**Electronic Journal of Statistics**
Vol. 16 (2022) 2796–2858
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2012

# Powerful multiple testing of paired null hypotheses using a latent graph model*

**Tabea Rebafka, Étienne Roquain and Fanny Villers**

*Laboratoire de Probabilités, Statistique et Modélisation*
*Sorbonne Université, Université de Paris & CNRS,*
*4, place Jussieu, 75005 Paris, France*
*e-mail:* tabea.rebafka@sorbonne-universite.fr
*e-mail:* etienne.roquain@sorbonne-universite.fr
*e-mail:* fanny.villers@sorbonne-universite.fr

**Abstract:** In this paper, we explore the multiple testing problem of paired null hypotheses, for which the data are collected on pairs of entities and tests have to be performed for each pair. Typically, for each pair $(i, j)$, we observe some interaction/association score between $i$ and $j$ and the aim is to detect the pairs with a significant score. In this setting, it is natural to assume that the true/false null constellation is structured according to an unobserved graph, where present edges correspond to a significant association score. The point of this work is to build an improved multiple testing decision by learning the graph structure. Our approach is in line with the seminal work of Sun and Cai [46], that uses the hidden Markov model to structure the dependencies between null hypotheses. Here, we adapt this strategy by considering the stochastic block model for the latent graph. Under appropriate assumptions, the new proposed procedure is shown to control the false discovery rate, up to remainder terms that vanish when the size of the number of hypotheses increases. The procedure is also shown to be nearly optimal in the sense that it is close to the procedure maximizing the true discovery rate. Numerical experiments reveal that our method outperforms state-of-the-art methods and is robust to model misspecification. Finally, the applicability of the new method is demonstrated on data concerning the usage of self-service bicycles in London.

**Keywords and phrases:** Multiple hypothesis testing, stochastic block model, false discovery rate, $q$-values, variational expectation-maximization algorithm.

Received October 2021.

## 1. Introduction

### *1.1. Context*

Multiple testing is a prominent research area of contemporary statistics, which is extensively used in various fields of applications, where multiple yes/no decisions have to be taken simultaneously. This methodology has essentially been

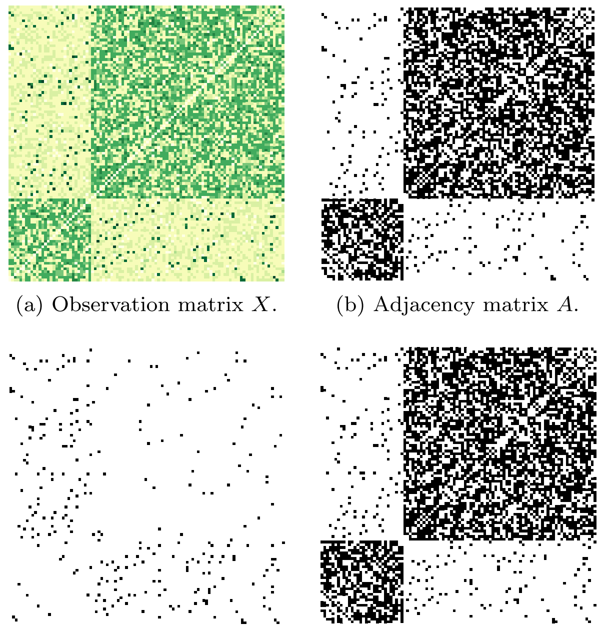
*The authors acknowledge the grants ANR-18-CE02-0010-01 (project EcoNet), ANR-16-CE40-0019 (project SansSouci) and ANR-17-CE40-0001 (BASICS) of the French National Research Agency ANR and the GDR ISIS through the “projets exploratoires” program (project TASTY).

developed for data collected per entity (e.g., genes or SNP in molecular biology or voxels in neuroimaging), that is, when data is structured in vectors. However, in a large variety of domains such as social, biological or information sciences, data are association/interaction scores collected on *pairs* of entities, and thus have a *matrix* form, potentially with large dimension. To mention only a few examples, these data can be proximity measurements based on mobile phones for contact tracing during an epidemic or studying social ties between humans [21]. They can also represent interactions between animals [28] in ecology, or any pairwise measurements according to one or two categorical variables, as the number of international migrants by country of origin and country of destination [34], or the journey counts between pairs of bicycle stations (the latter being our specific worked-out application).

While the case of vector-based multiple testing inference is ubiquitous, matrix-based datasets are far less understood in the statistical literature. To the best of our knowledge, it has only been studied when the matrix is built upon pairwise comparisons between coordinates of the same observed vector, as it is the case with marginal or partial correlations, see [20, 27, 12] among others. In particular, the works [27, 12] focus on applying a variation of the well-known Benjamini-Hochberg procedure in this context [4], and provide the control of the false discovery rate under a sparsity assumption. However, it does not take into account the structural information in the data.

In this paper, we depart from this setting by assuming that the data are directly collected in a matrix-wise fashion and we incorporate the structural information in our inference. For this, we follow the line of research based on the classical two-group mixture model introduced in [23]. The seminal works [22, 45] show how to control the false discovery rate while improving on Benjamini-Hochberg by consistently estimating the signal proportion, the null and alternative distributions. Further significant power enhancement can be obtained by incorporating some latent structure in the model, see [46] for group structure and [13, 26] for Markov structure. Here, we adapt these methods for the case of a latent graph model.

More precisely, the aim is to perform tests simultaneously for all entries of the data matrix, say $X$, and thus to infer the binary matrix, say $A$, for which $A_{i,j} = 1$ when the null hypothesis on pair $(i, j)$ is false. The point of our work is to consider this binary matrix $A$ as the adjacency matrix of a latent graph, where nodes correspond to the entities and present edges to false null hypotheses. Now, our approach consists in learning a statistical model for the latent graph and incorporate this information in the testing procedure. For this, we should model the connecting behaviors of the nodes, that is, the way a null hypothesis is likely to be true or not. This is done with the popular stochastic block model [24, 37, 40]. In a nutshell, given a clustering of the nodes, it assumes that the connection probability between two nodes depends on their respective cluster memberships. We refer the reader to [32] for an introduction to the stochastic block model and an overview of its numerous variants. Once $A$ is determined, the observation $X$, containing the paired observations, is modeled as a perturbation or noisy version of $A$. With this graph modeling, our paired



(a) Observation matrix $X$. (b) Adjacency matrix $A$.

(c) BH procedure. (d) New procedure.

Fig 1. *(a) Real-valued observation matrix $X$ with block structure. (b) Adjacency matrix $A$ coding for the true (white) and false (black) null hypotheses. (c) Rejections of the Benjamini-Hochberg procedure. (d) Rejections of our new procedure based on a latent graph.*

multiple testing problem is coupled with a dimension reduction of the true/false null constellation, that is, in the network, individuals are clustered according to their ability to be connected. Doing so, the detection performance of our procedure is drastically increased.

### *1.2. Intuition of the new method*

To illustrate our approach, consider the toy example in Figure 1. The observed symmetric matrix $X$ in (a) results from a perturbation of the binary matrix $A$ in (b), where black entries represent false null hypotheses. Nodes are ordered according to a partition of the nodes into two groups such that the underlying block structure appears. We observe that true null hypotheses give rise to pure noise in $X$ (light-green points in (a)) and false null hypotheses result in signal (green to dark-green points). We note that the signal strength depends on the group memberships of the nodes, that is, intermediate (green) signal is observed for intra-group edges, whereas the strongest signals occur for the few inter-group edges. The classical Benjamini-Hochberg procedure [4] (at level 5%), that works with a common threshold for all observations, recovers only a small part of the signal (134 rejections among 1793 false nulls in this simulated data) and thus is overly conservative, see (c). By contrast, with the new method, the threshold is locally adapted to each pair of entities using information provided by the learned graph topology. As one can see from (d), the new approach (at the same level 5%) recovers the signal almost perfectly (1793 correct rejections and 10 false discoveries).

Extensive numerical experiments show that similar results are obtained in numerous settings. In particular, it will be shown that our procedure is more efficient and powerful than other state-of-the-art methods.

### *1.3. Contributions and organization of the paper*

The contributions of this paper are as follows. First, we introduce a new setting for multiple testing of paired null hypotheses via a random graph model, called the noisy stochastic block model (Section 2). Second, model identifiability is established and a variational expectation-maximization algorithm for parameter estimation is developed (Section 3). Third, the estimation algorithm provides for each pair estimates of the posterior probability that the null hypothesis is true, also called $\ell$-values (or local fdr values [23]). Based on these $\ell$-values we propose a new multiple testing procedure relying on the general $q$-value approach (Section 4). Fourth, under mild model assumptions, the new procedure is shown to control the false discovery rate, up to a small remainder term. Moreover, the procedure is nearly optimal in the sense that it is close to the method maximizing the true discovery rate (Section 5). We underline that these theoretical results are non-asymptotic with respect to the number of tests, which is new to our knowledge compared to existing multiple testing literature in mixture models. Next, numerical experiments support the validity of our approach and demonstrate its robustness with respect to model assumptions (Section 6). The applicability of the new method is illustrated on data on the usage of self-service bikes in London (Section 7). Finally, detailed proofs and auxiliary results are deferred to the Appendices. The R code is available on CRAN via the package `noisySBM`.

### *1.4. Related works*

As the goal is to infer $A$ from its noisy version $X$, our multiple testing problem can be seen as a network inference problem. There is a rich literature dealing with network reconstruction, accounting for the uncertainty of available data in network analysis. For instance, in [17, 35, 36, 25, 39, 50] the uncertainty comes from a binary blurring mechanism of the underlying true network, that erroneously removes or adds edges according to some probabilities. In [30], uncertainty of edge presence delivers data under the form of connection probabilities and the authors use a modeling based on beta distributions. While these models are similar in spirit, the derived methods are markedly different, because our purpose is to control the false discovery rate (average proportion of erroneous discovered egdes), which has been ignored so far in this line of research to our knowledge.

Next, the new noisy stochastic block model is related to the weighted stochastic block model, first introduced in [29] (for which the weights come from a parametric exponential family) and also considered in [1, 38] (including a non-zero value only between connecting nodes). Our model can be seen as an instance of the weighted stochastic block model, with specific mixture distributions for the distribution of the edge weights. However, it additionally incorporates the modelling of true and false null hypotheses, which is not the case in the general weighted stochastic block model. This is crucial here because our primary interest is testing and not modelling.

## 2. Multiple testing framework for pairs

In this section, we present a new framework for multiple testing of paired null hypotheses. For this, the observed data matrix is viewed as a perturbation of a latent binary graph, where absent edges produce pure noise, while present edges generate a "signal plus noise" measure. The latent graph codes for the true/false constellation of null hypotheses and is assumed to follow a stochastic block model. Together, these two layers define the full model, which we call the noisy stochastic block model.

### 2.1. Setting

Let $n \geq 2$ be the number of entities or individuals in the observed population. We denote by $\mathcal{A} = \{(i,j) \ : \ 1 \leq i < j \leq n\}$ the set of all possible pairs that we would like to test. We observe $X = (X_{i,j})_{(i,j)\in\mathcal{A}}$ a real-valued upper-diagonal matrix, for which each $X_{i,j} \in \mathbb{R}$ corresponds to a measurement for the pair $(i,j) \in \mathcal{A}$, which typically is an association score between $i$ and $j$. We assume that a null hypothesis and an alternative is given for each pair, and we record the trueness/falseness of these null hypotheses in an unobserved binary upper-diagonal matrix $A = (A_{i,j})_{(i,j)\in\mathcal{A}}$, for which $A_{i,j} = 0$ if and only if the null hypothesis is true for the pair $(i,j)$. Specifically, this corresponds to the multiple testing setting where we test

$$H_{0,i,j} : \text{“}A_{i,j} = 0\text{” versus } H_{1,i,j} : \text{“}A_{i,j} = 1\text{”},$$

simultaneously for all $(i,j) \in \mathcal{A}$. To define a multiple testing model, we now specify the distribution of $X$ given $A$. Our modeling, to be presented in detail in the next section, will assume that:

- $A$ is structured according to a latent graph, itself built on a clustering $Z = (Z_i)_{1\leq i\leq n}$ of the entities. Here, the $Z_i$'s are categorical variables representing cluster memberships of the individuals;
- $(X_{i,j})_{(i,j)\in\mathcal{A}}$ are independent conditionally on $A, Z$;

- the distribution of $X_{i,j}$ conditionally on $A, Z$ is driven by some common null density if $A_{i,j} = 0$, and by some specific alternative density depending on $Z_i, Z_j$ if $A_{i,j} = 1$.

