  and  $n = 100$ . Horizontal lines represent the nominal level  $\alpha$ , with  $\alpha = 0.1$  in the SBM and the hub setting and  $\alpha = 0.05$  in the band and scale-free cases. The crosses in the boxplots correspond to the sample FDR (resp. TDR) defined as the mean of the FDP (resp. TDP).

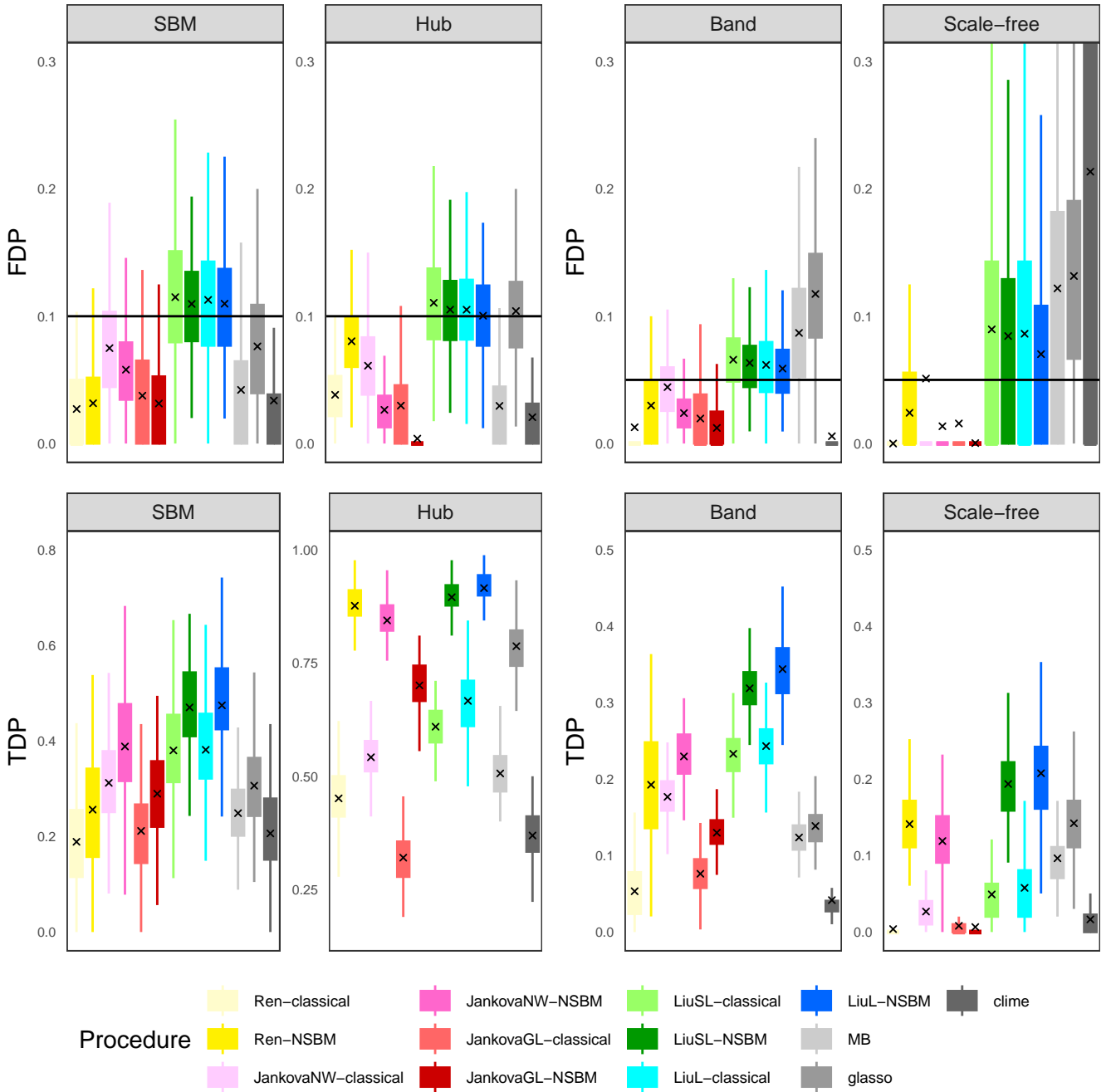


Figure S2: Boxplots of the FDP and TDP for all procedures described in Section 4.2 of the main paper for the settings SBM, hub, band and scale-free for 200 simulated datasets, where  $p = 100$  and  $n = 100$ . Horizontal lines represent the nominal level  $\alpha$ , with  $\alpha = 0.1$  in the SBM and the hub setting and  $\alpha = 0.05$  in the band and scale-free cases. The crosses in the boxplots correspond to the sample FDR (resp. TDR) defined as the mean of the FDP (resp. TDP).