Hence, while our multiple testing model is intrinsically based on an independence assumption, the null hypotheses are dependent, which induce an (unconditional) dependence between the measurements $X_{i,j}$. This is in line with the series of works [46, 13, 26] for which a latent structure is given for the nulls. This imposed structure will be crucial to build an efficient multiple testing procedure recovering $A$ from $X$.

### *2.2. Noisy stochastic block model*

Recall that the unobserved binary matrix $A = (A_{i,j})_{(i,j)\in\mathcal{A}}$ carries the true/false null hypothesis constellation in that $A_{i,j} = 0$ if and only if the null hypothesis is true for the pair $(i, j)$. Here, we interpret the matrix $A$ as the adjacency matrix of an undirected graph without self-loops, where nodes correspond to entities $i \in \{1, \dots, n\}$ and where there is an edge between two nodes $i$ and $j$ if and only if $A_{i,j} = 1$, that is, if the null hypothesis is false for the pair $(i, j)$. Now, we can borrow the classical, and widely used, *stochastic block model* to define a latent structure on the matrix $A$.

Namely, we assume that the nodes are randomly partitioned into $Q \geq 2$ groups. Let $Z = (Z_1, \dots, Z_n)$ be the block memberships of the nodes and $\pi = (\pi_q)_{1\leq q\leq Q} \in (0, 1)^Q$ with $\sum_{q=1}^{Q} \pi_q = 1$ the group probabilities. That is, $(Z_i)_{1\leq i\leq n}$ are independent and $\mathbb{P}(Z_i = q) = \pi_q, 1 \leq q \leq Q, i = 1, \dots, n$. Then, conditionally on $Z$, the variables $A_{i,j}$, $(i, j) \in \mathcal{A}$, are independent Bernoulli variables with parameter $w_{Z_i,Z_j}$, for a connectivity parameter $w = (w_{q,\ell})_{1\leq q,\ell\leq Q} \in (0, 1)^{Q\times Q}$. As the graph is undirected, $w$ is symmetric.

Now, for the distribution of the observations $X_{i,j}$ we introduce two parametric distribution families, $\{g_{0,u}, u \in \mathcal{V}_0\}$ and $\{g_u, u \in \mathcal{V}\}$. Both families are parametric with $\mathcal{V}_0 \subset \mathbb{R}^{d_0}$ and $\mathcal{V} \subset \mathbb{R}^{d_1}$, where $d_0$ and $d_1$ are the dimensions of the parameter spaces. The relation between $A$ and the observation $X$ is that missing edges ($A_{i,j} = 0$) are replaced by pure random noise, modeled by the null density $g_{0,\nu_0}$ for some $\nu_0 \in \mathcal{V}_0$, whereas in place of present edges ($A_{i,j} = 1$) there is a signal, modeled by an alternative density $g_{\nu_{Z_i,Z_j}}$ with parameters $\nu_{q,\ell} \in \mathcal{V}$. The latter depends on the block membership of the interacting nodes in the underlying stochastic block model, such that the signal strength can be modulated locally.

The unknown global model parameter is $\theta = (\pi, w, \nu_0, \nu)$, where $\pi$ and $w$ come from the stochastic block model, $\nu_0$ denotes the null parameter and $\nu = (\nu_{q,\ell})_{1\leq q,\ell\leq Q} \in \mathcal{V}^{Q\times Q}$ denotes the parameters of the signals. Again, since the graph is undirected, $\nu$ is symmetric. The parameter space is denoted by $\Theta$ and can be used to define further restrictions on $\theta$. The distribution of $(X, A, Z)$ in the noisy stochastic block model is denoted by $P_\theta$. This notation omits the external parameters $n$ and $Q$ for the sake of simplicity.

**Algorithm 1:** Data generation in the noisy stochastic block model

**Input**: Parameter $\theta = (\pi, w, \nu_0, \nu)$, number $n$ of entities.
1. Draw $Z_i \sim \mathcal{M}(1, \pi), 1 \leq i \leq n$ independently.
2. Conditionally on $Z = (Z_1, \dots, Z_n)$, generate independently

$$A_{i,j} \,|Z \;\sim\; \mathcal{B}(w_{Z_i,Z_j}), \quad (i,j) \in \mathcal{A}.$$

3. Conditionally on $Z$ and $A = (A_{i,j})_{(i,j)\in\mathcal{A}}$, generate independently

$$X_{i,j} \mid Z, A \;\sim\; (1 - A_{i,j})g_{0,\nu_0} + A_{i,j}g_{\nu_{Z_i,Z_j}}, \quad (i,j) \in \mathcal{A}.$$

**Output**: Observation matrix $X$, latent graph $A$, latent block memberships $Z$.

Algorithm 1 summarizes the data generation process in the noisy stochastic block model. Note that in case that $X$ is a non-symmetric $n \times n$ matrix, the noisy stochastic block model is easily adapted by using directed graphs, that is, $A$ is not symmetric and $\mathcal{A} = \{(i,j) \,:\, 1 \leq i,j \leq n\}$. Also, non-zero diagonal elements can be included by admitting self-loops in the stochastic block model.

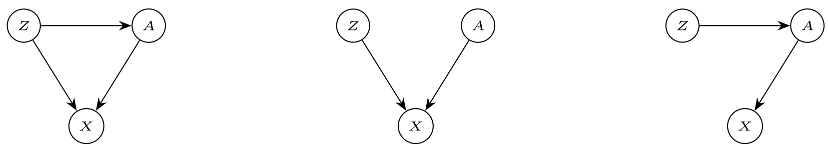
An inspection of the noisy stochastic block model reveals that there are two levels where the block memberships influence the distribution of the observed graph $X$. First, the latent graph $A$ may have block structure due to distinct Bernoulli parameters $w_{q,\ell}$. Second, the distribution of the observed values $X_{i,j}$ depends on the block memberships, see Figure 2 (a). However, it is possible to consider settings where either all Bernoulli parameters $w_{q,\ell}$ are equal, see (b), or all distributions under the alternative are equal, see (c). Then the latent block structure has an impact only on one of the random layers. Therefore, the noisy stochastic block model is a flexible model that can capture a large variety of situations.

In this paper, the leading example is the Gaussian case. It is particularly suitable when the observations $X_{i,j}$ correspond to real-valued test statistics that are known to be approximately Gaussian (e.g., asymptotically in a given parameter). The *Gaussian noisy stochastic block model* corresponds to the following choice of the parametric density families:

$$\{g_{0,u}, u \in \mathcal{V}_0\} = \{\mathcal{N}(0, \sigma_0^2), \sigma_0 > 0\}, \quad \{g_u, u \in \mathcal{V}\} = \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma > 0\}. \tag{1}$$

### *2.3. Multiple testing criteria*

Let us recall the classical criteria used in multiple testing, that we will also use in our paired framework. First, a multiple testing procedure is any measurable function $\varphi(X) \in \{0,1\}^{\mathcal{A}}$ with the convention that $\varphi_{i,j}(X) = 1$ if and only if the null-hypothesis on $(i,j)$ is rejected. Then the *false discovery rate* (FDR) of a given multiple testing procedure $\varphi$ is the average proportion of errors among




(a) General NSBM. (b) No block structure of $A$. (c) Single alternative distribution.

FIG 2. *Dependency structure in the general noisy stochastic block model (a) and special cases of the model with $w_{q,\ell}$ all equal (b) and $\nu_{q,\ell}$ all equal (c).*

the discoveries, defined as

$$\mathrm{FDR}_\theta(\varphi) = \mathbb{E}_\theta\left[\frac{\sum_{(i,j)\in\mathcal{A}}(1-A_{i,j})\varphi_{i,j}(X)}{\left(\sum_{(i,j)\in\mathcal{A}}\varphi_{i,j}(X)\right)\vee 1}\right], \tag{2}$$

where $\mathbb{E}_\theta$ refers to the expectation in our noisy stochastic block model with parameter $\theta$. Moreover, the power of $\varphi$ is defined as the *true discovery rate* (TDR), that is, the ratio of the average number of true discoveries to the average total number of alternatives given by

$$\mathrm{TDR}_\theta(\varphi) = \frac{\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} A_{i,j}\varphi_{i,j}(X)\right]}{\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} A_{i,j}\right]}. \tag{3}$$

Finally, we also introduce the *marginal false discovery rate* (MFDR) defined as

$$\mathrm{MFDR}_\theta(\varphi) = \frac{\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}}(1-A_{i,j})\varphi_{i,j}(X)\right]}{\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}}\varphi_{i,j}(X)\right]}, \tag{4}$$

with the convention $0/0 = 0$. The MFDR is a handy substitute for the FDR, because it involves the ratio of the expectations rather than the expectation of the ratio. Both quantities, the FDR and the MFDR, are close when the numerator and denominator concentrate around their respective expectation.

## 3. Estimation in noisy stochastic block model

This section presents materials for inference in the noisy stochastic block model: identifiability properties are given in Section 3.1 and parameter estimators are derived in Section 3.2.

### *3.1. Identifiability*

To make sure that the parametrization of our model is suitable, we show that the model is identifiable under mild assumptions.

**Assumption 1.** Any finite mixture of elements of $\{g_u, u \in \mathcal{V}\}$ is identifiable up to label swapping.

For instance, the family of Gaussian densities satisfies Assumption 1, see [48]. General conditions on the identifiability of mixtures of a one-parameter family of distributions are provided in [47].

**Theorem 2.** *Let $n \geq 3$ and $Q \geq 2$. Let Assumption 1 be satisfied. Let $\{g_{0,u}, u \in \mathcal{V}_0\}$ be a subset of $\{g_u, u \in \mathcal{V}\}$. Assume that parameters $\nu_0 \in \mathcal{V}_0 \subset \mathcal{V}$ and $\nu_{q,\ell} \in \mathcal{V}$ for $q \leq \ell$ are all distinct. Then all parameters of the associated noisy stochastic block model are identifiable up to label swapping, that is, up to permutation of the block membership labels $\{1, \dots, Q\}$.*

Our result is a generalization of Theorem 12 in [2] that states identifiability of the parametric random graph mixture model with weighted edges. Compared to their model, the noisy stochastic block model replaces the point mass in 0 by the null distribution $g_{0,\nu_0}$ with unknown parameter $\nu_0$. A detailed proof of Theorem 2 is provided in Appendix A.1. As a corollary of the theorem, the *Gaussian* noisy stochastic block model is identifiable under the constraint that all elements of $\{(0, \sigma_0), (\mu_{q,\ell}, \sigma_{q,\ell}), 1 \leq q \leq \ell \leq Q\}$ are distinct.

In general, the constraint that all parameters $\nu_0$ and $\nu_{q,\ell}$ are pairwise distinct is not a necessary condition. It is for instance possible to show identifiability also in an affiliation-type configuration of the noisy stochastic block model, see Appendix A.2 for details.

In the literature, different approaches exist to prove identifiability of stochastic block models under different assumptions. In [16], for instance, it is only required that the connectivity matrix $w$ has distinct rows. However, the algebraic arguments used in the proof are not easily adapted to the noisy stochastic block model. Hence, we have chosen to follow the approach by [2] as it adapts to our model in a quite natural way.

### *3.2. Parameter estimation and node clustering*

As in most latent variable models, parameter estimation is a difficult task in the noisy stochastic block model. Indeed, the complete-data likelihood function $\theta \mapsto \mathcal{L}(X, A, Z; \theta)$ has a simple expression, and the observed likelihood function $\theta \mapsto \mathcal{L}(X; \theta)$ is then obtained by integrating over all possible configurations of the latent variables $(A, Z) \in \{0,1\}^{\mathcal{A}} \times \{1, \dots, Q\}^n$. This is prohibitive for any reasonable values of $n$ and $Q$ due to the size of the latter set. As a consequence, the maximum likelihood estimator cannot be computed by directly maximizing the observed likelihood function, but an approximation can be obtained by an EM-type algorithm.

We propose an estimation algorithm, that is similar to the variational EM algorithm for the standard stochastic block model [19], which, in addition to parameter estimates, provides estimates of the group memberships. In this section, only the main ideas of the algorithm are presented. A full description is provided in Appendix B. The corresponding R code is available in the package `noisySBM` on CRAN.

The classical EM algorithm alternates until convergence an M-step, that updates the parameter estimate $\hat{\theta}$, with an E-step, that determines the posterior distribution $P_{A,Z|X;\hat{\theta}}$ of the latent variables $(A, Z)$ under the current value $\hat{\theta}$. As in the standard stochastic block model, the latter step is intractable due to the involved dependence structure of the model, but a mean-field approximation of $P_{A,Z|X;\hat{\theta}}$ by a factorized probability distribution, say $\tilde{P}_\tau$, depending on some parameter $\tau$, can be used. Thus, the variational E-step consists in searching the best variational parameter $\hat{\tau}$ that minimizes the Kullback-Leibler divergence between $\tilde{P}_\tau$ and $P_{A,Z|X;\hat{\theta}}$. In practice, this amounts to solve a fixed point equation numerically. Interestingly, in addition to $\hat{\theta}$, the variational parameter $\hat{\tau}$ obtained at the end of the variational EM algorithm provides a meaningful node clustering.

In applications, the number of latent blocks $Q$ is generally unknown, but can be estimated from the data via an adaptation of the classical integrated classification likelihood (ICL) approach [8]. It relies on a penalized observed likelihood criterion, where the penalty involves the traditional BIC penalty and the entropy of the latent variable distribution.

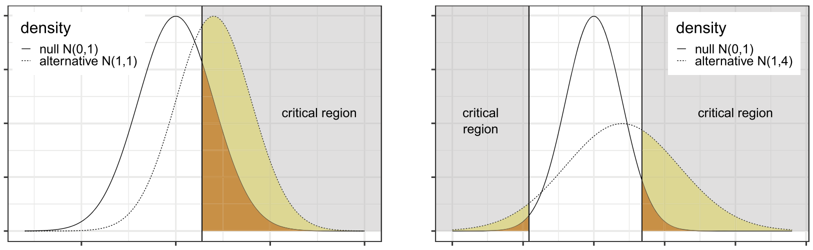
## 4. New procedure

In this section, we introduce our new method for paired multiple testing in the aforementioned noisy stochastically block model.

### 4.1. *Oracle procedure*

For a given level $\alpha \in (0, 1)$, we pursue the general aim of building a multiple testing procedure $\varphi$ that controls the FDR at level $\alpha$ while having a TDR as large as possible. Conventional procedures, such as BH procedure, base the inference of $A_{i,j}$ only on the single observation $X_{i,j}$, but do not take into account the specific dependency structure among the observations. By contrast, it is known from previous work [45, 49, 14] that maximizing the TDR while controlling the MFDR can be done via the multivariate test statistics $\mathbb{P}_\theta(A_{i,j} = 0|X)$, $(i, j) \in \mathcal{A}$, often called *local FDR* or *$\ell$-values* [23, 22, 15]. However, these quantities are intractable in our context, because their computation involves a sum over all possible values of the latent variable $Z$, which boils down to the well known problem of computing the likelihood in a mixture model. To circumvent this problem, we consider *structured $\ell$-values* that use an additional conditioning with respect to the membership structure $Z$: for all $(i, j) \in \mathcal{A}$, $z \in \{1, \dots, Q\}^n$, $\theta = (\pi, w, \nu_0, \nu) \in \Theta$, we define

$$\ell_{i,j}(X, z, \theta) = \mathbb{P}_\theta(A_{i,j} = 0 \mid X, Z = z) = \boldsymbol{\ell}(X_{i,j}, z_i, z_j, \theta), \tag{5}$$




(a) One-sided critical region. (b) Two-sided critical region.

FIG 3. *Null and alternative densities in the Gaussian case when $z_i = q, z_j = \ell$ and critical region $\{x : \boldsymbol{\ell}(x, q, \ell; \theta) \le t\}$ for some values of $t$ and $w_{q,\ell}$. The size of the orange area corresponds to $\boldsymbol{L}_0(t, q, \ell; \theta, \theta)$, the sum of the orange and the yellow area gives $\boldsymbol{L}_1(t, q, \ell; \theta, \theta)$.*

where $\boldsymbol{\ell}(\cdot)$ is defined as

$$\boldsymbol{\ell}(x, q, \ell, \theta) = \frac{(1 - w_{q,\ell})g_{0,\nu_0}(x)}{(1 - w_{q,\ell})g_{0,\nu_0}(x) + w_{q,\ell}g_{\nu_{q,\ell}}(x)}. \tag{6}$$

(Note that our model assumption implies that $\ell_{i,j}(X, z, \theta)$ in (5) indeed depends on $z$ only through $z_i, z_j$.) The structured $\ell$-value $\ell_{i,j}(X, z, \theta)$ incorporates the group information $z_i = q$ and $z_j = \ell$ via $w_{q,\ell}$ and $\nu_{q,\ell}$, which provides much more information than the single value $X_{i,j}$ and will considerably help to make the final decision. As illustrated in Figure 3 in the Gaussian case, the rejection region for $X_{i,j}$ corresponding to $\ell_{i,j}(X, Z; \theta) \le t$ is not necessarily of the form $|X_{i,j}| \ge c$, $c > 0$. For instance, depending on the model parameters, it can be one-sided (a), or two-sided with unbalanced sides (b) (other shapes are possible and are discussed in Appendix D). Hence, unlike methods based on $p$-values that are solely functions of $|X_{i,j}|$, this approach is more flexible and allows for various shapes of rejection regions, which can lead to a substantial power improvement.

Now, we choose the threshold $t$ in the decision rule $\ell_{i,j}(X, Z, \theta) \le t$ so that its MFDR is equal to $\alpha$. In the NSBM with true parameter $\theta^* = (\pi^*, w^*, \nu_0^*, \nu^*) \in \Theta$, the MFDR of this procedure is denoted, with some abuse of notation, by

$$\begin{aligned} \mathrm{MFDR}_{\theta^*}(\theta, t) &= \frac{\mathbb{E}_{\theta^*}\left[\sum_{(i,j)\in\mathcal{A}}(1 - A_{i,j})\mathbb{1}\{\ell_{i,j}(X, Z, \theta) \le t\}\right]}{\mathbb{E}_{\theta^*}\left[\sum_{(i,j)\in\mathcal{A}}\mathbb{1}\{\ell_{i,j}(X, Z, \theta) \le t\}\right]} \\ &= \frac{\sum_{q,\ell}\pi_q^*\pi_\ell^*(1 - w_{q,\ell}^*)\boldsymbol{L}_0(t, q, \ell; \theta^*, \theta)}{\sum_{q,\ell}\pi_q^*\pi_\ell^*[(1 - w_{q,\ell}^*)\boldsymbol{L}_0(t, q, \ell; \theta^*, \theta) + w_{q,\ell}^*\boldsymbol{L}_1(t, q, \ell; \theta^*, \theta)]}, \end{aligned} \tag{7}$$

where for $\delta \in \{0, 1\}$, $q, \ell \in \{1, \dots, Q\}$, we let

$$\begin{aligned} \boldsymbol{L}_\delta(t, q, \ell; \theta^*, \theta) &= \mathbb{P}_{\theta^*}(\ell_{i,j}(X, Z; \theta) \le t \mid Z, Z_i = q, Z_j = \ell, A_{i,j} = \delta) \\ &= \mathbb{P}_{\theta^*}(\boldsymbol{\ell}(X_{i,j}, q, \ell; \theta) \le t \mid Z_i = q, Z_j = \ell, A_{i,j} = \delta). \end{aligned} \tag{8}$$

Note that the functionals $\boldsymbol{L}_0$ and $\boldsymbol{L}_1$ do not depend on $(i,j) \in \mathcal{A}$. For $\theta = \theta^*$, the quantities $\boldsymbol{L}_0(t,q,\ell;\theta,\theta)$ and $\boldsymbol{L}_1(t,q,\ell;\theta,\theta)$ correspond to the size of the area of the rejection region under the null and the alternative, respectively, see Figure 3. In the Gaussian NSBM, these quantities have closed-form expressions provided in Appendix D.

Based on the above, in the NSBM with true parameter $\theta^* \in \Theta$ and latent variable $Z$, we reject the nulls satisfying $\ell_{i,j}(X,Z,\theta^*) \le t$ with a threshold $t$ chosen such that $\mathrm{MFDR}_{\theta^*}(\theta^*,t) = \alpha$. To circumvent the explicit calculation of such a threshold, we first introduce the quantities

$$q_{i,j}(X,z;\theta) = \mathrm{MFDR}_{\theta}(\theta, \ell_{i,j}(X,z;\theta)),\ \ (i,j) \in \mathcal{A}. \tag{9}$$

Next, composing by the function $\mathrm{MFDR}_{\theta^*}(\theta^*,\cdot)$ both sides of the inequality $\ell_{i,j}(X,Z,\theta^*) \le t$, we rewrite the above procedure as

$$\varphi^*_{i,j} = \mathbb{1}\{q_{i,j}(X,Z;\theta^*) \le \alpha\},\ \ (i,j) \in \mathcal{A}, \tag{10}$$

which is called the *oracle multiple testing procedure*. The quantity $q_{i,j}(X,Z;\theta^*)$ is often referred to as the *q-value*, a term popularized in [44]. Note that $\ell$-value thresholding and $q$-value thresholding are equivalent in our context, see also (16) below.

The procedure $\varphi^*$ has the following optimality property: under appropriate assumptions and for a convenient nominal level $\alpha$, the procedure $\varphi^*$ controls the MFDR, that is, $\mathrm{MFDR}_{\theta^*}(\varphi^*) \le \alpha$. Moreover, $\varphi^*$ has maximal power among all procedures controlling the MFDR, that is, for any multiple testing procedure $\varphi$ such that $\mathrm{MFDR}_{\theta^*}(\varphi) \le \alpha$, $\varphi^*$ satisfies $\mathrm{TDR}_{\theta^*}(\varphi^*) \ge \mathrm{TDR}_{\theta^*}(\varphi)$, see Theorem 6 below for a rigorous statement. The assumptions of the theorem are mainly regularity assumptions which are precisely detailed in Section 5.1.

### *4.2. Data-driven procedure*

The optimal procedure $\varphi^*$ cannot be used in general because the true model parameter $\theta^*$ and the block memberships $Z$ are unknown. To approximate the oracle by a feasible procedure, we use a plug-in approach, where $\theta^*$ is replaced by an estimator $\widehat{\theta} = (\widehat{\pi}, \widehat{w}, \widehat{\nu}_0, \widehat{\nu})$ and $Z$ by some clustering $\widehat{Z}$. One can use the estimates provided by the VEM algorithm for the NSBM, as pointed out in Section 3.2, but actually any appropriate estimates $(\widehat{\theta}, \widehat{Z})$ are allowed here (to ensure FDR control, some sufficient conditions will be exhibited further on, see Section 5.2). Therefore, we define the estimated version of the $q$- and $\ell$-values as

$$\widehat{\ell}_{i,j}(X) = \ell_{i,j}(X,\widehat{Z};\widehat{\theta}) = \frac{(1-\widehat{w}_{\hat{Z}_i,\hat{Z}_j})g_{0,\hat{\nu}_0}(X_{i,j})}{(1-\widehat{w}_{\hat{Z}_i,\hat{Z}_j})g_{0,\hat{\nu}_0}(X_{i,j}) + \widehat{w}_{\hat{Z}_i,\hat{Z}_j} g_{\hat{\nu}_{\hat{Z}_i,\hat{Z}_j}}(X_{i,j})}, \tag{11}$$

$$\widehat{q}_{i,j}(X) = q_{i,j}(X,\widehat{Z};\widehat{\theta}) = \mathrm{MFDR}_{\hat{\theta}}(\widehat{\theta}, \widehat{\ell}_{i,j}(X)), \tag{12}$$

and obtain the following feasible multiple testing procedure

$$\varphi^{\mathrm{New}}_{i,j} = \mathbb{1}\{\widehat{q}_{i,j}(X) \le \alpha\}, \quad (i,j) \in \mathcal{A}.$$

**Algorithm 2:** New paired multiple testing procedure

**Input**: Observations $X_{i,j}$, $(i,j) \in \mathcal{A}$, nominal level $\alpha$.
**Output**: Procedure $\varphi^{\mathrm{New}}_{i,j}$, $(i,j) \in \mathcal{A}$.
Compute a parameter estimate $\widehat{\theta}$ and a node clustering $\widehat{Z}$ for the observed data $X$.
Compute the $\ell$-values $\widehat{\ell}_{i,j}(X)$ according to (11).
Compute the $q$-values $\widehat{q}_{i,j}(X)$ according to (12).
Infer the decision for each pair by setting $\varphi^{\mathrm{New}}_{i,j} = \mathbb{1}\{\widehat{q}_{i,j}(X) \leq \alpha\}$, $(i,j) \in \mathcal{A}$.

Algorithm 2 presents the different steps to implement this procedure.

## 5. Theoretical properties of the new procedure

This section presents a theoretical study showing that our procedure $\varphi^{\mathrm{New}}$ correctly controls the FDR and has a TDR close to the one of the oracle procedure $\varphi^*$ defined in (10), which is shown to be optimal. The proofs of all results as well as further details on the assumptions are provided in Appendix C.

### *5.1. Assumptions and notation*

We introduce in this section the numerous and necessary assumptions to state our main result. More illustrations are provided in Appendix D in the Gaussian case, see also Examples 1 and 2 below.

According to (12), the behavior of the function $(t, \theta, \theta^*) \mapsto \mathrm{MFDR}_{\theta^*}(\theta, t)$ is crucial to study the properties of $\varphi^{\mathrm{New}}$. Since the former is related to the functionals $\boldsymbol{L}_0(\cdot)$ and $\boldsymbol{L}_1(\cdot)$ via (7), we introduce the following assumption.

**Assumption 3** (Continuity conditions for $\boldsymbol{L}_0$ and $\boldsymbol{L}_1$)**.** For all $q, \ell \in \{1, \dots, Q\}$,

- the functions $(t, \theta, \theta^*) \in [0,1] \times \Theta^2 \mapsto \boldsymbol{L}_0(t, q, \ell; \theta^*, \theta)$ and $(t, \theta, \theta^*) \in [0,1] \times \Theta^2 \mapsto \boldsymbol{L}_1(t, q, \ell; \theta^*, \theta)$ are continuous on $[0,1] \times \Theta^2$;
- there exist functions $\theta \in \Theta \mapsto t_{1,q,\ell}(\theta) \in [0,1]$ and $\theta \in \Theta \mapsto t_{2,q,\ell}(\theta) \in [0,1]$ with $t_{1,q,\ell}(\theta) < t_{2,q,\ell}(\theta)$ for all $\theta$, such that, for any $\theta^* \in \Theta$, the maps $t \in [0,1] \mapsto \boldsymbol{L}_0(t, q, \ell; \theta^*, \theta)$ and $t \in [0,1] \mapsto \boldsymbol{L}_1(t, q, \ell; \theta^*, \theta)$ are both continuous on $[0,1]$, with value 0 on $[0, t_{1,q,\ell}(\theta)]$, increasing on $[t_{1,q,\ell}(\theta), t_{2,q,\ell}(\theta)]$ and value 1 on $[t_{2,q,\ell}(\theta), 1]$.

When Assumption 3 holds, we let

$$t_1(\theta) = \min_{1 \leq q,\ell \leq Q} \{t_{1,q,\ell}(\theta)\} \quad \text{and} \quad t_2(\theta) = \max_{1 \leq q,\ell \leq Q} \{t_{2,q,\ell}(\theta)\}, \tag{13}$$

for any $\theta \in \Theta$. We also denote by

$$e_0(\theta) = \mathrm{MFDR}_\theta(\theta, 1) = \sum_{q,\ell} \pi_q \pi_\ell (1 - w_{q,\ell}) \tag{14}$$

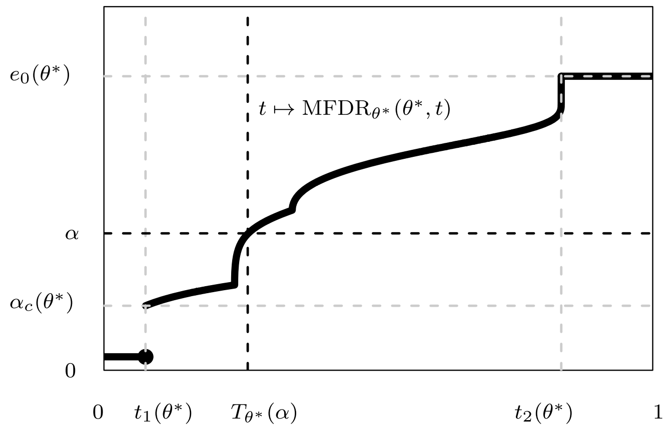



FIG 4. *Graphical illustration for* $t \mapsto \mathrm{MFDR}_{\theta^*}(\theta^*, t)$, $\alpha_c(\theta^*)$ *and* $T_{\theta^*}(\alpha)$.

the maximal value of $t \mapsto \mathrm{MFDR}_\theta(\theta, t)$, which corresponds to the average proportion of true nulls (or of absent edges in the graph). We sometimes write $e_0$ instead of $e_0(\theta)$ to lighten notation.

One can show that the function $t \mapsto \mathrm{MFDR}_\theta(\theta, t)$ is zero on $[0, t_1(\theta)]$, continuous increasing on $(t_1(\theta), t_2(\theta)]$ and is equal to $e_0(\theta)$ on $t \in [t_2(\theta), 1]$, see Figure 4 for an illustration. Moreover, $t \in [0,1] \mapsto \mathrm{MFDR}_\theta(\theta, t)$ is always non-decreasing and left-continuous so that we can define its generalized inverse in $\alpha \in [0, e_0]$ by

$$T_\theta(\alpha) = \max\{t \in [0,1] \,:\, \mathrm{MFDR}_\theta(\theta, t) \le \alpha\}, \;\; \theta \in \Theta. \tag{15}$$

Doing so, the oracle procedure given in (10) can be written as

$$\varphi^*_{i,j} = \mathbb{1}\{\ell_{i,j}(X, Z, \theta^*) \le T_{\theta^*}(\alpha)\}, \tag{16}$$

that is, as a $\ell$-value thresholding procedure with threshold $T_{\theta^*}(\alpha)$.

Let $\theta^* \in \Theta$ be the true model parameter. To avoid the regime where the quantity $\mathrm{MFDR}_{\theta^*}(\theta^*, t)$ is zero, we consider a level $\alpha$ above the so-called critical level $\alpha_c(\theta^*)$ introduced in [18] and defined in our context as

$$\alpha_c(\theta^*) = \lim_{t \to t_1(\theta^*)^+} \{\mathrm{MFDR}_{\theta^*}(\theta^*, t)\} \in [0, e_0). \tag{17}$$

It corresponds to the infimum of the non-zero values of $t \mapsto \mathrm{MFDR}_{\theta^*}(\theta^*, t)$. While $\alpha_c(\theta^*) = 0$ is the typical case, it occurs that $\alpha_c(\theta^*) > 0$ when $t_1(\theta^*) > 0$. Sometimes, we denote $\alpha_c(\theta^*)$ by $\alpha_c$ for short. Throughout this section, we thus fix a super-critical nominal level $\alpha \in (\alpha_c, e_0)$ with corresponding threshold $T_{\theta^*}(\alpha)$, as illustrated in Figure 4.

In addition, the following assumption ensures the uniform concentration of the underlying $\ell$-value empirical processes.

**Assumption 4** (Low-complexity for $\ell$-value thresholding)**.** There is an integer $M \geq 1$ such that for any $(i,j) \in \mathcal{A}$, $t \in [0,1]$, $\theta \in \Theta$, up to removing a set of $P_\theta$-probability measure 0, we have $\{\ell_{i,j}(X,Z,\theta) \leq t\} = \{\boldsymbol{\ell}(X_{i,j}, Z_i, Z_j, \theta) \leq t\} = \{X_{i,j} \in I\}$ for some $I \subset \mathbb{R}$ which only depends on $Z_i, Z_j, \theta, t$ and such that $I$ can be written as the union of at most $M$ non-empty open intervals of $\mathbb{R}$.

**Example 1.** Assumptions 3 and 4 are both satisfied in the Gaussian NSBM whenever $\theta^* = (\pi^*, w^*, \sigma_0^*, \mu^*, \sigma^*)$ is such that $(0, \sigma_0^*) \neq (\mu^*_{q,\ell}, \sigma^*_{q,\ell})$, for all $1 \leq q, \ell \leq Q$. In addition, $\alpha_c(\theta^*) = 0$ if and only if $\max_{q,\ell} \sigma^*_{q,\ell} \geq \sigma^*_0$, that is, if there is at least one variance under the alternative that is larger or equal to the variance under the null, see Appendix D.

While Assumption 3 is sufficient to establish consistency results for $\varphi^{\text{New}}$, the following regularity conditions are useful to obtain convergence rates.

**Assumption 5** (Lipschitz-type conditions for $\boldsymbol{L}_1$, MFDR and $T_{\theta^*}$)**.** For some compact interval $\mathcal{K} \subset [(\alpha_c + \alpha)/2, (\alpha + e_0)/2]$, there are three constants $C_1 = C_1(\theta^*, \alpha, \mathcal{K}) > 0$, $C_2 = C_2(\theta^*, \alpha, \mathcal{K}) > 0$ and $C_3 = C_3(\theta^*, \alpha, \mathcal{K}) > 0$ such that

(i) $\sup_{q,\ell} \sup_{t \in T_{\theta^*}(\mathcal{K})} \sup_{\delta \in \{0,1\}} \sup_{\theta, \theta' \in \Theta \setminus \{\theta^*\}} \left\{ \frac{\left| \boldsymbol{L}_\delta(t,q,\ell;\theta',\theta) - \boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta^*) \right|}{\|\theta - \theta^*\|_\infty \vee \|\theta' - \theta^*\|_\infty} \right\} \leq C_2$;

(ii) $\sup_{q,\ell} \sup_{t \in T_{\theta^*}(\mathcal{K}) \setminus \{T_{\theta^*}(\alpha)\}} \left\{ \frac{|\boldsymbol{L}_1(t,q,\ell;\theta^*,\theta^*) - \boldsymbol{L}_1(T_{\theta^*}(\alpha),q,\ell;\theta^*,\theta^*)|}{|t - T_{\theta^*}(\alpha)|} \right\} \leq C_1$;

(iii) $\sup_{y \in \mathcal{K} \setminus \{\alpha\}} \left\{ \frac{|T_{\theta^*}(y) - T_{\theta^*}(\alpha)|}{|y - \alpha|} \right\} \leq C_3$.

**Example 2.** Assumptions 3, 4 and 5 (i) (ii) are all satisfied in the Gaussian NSBM, in either one of the two following cases:

- Equal variances: for all $1 \leq q, \ell \leq Q$, $\sigma^*_{q,\ell} = \sigma^*_0$, $\mu^*_{q,\ell} \neq 0$;
- Larger variances under alternatives: for all $1 \leq q, \ell \leq Q$, $\sigma^*_{q,\ell} > \sigma^*_0$.

Assumption 5 (iii) also holds in this case, provided that $\alpha \in (0,1)$ is taken outside a set of Lebesgue measure 0. We also refer to Appendix D for weaker constraints on the parameter $\theta^*$.

Obviously, the behavior of $\varphi^{\text{New}}$ relies on the quality of the estimator $\widehat{\theta}$ and the clustering $\widehat{Z}$. To state our results on $\varphi^{\text{New}}$, we introduce the following risk probability defined for any $\theta^* \in \Theta$ and $\varepsilon > 0$ by

$$\eta(\theta^*, \varepsilon) = \mathbb{P}_{\theta^*}(\widehat{Z} \neq Z \text{ or } \|\widehat{\theta} - \theta^*\|_\infty > \varepsilon), \tag{18}$$

which corresponds to the probability that, either the clustering makes at least one mistake, or the estimator $\widehat{\theta}$ is more than $\varepsilon$ away from the true parameter $\theta^*$. Clearly, the norm and clustering inequalities in (18) hold up to label switching. Namely, the event $\{\|\widehat{\theta} - \theta^*\|_\infty > \varepsilon \text{ or } \widehat{Z} \neq Z\}$ means that for any permutation $\sigma$ of $\{1, \ldots, Q\}$ we have $\sigma(Z) \neq \widehat{Z}$ or $\|\widehat{\theta}^\sigma - \theta^*\|_\infty > \varepsilon$, where $\widehat{\theta}^\sigma = (\hat{\pi}^\sigma, \hat{w}^\sigma, \hat{\nu}_0, \hat{\nu}^\sigma)$ with $\hat{\pi}^\sigma = (\hat{\pi}_{\sigma(q)})_{1 \leq q \leq Q}$, $\hat{w}^\sigma = (\hat{w}_{\sigma(q),\sigma(\ell)})_{1 \leq q,\ell \leq Q}$, $\hat{\nu}^\sigma = (\hat{\nu}_{\sigma(q),\sigma(\ell)})_{1 \leq q,\ell \leq Q}$.

Since the pioneer paper of Celisse et al. [16], several studies have suggested that, in various SBM-type models, under appropriate restrictions on the parameter set $\Theta$, the order of the risk probability $\eta(\theta^*, \varepsilon)$ becomes small when $n$

increases, see, e.g., [7, 10]. This is proved for the maximum likelihood estimator, or alternatively for its variational approximation, and for a clustering based upon a maximum *a posteriori* approach, as used in our algorithm in Section 3.2. In particular, [10] suggests that $\eta(\theta^*, \varepsilon_n)$ should be small in a valued SBM, even when $\varepsilon = \varepsilon_n$ tends to zero at some rate. However, the framework used therein does not encompass the new NSBM. As the evaluation of $\eta(\theta^*, \varepsilon_n)$ in the NSBM would require an entirely new study, which is beyond the scope of this paper, we leave this task for future investigations and express all our results in terms of the implicit risk probability $\eta(\theta^*, \varepsilon)$.

### *5.2. Results*

To start with, we state the optimality of $\varphi^*$ in terms of MFDR, FDR and TDR.

**Theorem 6** (Optimality of $\varphi^*$)**.** *Let Assumption 3 be true and $\theta^* \in \Theta$. Let $\alpha \in (\alpha_c, e_0)$ with $\alpha_c = \alpha_c(\theta^*)$ given by (17) and $e_0 = e_0(\theta^*)$ defined by (14). Then, $\mathrm{MFDR}_{\theta^*}(\varphi^*) = \alpha$, and for any multiple testing procedure $\varphi$ such that $\mathrm{MFDR}_{\theta^*}(\varphi) \leq \alpha$, we have $\mathrm{TDR}_{\theta^*}(\varphi^*) \geq \mathrm{TDR}_{\theta^*}(\varphi)$. Moreover, under the additional Assumption 4, we have $\limsup_n \{\mathrm{FDR}_{\theta^*}(\varphi^*)\} \leq \alpha$, where parameter $\theta^*$ is fixed and does not depend on $n$.*

Now, we present our two main results, showing that $\varphi^{\mathrm{New}}$ mimics the behavior of $\varphi^*$, both in terms of FDR and TDR, up to remainder terms. First, the following consistency result holds when the sample size $n$ increases and model parameter $\theta^*$ is fixed with $n$.

**Theorem 7** (Consistency)**.** *Let Assumptions 3 and 4 be true. Let $\theta^* \in \Theta$ be a fixed parameter, that does not depend on the sample size $n$. Choose $\alpha \in (\alpha_c, e_0)$ where $\alpha_c = \alpha_c(\theta^*)$ given by (17) and $e_0 = e_0(\theta^*)$ given by (14). Consider the procedures $\varphi^*$ defined by (10) and $\varphi^{New}$ of Algorithm 2 for some estimator $\widehat{\theta}$ and clustering $\widehat{Z}$. Assume that the estimator $\widehat{\theta}$ and clustering $\widehat{Z}$ are consistent, that is, $\eta(\theta^*, \varepsilon)$ given by (18) converges to 0 for any $\varepsilon > 0$ as $n$ tends to infinity. Then, we have*

$$\limsup_n \{\mathrm{FDR}_{\theta^*}(\varphi^{New})\} \leq \alpha, \quad \liminf_n \{\mathrm{TDR}_{\theta^*}(\varphi^{New}) - \mathrm{TDR}_{\theta^*}(\varphi^*)\} \geq 0.$$

Theorem 7 shows that, in the graph-structured NSBM, it is possible to construct a procedure that controls the FDR and which is asymptotically optimal with respect to the TDR, provided that a consistent estimator $\widehat{\theta}$ and clustering $\widehat{Z}$ are available. Here, consistency means that $\eta(\theta^*, \varepsilon)$ converges to 0, see the end of Section 5.1 for more discussions on that condition. This results is in line with the state-of-the-art consistency results for the FDR and TDR in mixture models with structured latent variables, see [45, 13, 46, 14]. Theorem 7 relies on Assumptions 3 and 4. For instance, they both hold in the Gaussian NSBM with a true parameter taken as in Example 1.

Now, in Appendix C.1, we have derived more general results (see Theorems 10 and 11 therein), which are non-asymptotic, that is, are valid for any $n \geq 2$. Here we only state an important consequence of this result concerning the rate of convergence of the FDR and the TDR of the new procedure (the true parameter $\theta^*$ being still fixed with $n$).

**Theorem 8** (Convergence rate)**.** *Let Assumptions 3 and 4 (with some constant $M > 0$) be true, $\theta^* \in \Theta$, $\alpha \in (\alpha_c, e_0)$ where $\alpha_c = \alpha_c(\theta^*)$ given by (17) and $e_0 = e_0(\theta^*)$ given by (14). Let Assumption 5 (i) be true for some compact set $\mathcal{K} \subset [(\alpha_c + \alpha)/2, (\alpha + e_0)/2]$. Consider the procedure $\varphi^{New}$ of Algorithm 2 for some estimator $\widehat{\theta}$ and some clustering $\widehat{Z}$ with associated risk probability $\eta(\theta^*, \cdot)$ defined by (18). Consider the oracle procedure $\varphi^*$ defined by (10). Then there exist constants $C = C(\theta^*, \alpha, \mathcal{K}, M)$ and $N = N(\theta^*, \alpha, \mathcal{K}, M) \in (0, 1)$ such that for all $n \geq N$, we have*

$$\mathrm{FDR}_{\theta^*}(\varphi^{New}) \leq \alpha + C\varepsilon_n + \eta(\theta^*, \varepsilon_n), \tag{19}$$

*for any sequence $\varepsilon_n \geq \sqrt{\frac{\log n}{n}}$. If, in addition, Assumptions 5 (ii)-(iii) hold (for the same set $\mathcal{K}$), then there exist constants $C = C(\theta^*, \alpha, \mathcal{K}, M)$ and $N = N(\theta^*, \alpha, \mathcal{K}, M) \in (0, 1)$ such that for all $n \geq N$, we have*

$$\mathrm{TDR}_{\theta^*}(\varphi^*) \leq \mathrm{TDR}_{\theta^*}(\varphi^{New}) + C\varepsilon_n + \eta(\theta^*, \varepsilon_n), \tag{20}$$

*for any sequence $\varepsilon_n \geq \sqrt{\frac{\log n}{n}}$.*

Above, $\varepsilon_n$ corresponds to an upper bound on the convergence rate of $\widehat{\theta}$. Given the derived bounds, it is desirable to choose $\varepsilon_n$ such that both $\varepsilon_n$ and $\eta(\theta^*, \varepsilon_n)$ tend to zero, and possibly at the same rate. The results given in [10] for other valued SBM-type models (see, e.g., Theorem 6.2 and Proposition 3 therein) suggest that $\eta(\theta^*, \varepsilon_n)$ could tend to zero by taking $\sqrt{n}\varepsilon_n \to \infty$. This is achieved by the condition $\varepsilon_n \geq \sqrt{\frac{\log n}{n}}$ in Theorem 8.

Finally, compared to the consistency result, Theorem 8 uses the additional Assumption 5. This assumption is valid in the Gaussian NSBM under restrictions on the model parameters and on the range of $\alpha$, see Example 2.

## 6. Numerical experiments

In this section, a simulation study evaluates the FDR and power of the new multiple testing procedure $\varphi^{\mathrm{New}}$. Moreover, we provide a comparison with standard test procedures in various settings. For reproducibility of results, the code is available on request.

### *6.1. Settings*

We consider four scenarios (a)–(d), that we now describe in detail. In all scenarios, undirected graphs with $n = 200$ nodes are generated and the standard normal distribution is chosen as the null.

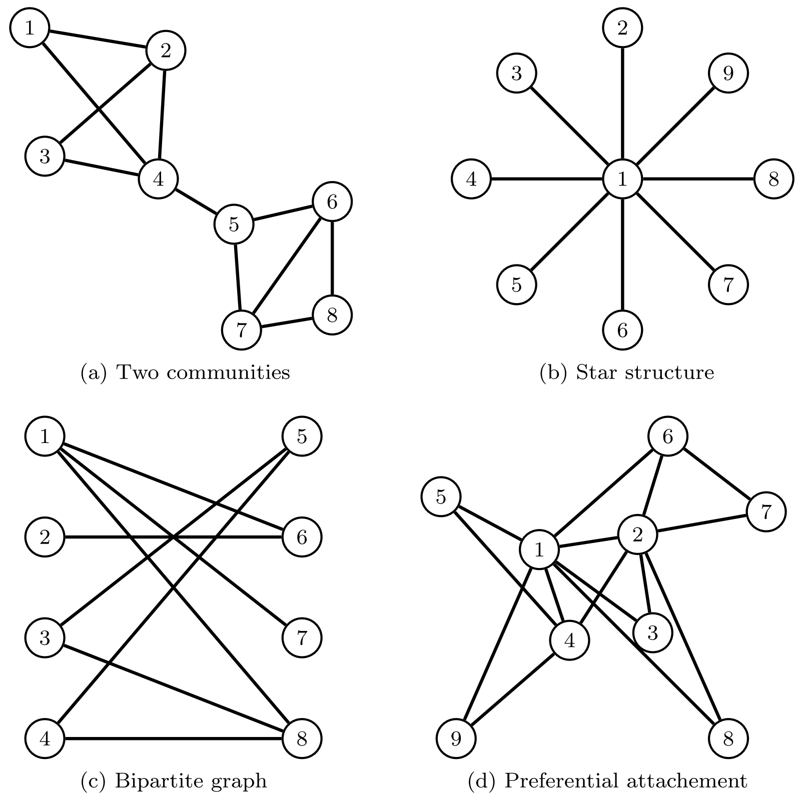



Fig 5. *Structure of the latent graph $A$ in the simulation scenarios (a)–(d).*

(a) Data are simulated from the Gaussian NSBM with $Q = 2$ latent groups and equal group probabilities ($\pi_1 = \pi_2 = 1/2$). The connectivity parameter $w$ is such that $A$ has two communities, namely $w_{1,1} = w_{2,2} = 0.8$ and $w_{1,2} = 0.1$. For the alternative distributions, we set Gaussian means to $\mu_{1,1} = \mu_{2,2} = 1$ and $\mu_{1,2} = 3$, so that there is a strong signal for inter-community connections. Gaussian variances are all equal, that is $\sigma^2_{q,\ell} = 1$ for all $q, \ell$.

The difference of the following three scenarios resides in the generation of the latent graph $A$. Then, given $A$, all observations $X_{i,j}$ are drawn independently from $\mathcal{N}(2, 1)$ if $A_{i,j} = 1$ and from $\mathcal{N}(0, 1)$ under the null.

(b) The latent graph $A$ is fixed and has the form of a star.
(c) In this setting, $A$ is a random bipartite graph. Nodes are partitioned into two groups of equal size. Pairs of nodes that belong to different groups connect with probability 0.5, while there is no edge ($A_{i,j} = 0$) whenever nodes $i$ and $j$ belong to the same group.
(d) In the last scenario, $A$ is generated according to a preferential attachment

or Barabási-Albert model [3]. Starting from a small root graph, the graph is grown sequentially by adding new nodes that connect randomly to a given number of nodes, by privileging connections to existing nodes with high degree. In this way, highly heterogeneous networks are obtained. Here, the root graph is an Erdos-Rényi graph with 40 nodes and density 0.5, and every new node connects to 30 existing nodes.

For all scenarios, the latent graphs $A$ are illustrated in Figure 5. Note that in (b), (c) and (d) the latent graph $A$ is not generated according to a SBM. Thus, the simulated data are not exactly from a NSBM. In this regard, our experiments also provide insights on the robustness of the NSBM with respect to model misspecifications.

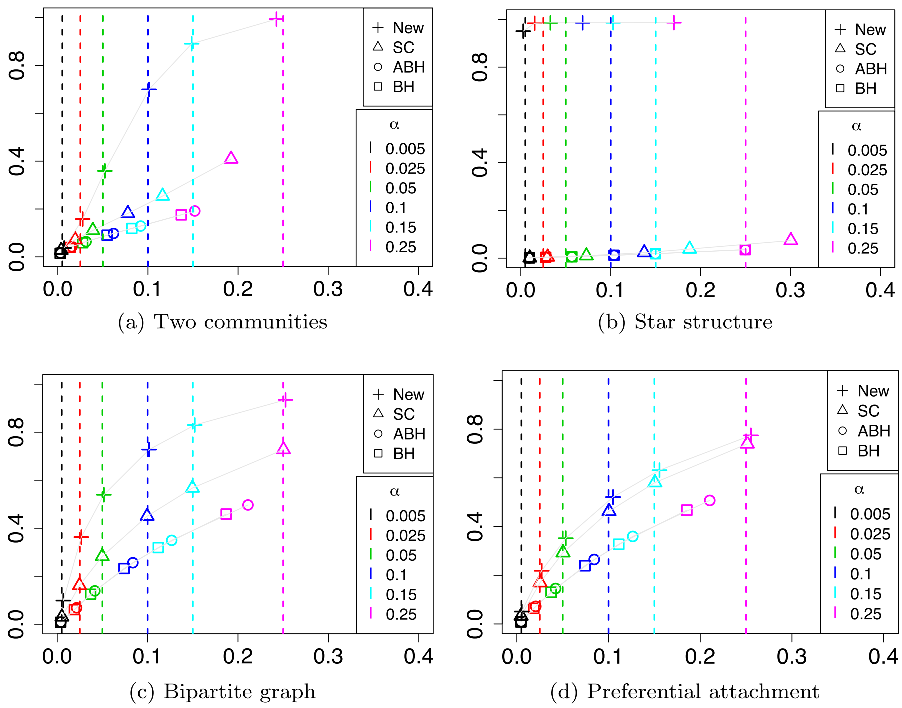
### 6.2. Procedures

All procedures that are considered in our study are known to control the FDR, either for finite $n$ or asymptotically, as $n$ tends to infinity. Most of them are especially suited for normally distributed data. As some methods require the knowledge of the null distribution, we provide this information to all procedures for reasons of equity. The following procedures are considered:

- The *Benjamini-Hochberg procedure (BH)* [4], used with $p$-values associated with the two-sided test statistics $|X_{i,j}|$, given by
$$p_{i,j}(X) = 2(1 - \Phi(|X_{i,j}|)), \quad (i,j) \in \mathcal{A}, \tag{21}$$
where $\Phi$ denotes the cumulative distribution function of the standard normal distribution. This procedure, also denoted $\varphi^{BH}$, has the property that $\mathrm{FDR}_\theta(\varphi^{BH}) = e_0\alpha \le \alpha$, since conditional independence holds in the NSBM [6].
- The *adaptive BH procedures (ABH)*, which corresponds to the BH procedure applied at level $\alpha/\widehat{e}_0$, where $\widehat{e}_0 = \frac{2}{m}(1 + \sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{p_{i,j}(X) > 1/2\})$ is an estimate of $e_0$ [42, 43]. This procedure is an improvement of BH that still ensures $\mathrm{FDR}_\theta(\varphi^{ABH}) \le \alpha$ [5, 9].
- The *Sun and Cai procedure (SC)*, which is based on thresholding cumulative means of $\ell$-values, which are computed by estimating the alternative density by a mixture distribution [45]. Note that the $\ell$-values used in SC do not take into account the graph structure. For computations, the available R code is used [11].
- The new procedure $\varphi^{\mathrm{New}}$ given by Algorithm 2 for the Gaussian NSBM with known null distribution and with the VEM algorithm. Recall that a data-driven model selection is done to select the number of node blocks.

### 6.3. Results

For every scenario, 500 datasets are simulated and all test procedures are applied with nominal FDR level $\alpha$ ranging in $\{0.005, 0.025, 0.05, 0.1, 0.15, 0.25\}$.




(a) Two communities

(b) Star structure

(c) Bipartite graph

(d) Preferential attachment

FIG 6. *Plot of* $(\widehat{\mathrm{FDR}}_\alpha, \widehat{\mathrm{TDR}}_\alpha)$ *for BH, ABH, SC and the new procedure* $\varphi^{New}$*. Dashed lines represent the nominal levels* $\alpha$*.*

We evaluate the sample FDR and the sample TDR, denoted by $\widehat{\mathrm{FDR}}_\alpha$ and $\widehat{\mathrm{TDR}}_\alpha$, respectively. Displaying the points $(\widehat{\mathrm{FDR}}_\alpha, \widehat{\mathrm{TDR}}_\alpha)$ for all $\alpha$ and all test procedures provides ROC-type curves, simultaneously illustrating whether the FDR is correctly controlled and evaluating the test power. All results are given in Figure 6.

In all considered scenarios, BH and its adaptive variant ABH behave very similarly: most of the time the FDR is controlled by the nominal level $\alpha$, however, the TDR is relatively low. A substantial improvement in power is obtained by the SC procedure, while the FDR is not always controlled in scenario (b). The improvement is due to the estimation of the alternative distribution, which leads to more meaningful rejection regions.

Concerning the new procedure, we observe that in all considered scenarios, $\varphi^{\mathrm{New}}$ has a uniformly better performance than the other methods. Its FDR is close to or lower than the nominal level $\alpha$. Moreover, $\varphi^{\mathrm{New}}$ outperforms the other procedures in terms of the TDR, sometimes substantially. This is in line with our theoretical results, namely Theorem 8.

In scenario (a), learning the communities of the latent graph $A$ helps in the detection of intragoup connections although the signal is rather small and noisy ($\mu_{11} = \mu_{22} = 1$). Likewise for the star structure, the TDR enhancement over

BH-like procedures and SC is spectacular. Here the model selection device in the VEM algorithm mostly finds $Q = 2$ groups, where the center of the star forms one group and all other nodes the second group. The probability of connection between the two groups is estimated close to 1 ($\widehat{w}_{1,2} \approx 1$), so that intergroup connections are efficiently discovered. Without learning the latent star structure, the other methods fail to detect false nulls. Similar results are obtained for the bipartite random graph, which is conveniently approximated by a NSBM with $Q = 2$ groups.

Scenario (b), (c) and (d) serve to evaluate robustness of the results in view of model mis-specification, as latent graph $A$ is not a SBM. For instance, in (d) the new procedure mainly selects a NSBM with $Q = 2$ groups to fit the data. One group contains the nodes with high degree, the other group the remaining nodes. Markedly, despite the model bias, the performance of our procedure is only weakly affected: the FDR control is essentially maintained, and the ROC curve still dominates the SC procedure even though the difference is less salient. The robustness of our method is due to the fact that the SBM is able to accommodate a wide spectrum of graph topologies, preferential attachement models included.

## 7. Application to bicycle-sharing network

Our goal is to study how the usage of a public bicycle hire scheme in London depends on the weekday. We use data for the year 2019 available from the web site https://cycling.data.tfl.gov.uk. For illustration, journey counts between pairs of stations for the week of April 3–9, 2019 are displayed in Figure 7 (a). We see that overall counts for weekdays are very similar, while a significant difference is observed on the weekend.

Our study focuses on a comparison of the traffic on Mondays and Tuesdays during the entire year 2019. While the global activity on these weekdays is very much alike, we aim at identifying the pairs of docking stations in London that have significantly different daily journey counts. The signal to be identified here is rather sparse.

These data can be modeled by a NSBM, for which nodes represent docking stations, and observations are the values of a test statistic that compares the traffic on Mondays and Tuesdays for a given pair of stations. To be explicit, for stations $i$ and $j$ and for every day $d$ of 2019, let $Y_{i,j}^{d}$ be the number of journeys realized between these two stations at this date. Denote by $\bar{Y}_{i,j}^{\mathrm{Mon}}$ (resp. $\bar{Y}_{i,j}^{\mathrm{Tue}}$) the empirical mean over the year 2019 of the day counts $Y_{i,j}^{d}$ of all Mondays (resp. Tuesdays). Let us assume that $\bar{Y}_{i,j}^{\mathrm{Mon}}$ (resp. $\bar{Y}_{i,j}^{\mathrm{Tue}}$) is the mean of i.i.d. Poisson variables of mean $\mu_{i,j}^{\mathrm{Mon}}$ (resp. $\mu_{i,j}^{\mathrm{Tue}}$). The observed difference between Mondays and Tuesdays is summarized in the following normalized statistics

$$X_{i,j} = \frac{\bar{Y}_{i,j}^{\mathrm{Mon}} - \bar{Y}_{i,j}^{\mathrm{Tue}}}{\sqrt{\bar{Y}_{i,j}^{\mathrm{Mon}}/n^{\mathrm{Mon}} + \bar{Y}_{i,j}^{\mathrm{Tue}}/n^{\mathrm{Tue}}}},$$

(with, by convention, $X_{i,j} = 0$ if the denominator is zero), where $n^{\mathrm{Mon}} = 52$

(a)

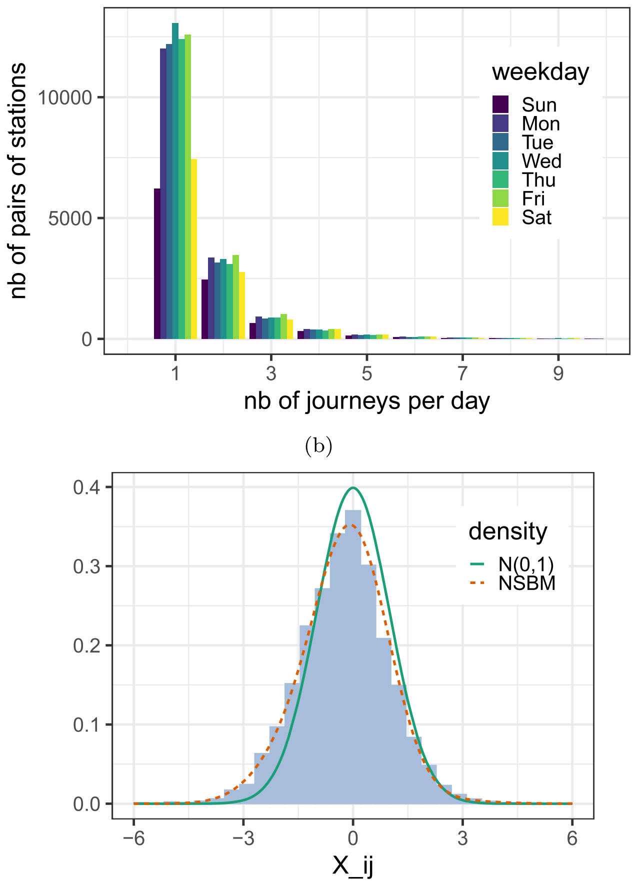



FIG 7. *(a) Journey counts per day between pairs of stations for April 3–9, 2019. (b) Histogram of the test statistics* $X_{i,j}$ *compared to the standard normal distribution and to the estimated marginal distribution of* $X_{i,j}$ *in the NSBM.*

(resp. $n^{\text{Tue}} = 53$) are the number of Mondays (resp. Tuesdays) in 2019. Provided that the sample sizes are large enough, the variables $X_{i,j}$ are approximately Gaussian with variance 1.

We would like to apply our procedure to the test statistics $X_{i,j}, (i,j) \in \mathcal{A}$ with a Gaussian model. However, the full data set contains 758 docking stations and most of them are so far from another that there are no journeys between them. Plenty of pairs of stations do not count a single journey over the whole year, resulting in a very sparse matrix $X$ (more than 62% of the entries are zero), which is not compatible with the Gaussian assumption for the $X_{i,j}$. Thus, we delete stations to achieve a rate of only 3% of zero entries of $X$, leading to a graph with $n = 152$ nodes and 11476 possible edges. From Figure 7 (b)

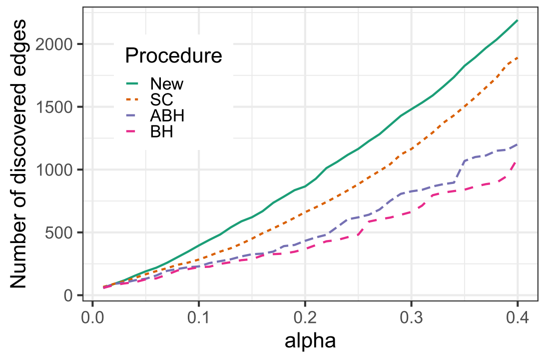



Fig 8. *Number of rejected edges as a function of the significance level $\alpha$ for the different procedures.*

we see that data look Gaussian, but do not come exactly from the standard normal distribution. This indicates that some observations come from a different alternative distribution and the interest of a mixture model.

On these pre-processed data, we run the new test procedure, that is, Algorithm 2 with the estimators of the VEM algorithm for a Gaussian NSBM with a data-driven selection of the number of latent blocks. We observe in Figure 7 (b) a very good fit of the estimated marginal distribution of $X_{i,j}$ to the data.

At $\alpha = 0.05$, the new procedure makes 190 discoveries out of 11476 possible pairs. This low rate of rejection (1.6%) supports the observation/assumption that the overall traffic is very similar on Mondays and Tuesdays. Figure 8 compares our multiple testing procedure to the BH, ABH and SC procedures in terms of number of rejections as a function of the nominal significance level $\alpha$. The new procedure detects more signal than the other methods for almost all levels $\alpha$. For instance, at $\alpha = 0.05$, SC finds only 167, ABH 131 and BH 126 discoveries. This is coherent with our theoretical and numerical findings. Hence, more station pairs with a significant difference in the traffic intensity are identified.

In addition, the values of the normal means $\hat{\mu}_{q,\ell}$ estimated by our algorithm provide further information on the alternative distributions: they indicate whether the traffic on Tuesdays is amplified or diminished compared to Mondays, and even quantify the amplitude of the phenomenon. In more detail, our algorithm selects a model with two latent blocks of almost equal size ($\hat{\pi} = (0.48, 0.52)$). According to the obtained parameter estimates, there are 3% of false nulls among stations belonging to the first group ($\hat{w}_{1,1} = 0.035$), with a large alternative mean ($\hat{\mu}_{1,1} = 2.95$). This means that there is significantly more traffic on Mondays than on Tuesdays for these dock stations. By contrast, within the second group, the proportion of false nulls is estimated to 30% ($\hat{w}_{2,2} = 0.308$) with a negative alternative mean ($\hat{\mu}_{2,2} = -1.41$). This means that significantly less bikes are used on Mondays than on Tuesdays. Finally, in the intergroup, the proportion of false nulls is moderate ($\hat{w}_{1,2} = 0.146$) and the alternative mean is $\hat{\mu}_{1,2} = -1.84$. So this is similar to the intra traffic of group 2, but with a different amount of signal in proportion and intensity. All in all, our method provides valuable informations that can help the practitioner when optimizing the distribution of bikes among the stations.

Interestingly, we can check that the decision is not only driven by the amplitude of $X_{i,j}$. For instance, there are pairs with moderately large positive values $X_{i,j}$ that are not declared as significant by the new procedure, because the interacting nodes are learned to belong to blocks such that under the alternative the signal takes a negative value. Thus, the observed value $X_{i,j}$ is more likely to be generated by the null distribution than under the alternative. In other words, by analyzing the structure of the graph as a whole and the positions of the stations therein, our method interprets some of the slightly extreme values $X_{i,j}$ of the test statistic as noise and concludes that there is no evidence for a difference between the bike usage on Mondays or Tuesdays.

## 8. Conclusion and discussion

We introduced a new procedure for multiple testing of paired null hypotheses, for which the data are collected as entries of a matrix. By viewing the multiple testing task as a graph inference problem with an NSBM structure, we have significantly enhanced the performance of the method compared to the state of the art. In addition to the false discovery rate control, the inferred decision comes with a clustering of the individuals which is a useful side information that increases interpretability. While the interest of the method has been demonstrated on a particular application, it can be applied broadly whenever the practitioner collects interaction measures between individuals (independently for each pair).

The developed method has been validated via FDR– and TDR–guarantees that are non-asymptotic in the number of tests under general regularity assumptions. This is a novelty with respect to the multiple testing literature in mixture models. As a case in point, Theorem 8 is to our knowledge the first result in a mixture-model context providing convergence rates for the FDR control and the TDR optimality. In addition, while our proof technics obviously rely on the specific NSBM structure, they can accommodate various structured frameworks, as standard gaussian mixtures, because they are based on concentration inequalities.

This work opens several interesting directions of research. First, to go further in the study of the convergence rates of the FDR and TDR, the quantification of the quality of the parameter estimator and the clustering obtained by the VEM algorithm in the NSBM has to be addressed. Namely, in Theorem 8, $\varepsilon_n$ is an upper bound on the convergence rate of $\widehat{\theta}$ that should be chosen so that $\eta(\theta^*, \varepsilon_n)$ still tends to zero. It is desirable to choose $\varepsilon_n$ balancing the quantities $\varepsilon_n$ and $\eta(\theta^*, \varepsilon_n)$, which requires to evaluate the risk probability $\eta(\theta^*, \varepsilon_n)$. This task meets several recent works on various versions of the SBM, see [16, 7, 10], but the proposed solutions are still incomplete to make our convergence rates explicit.

Second, our consistency results are established when the 'true' parameter $\theta^*$ is kept fixed with the number $n$ of individuals. While our non-asymptotic statements allow to get some bounds when the parameters depend on $n$, it is unclear whether our consistency result holds in that situation, especially when some $w_{q,\ell}$ tend to zero at some rate, that is, when the true interaction probability vanishes. Studying such a sparse situation is an interesting avenue for future research.

Finally, our numerical experiments (our scenario (b)) suggest that the FDR and TDR results might hold *outside the NSBM*, namely in the case where the adjacency matrix $A$ (that is, the true/false constellation of nulls) is fixed and deterministic, rather than generated randomly according to a SBM. Related to this, it is interesting to adopt a Bayesian point of view on our modeling: the parameter of interest being $A$, the SBM can be seen as a prior distribution on the parameter, while the underlying frequentist model is the one with a deterministic matrix $A$. In this view, the new VEM-based procedure is an empirical Bayes procedure that fits the hyper-parameters $\pi$ and $w$ by a marginal maximum likelihood approach. Recent literature on frequentist properties of Bayesian multiple testing procedures, see, e.g., [15], suggests that using Cauchy alternative distributions, that is, Cauchy *slabs*, rather than Gaussians might be of interest, especially under sparsity of the signal. Studying the FDR and TDR properties of the resulting inference procedure is thus another interesting direction for future work.

## Appendix A: Identifiability of the NSBM

### A.1. Proof of Theorem 2

The proof of Theorem 2 uses a technique similar to the one in the identifiability proof of Theorem 12 in [2].

*Proof.* We consider the joint distribution of the edges relating the first three nodes $K_3 = (X_{1,2}, X_{2,3}, X_{3,1})$ to identify all model parameters, up to label swapping. Denote $[\![1,Q]\!] = \{1,\dots,Q\}$, and $\bar{w}_{q,\ell} = 1 - w_{q,\ell}$. Let $G_u$ be the cumulative distribution function associated with density $g_u$ for all $u \in \mathcal{V}$. The cumulative distribution function $F_{K_3}$ of the triplet $K_3$ is given by

$$\begin{aligned}
&F_{K_3}(x_1,x_2,x_3)\\
&\quad= \sum_{a_1,a_2,a_3\in\{0,1\}^3}\ \sum_{q,\ell,m\in[\![1,Q]\!]^3} \mathbb{P}(X_{1,2}\le x_1, X_{2,3}\le x_2, X_{3,1}\le x_3,\\
&\qquad\qquad A_{1,2}=a_1, A_{2,3}=a_2, A_{3,1}=a_3, Z_1=q, Z_2=\ell, Z_3=m)\\
&\quad= \sum_{a_1,a_2,a_3}\sum_{q,\ell,m} \pi_q\pi_\ell\pi_m[\bar{w}_{q,\ell}G_{\nu_0}(x_1)]^{1-a_1}[w_{q,\ell}G_{\nu_{q,\ell}}(x_1)]^{a_1}[\bar{w}_{\ell,m}G_{\nu_0}(x_2)]^{1-a_2}\\
&\qquad\qquad \times [w_{\ell,m}G_{\nu_{\ell,m}}(x_2)]^{a_2}[\bar{w}_{q,m}G_{\nu_0}(x_3)]^{1-a_3}[w_{q,m}G_{\nu_{q,m}}(x_3)]^{a_3},
\end{aligned}$$

TABLE 1

*List of mixture component distributions with corresponding mixing weight of the distribution of $K_3$. The last two columns give the possible values of the variables indicating the presence or absence of edges $\mathbf{A}_{1,2,3} = (A_{1,2}, A_{2,3}, A_{3,1})$ and the group memberships $(Z_1, Z_2, Z_3)$ for the corresponding mixture component distribution.*

| For | Component | Mixing weight | $\mathbf{A}_{1,2,3}$ | $(Z_1, Z_2, Z_3)$ |
|---|---|---|---|---|
| $q \in [\![1, Q]\!]$ | $G_{\nu_{q,q}} \otimes G_{\nu_{q,q}} \otimes G_{\nu_{q,q}}$ | $\pi_q^3 w_{q,q}^3$ | (1,1,1) | $(q, q, q)$ |
| $q \neq \ell$ | $G_{\nu_{q,q}} \otimes G_{\nu_{q,\ell}} \otimes G_{\nu_{q,\ell}}$ | $\pi_q^2 \pi_\ell w_{q,q} w_{q,\ell}^2$ | (1,1,1) | $(q, q, \ell)$ |
| $q \neq \ell$ | $G_{\nu_{q,\ell}} \otimes G_{\nu_{q,q}} \otimes G_{\nu_{q,\ell}}$ | $\pi_q^2 \pi_\ell w_{q,q} w_{q,\ell}^2$ | (1,1,1) | $(\ell, q, q)$ |
| $q \neq \ell$ | $G_{\nu_{q,\ell}} \otimes G_{\nu_{q,\ell}} \otimes G_{\nu_{q,q}}$ | $\pi_q^2 \pi_\ell w_{q,q} w_{q,\ell}^2$ | (1,1,1) | $(q, \ell, q)$ |
| $q \neq \ell \neq m$ | $G_{\nu_{q,\ell}} \otimes G_{\nu_{\ell,m}} \otimes G_{\nu_{m,q}}$ | $\pi_q \pi_\ell \pi_m w_{q,\ell} w_{\ell,m} w_{m,q}$ | (1,1,1) | $(q, \ell, m)$ |
| $q, \ell, m \in [\![1, Q]\!]^3$ | $G_{\nu_{q,\ell}} \otimes G_{\nu_{\ell,m}} \otimes G_{\nu_0}$ | $\pi_q \pi_\ell \pi_m w_{q,\ell} w_{\ell,m} \bar{w}_{m,q}$ | (1,1,0) | $(q, \ell, m)$ |
| $q, \ell, m \in [\![1, Q]\!]^3$ | $G_{\nu_{q,\ell}} \otimes G_{\nu_0} \otimes G_{\nu_{\ell,m}}$ | $\pi_q \pi_\ell \pi_m w_{q,\ell} w_{\ell,m} \bar{w}_{m,q}$ | (1,0,1) | $(\ell, q, m)$ |
| $q, \ell, m \in [\![1, Q]\!]^3$ | $G_{\nu_0} \otimes G_{\nu_{q,\ell}} \otimes G_{\nu_{\ell,m}}$ | $\pi_q \pi_\ell \pi_m w_{q,\ell} w_{\ell,m} \bar{w}_{m,q}$ | (0,1,1) | $(m, q, \ell)$ |
| $q \in [\![1, Q]\!]$ | $G_{\nu_{q,q}} \otimes G_{\nu_0} \otimes G_{\nu_0}$ | $\pi_q^2 w_{q,q} \sum_\ell \pi_\ell \bar{w}_{q,\ell}^2$ | (1,0,0) | $(q, q, [\![1, Q]\!])$ |
| $q < \ell$ | $G_{\nu_{q,\ell}} \otimes G_{\nu_0} \otimes G_{\nu_0}$ | $2\pi_q \pi_\ell w_{q,\ell} \sum_m \pi_m \bar{w}_{\ell,m} \bar{w}_{m,q}$ | (1,0,0) | $(q, \ell, [\![1, Q]\!]) \cup (\ell, q, [\![1, Q]\!])$ |
| $q \in [\![1, Q]\!]$ | $G_{\nu_0} \otimes G_{\nu_{q,q}} \otimes G_{\nu_0}$ | $\pi_q^2 w_{q,q} \sum_\ell \pi_\ell \bar{w}_{q,\ell}^2$ | (0,1,0) | $([\![1, Q]\!], q, q)$ |
| $q < \ell$ | $G_{\nu_0} \otimes G_{\nu_{q,\ell}} \otimes G_{\nu_0}$ | $2\pi_q \pi_\ell w_{q,\ell} \sum_m \pi_m \bar{w}_{\ell,m} \bar{w}_{m,q}$ | (0,1,0) | $([\![1, Q]\!], q, \ell) \cup ([\![1, Q]\!], \ell, q)$ |
| $q \in [\![1, Q]\!]$ | $G_{\nu_0} \otimes G_{\nu_0} \otimes G_{\nu_{q,q}}$ | $\pi_q^2 w_{q,q} \sum_\ell \pi_\ell \bar{w}_{q,\ell}^2$ | (0,0,1) | $(q, [\![1, Q]\!], q)$ |
| $q < \ell$ | $G_{\nu_0} \otimes G_{\nu_0} \otimes G_{\nu_{q,\ell}}$ | $2\pi_q \pi_\ell w_{q,\ell} \sum_m \pi_m \bar{w}_{\ell,m} \bar{w}_{m,q}$ | (0,0,1) | $(q, [\![1, Q]\!], \ell) \cup (\ell, [\![1, Q]\!], q)$ |
| | $G_{\nu_0} \otimes G_{\nu_0} \otimes G_{\nu_0}$ | $\sum_{q,\ell,m} \pi_q \pi_\ell \pi_m \bar{w}_{q,\ell} \bar{w}_{\ell,m} \bar{w}_{m,q}$ | (0,0,0) | $[\![1, Q]\!]^3$ |

using the conditional independence property of the NSBM. From the above expression, it is clear that $K_3$ has a (3-dimensional) mixture distribution with numerous terms, all of them having independent coordinates. More precisely, Table 1 summarizes all mixture component distributions with corresponding mixing weights. Furthermore, the last two columns of the table give the possible values of the variables indicating the presence or absence of edges $\mathbf{A}_{1,2,3} = (A_{1,2}, A_{2,3}, A_{3,1})$ and the group memberships $(Z_1, Z_2, Z_3)$ for the corresponding mixture component.

It is important to note that due to the latent structure of the model and since the parameters $\{\nu_0\} \cup \{\nu_{q,\ell}, q \leq \ell\}$ are supposed to be pairwise distinct, not any combination of three distributions $G_{\nu_{q,\ell}}$ is a mixture component. For example, $G_{\nu_{1,1}} \otimes G_{\nu_{2,2}} \otimes G_{\nu_{3,3}}$ is not a mixture component, since the first marginal distribution implies that node 2 belongs to group 1, i.e. $Z_2 = 1$, while the second marginal implies that node 2 belongs to group 2, i.e. $Z_2 = 2$. Indeed, to identify model parameters we heavily rely on the specific structure of the mixture distribution.

By Assumption 1 and since the parameters $\{\nu_0\} \cup \{\nu_{q,\ell}, q \leq \ell\}$ are supposed to be pairwise distinct, the mixture contains exactly $4Q^3 + 3Q(Q+1)/2 + 1$ components.

According to Theorem 1 in [48], the mixture parameters of the finite mixture $F_{K_3}$ are identifiable under the assumptions of Theorem 2. That is, we identify the parameters, say $(u_a^1, u_a^2, u_a^3) \in \mathcal{V}^3$, of each mixing component with associated mixture weight, say $p_a$. In other words, we identify the parameter set $\mathcal{B}$ given by

$$\mathcal{B} = \left\{ (u_a^1, u_a^2, u_a^3, p_a), 1 \leq a \leq 4Q^3 + 3Q(Q+1)/2 + 1 \right\},$$

where elements of $\mathcal{B}$ are unordered. Now, the first step is to identify the parameter $u_a^k$ for $k \in \{1,2,3\}$ which corresponds to the parameter $\nu_0$ of the null distribution.

First, note that the mixture distribution of $K_3$ contains exactly $Q+1$ components with i.i.d. coordinates, which are $G_{\nu_{q,q}} \otimes G_{\nu_{q,q}} \otimes G_{\nu_{q,q}}$ for $1 \leq q \leq Q$ and $G_{\nu_0} \otimes G_{\nu_0} \otimes G_{\nu_0}$. All other mixture components contain at least two coordinates with different distributions as the parameters $\nu_0$ and $\nu_{q,\ell}$ are assumed to be all distinct. In other words, we identify the parameters of the intragroup distributions $\nu_{q,q}$ and the parameter $\nu_0$ of the null with corresponding mixing weights. That is, we identify the set

$$\begin{aligned}\mathcal{C} &= \{(u_b, s_b), 1 \leq b \leq Q+1\} \\ &= \{(\nu_{q,q}, \pi_q^3 w_{q,q}^3), 1 \leq q \leq Q\} \cup \Big\{\Big(\nu_0, \sum_{q,\ell,m} \pi_q \pi_\ell \pi_m \bar{w}_{q,\ell} \bar{w}_{\ell,m} \bar{w}_{m,q}\Big)\Big\}.\end{aligned}$$

To start with, we identify the element of $\mathcal{C}$ that corresponds to the null distribution. For this, we use that the null is the only distribution that can mix with any other distribution $G_{\nu_{q,q}}$. More precisely, consider the components of the mixture distribution $F_{K_3}$ where the coordinates have two different distributions with parameters $u_b$ belonging to $\mathcal{C}$, that is, consider all mixture components of the form $G_{u_1} \otimes G_{u_1} \otimes G_{u_2}$ with $u_1, u_2 \in \mathcal{C}$ and $u_1 \neq u_2$. As stated above, it is impossible to get the component $G_{\nu_{q,q}} \otimes G_{\nu_{q,q}} \otimes G_{\nu_{\ell,\ell}}$ with $q \neq \ell$ due to the topology induced by the latent structure of the NSBM. Thus, all components of the form $G_{u_1} \otimes G_{u_1} \otimes G_{u_2}$ are either $G_{\nu_{q,q}} \otimes G_{\nu_{q,q}} \otimes G_{\nu_0}$ or $G_{\nu_0} \otimes G_{\nu_0} \otimes G_{\nu_{q,q}}$, for some $1 \leq q \leq Q$. As a result, $\nu_0$ is the only element $u \in \mathcal{C}$ such that the components $G_{u_1} \otimes G_{u_1} \otimes G_u$ and $G_{u_2} \otimes G_{u_2} \otimes G_u$ both appear in the mixture for two different elements $u_1, u_2 \in \mathcal{C}$. Hence, as $Q \geq 2$, we are able to recognize the null parameter $\nu_0$. Next, we also identify $\nu_{q,q}, 1 \leq q \leq Q$, up to label swapping of the remaining parameters $u_b$ in $\mathcal{C}$.

Now fix $1 \leq q \leq Q$. As the parameter $\nu_{q,q}$ has already been identified, we can consider the mixing weight, say $s$, associated with mixture component $G_{\nu_{q,q}} \otimes G_{\nu_{q,q}} \otimes G_{\nu_{\nu_{q,q}}}$, that corresponds to the term $\pi_q^3 w_{q,q}^3$. Similarly, the mixing weight, say $t$, associated with mixture component $G_{\nu_{q,q}} \otimes G_{\nu_{q,q}} \otimes G_{\nu_0}$ corresponds to $\pi_q^3 w_{q,q}^2 \bar{w}_{q,q}$. Hence, $s/(s+t)$ is equal to the parameter $w_{q,q}$. In this way, we identify $w_{q,q}$ for all $1 \leq q \leq Q$. Next, we note that $s^{1/3}/w_{q,q}$ equals $\pi_q$.

To summarize, at this stage we have identified $\nu_0$ and the vectors $(\nu_{q,q}, \pi_q, w_{q,q})$ for $1 \leq q \leq Q$ up to label swapping.

Now fix $1 \leq q < \ell \leq Q$. We identify parameter $\nu_{q,\ell}$ as the only parameter $u \neq \nu_0$ such that there are two components of the form $G_{\nu_{q,q}} \otimes G_u \otimes G_u$ and $G_{\nu_{\ell,\ell}} \otimes G_u \otimes G_u$ in the global mixture. To see this, note that a mixture component of the form $G_{\nu_{q,q}} \otimes G_u \otimes G_u$ with $u \neq \nu_0$ implies that the first two nodes belong to the latent group $q$, i.e. $Z_1 = Z_2 = q$. Hence, the other two coordinates are interactions of a node in group $q$ with another node. So parameter $t$ belongs to the set $\mathcal{D}_q = \{\nu_{q,m}, q \leq m \leq Q\} \cup \{\nu_{m,q}, 1 \leq m \leq q\}$. Likewise, the second and third coordinates of the component $G_{\nu_{\ell,\ell}} \otimes G_u \otimes G_u$ with $u \neq \nu_0$ describe interactions of a node in group $\ell$ with a node in another group. Thus, $u$ belongs to the set $\mathcal{D}_\ell = \{\nu_{\ell,m}, \ell \leq m \leq Q\} \cup \{\nu_{m,\ell}, 1 \leq m \leq \ell\}$. As $\mathcal{D}_q \cap \mathcal{D}_\ell = \nu_{q,\ell}$, it follows that $u = \nu_{q,\ell}$. This proves identifiability in parameter $\nu_{q,\ell}$.

Finally, since the component $G_{\nu_{q,q}} \otimes G_{\nu_{q,\ell}} \otimes G_{\nu_{q,\ell}}$ is identifiable, the associated mixing weight $y = \pi_q^2 \pi_\ell w_{q,q} w_{q,\ell}^2$ is also identifiable. Hence, parameter $w_{q,\ell}$ is given by $\left(y/(\pi_q^2 \pi_\ell w_{q,q})\right)^{1/2}$. This concludes the proof. □

Theorem 2 also holds for directed graphs. One of the differences in the proof concerns the identification of parameters $\nu_{q,\ell}$. Considering the two mixture components of the form $G_{\nu_{q,q}} \otimes G_{u_1} \otimes G_{u_2}$ and $G_{\nu_{\ell,\ell}} \otimes G_{u_2} \otimes G_{u_1}$ with $u_1, u_2 \neq \nu_0$, it follows that $u_1 = \nu_{q,\ell}$ and $u_2 = \nu_{\ell,q}$.

### *A.2. Affiliation model*

To illustrate that identifiability can be obtained in settings different from Theorem 2, we consider here an NSBM with an affiliation structure, see, e.g., [33]. In particular, we define the *affiliation-NSBM* as the NSBM where

$$w_{q,\ell} = \begin{cases} w_{\text{in}} & \text{if } q = \ell \\ w_{\text{out}} & \text{if } q \neq \ell \end{cases} ; \qquad g_{\nu_{q,\ell}} = \begin{cases} g_{\text{in}} & \text{if } q = \ell \\ g_{\text{out}} & \text{if } q \neq \ell \end{cases} .$$

Here the model parameters are $w_{\text{in}}, w_{\text{out}}$, the group proportions $\pi_q, 1 \leq q \leq Q$ and the distributions $g_0, g_{\text{in}}, g_{\text{out}}$. The following theorem shows that the affiliation-NSBM is identifiable.

**Theorem 9.** *Let $n \geq Q$ and $Q \geq 2$. Suppose that $\mathcal{G} = \{g_u, u \in \mathcal{V}\}$ satisfies Assumption 1. Let $g_0, g_{in}, g_{out} \in \mathcal{G}$ be three pairwise distinct distributions. Assume that $w_{in} < 1/2$ and $w_{out} < 1/2$. Then the parameters $w_{in}, w_{out}$ and the distributions $g_0, g_{in}, g_{out}$ of the associated affiliation-NSBM are identifiable, and the group proportions $\pi_q, 1 \leq q \leq Q$ are identifiable up to label swapping.*

*Proof.* Similarly to the proof of Theorem 2 we first study the distribution of the triplet $K_3 = (X_{1,2}, X_{2,1}, X_{3,1})$ to identify the distributions $G_0$, $G_{\text{in}}$ and $G_{\text{out}}$ and the connectivity parameters $w_{\text{in}}$ and $w_{\text{out}}$. However, to identify the group proportions $\pi_q$ we have to proceed differently.

First, note that the distribution of $K_3$ is a three-dimensional finite mixture. All mixture components have independent coordinates. The mixture has 24 components.

From the distribution of $K_3$ we identify three components with i.i.d. coordinates that correspond to $G_0 \otimes G_0 \otimes G_0$, $G_{\text{in}} \otimes G_{\text{in}} \otimes G_{\text{in}}$ and $G_{\text{out}} \otimes G_{\text{out}} \otimes G_{\text{out}}$ and corresponding mixing weights $p_1 := \bar{w}_{\text{in}}^3 \sum_{q=1}^Q \pi_q^3 + 3\bar{w}_{\text{in}} \bar{w}_{\text{out}}^2 \sum_{q \neq \ell} \pi_q^2 \pi_\ell + \bar{w}_{\text{out}}^3 \sum_{q \neq \ell \neq m} \pi_q \pi_\ell \pi_m$, $p_2 := w_{\text{in}}^3 \sum_{q=1}^Q \pi_q^3$ and $p_3 := w_{\text{out}}^3 \sum_{q \neq \ell \neq m} \pi_q \pi_\ell \pi_m$. In other words, we identify the unordered set given by $\mathcal{B} = \{(G_a, t_a), a \in [\![1, 3]\!]\} = \{(G_0, p_1), (G_{\text{in}}, p_2), (G_{\text{out}}, p_3)\}$. By the assumption $w_{\text{in}} < 1/2$ and $w_{\text{out}} < 1/2$ it is clear that $G_0$ is the distribution with largest associated mixing weight, that is $p_1 = \max\{p_1, p_2, p_3\}$. This gives us $G_0$. To decide which of the remaining distributions in $\mathcal{B}$, say $G_1$ and $G_2$, corresponds to $G_{\text{in}}$, we remark that only one of the following two components appears in the mixture of $K_3$: either $G_1 \otimes G_2 \otimes G_2$

or $G_2 \otimes G_1 \otimes G_1$. This component coincides with $G_{\text{in}} \otimes G_{\text{out}} \otimes G_{\text{out}}$. Thus, we identify the distributions $G_{\text{in}}$ and $G_{\text{out}}$.

Denote by $p_4 = w_{\text{in}}^2 \bar{w}_{\text{in}} \sum_{q=1}^{Q} \pi_q^3$ the weight of component $G_{\text{in}} \otimes G_{\text{in}} \otimes G_0$. Then we see that $p_2/(p_2 + p_4) = w_{\text{in}}$. We also derive the value of $\sum_{q=1}^{Q} \pi_q^3$ as $p_2/w_{\text{in}}^3$.

Finally, denote by $p_5$ the weight of component $G_{\text{in}} \otimes G_0 \otimes G_0$, that is $p_5 = w_{\text{in}} \bar{w}_{\text{in}}^2 \sum_{q=1}^{Q} \pi_q^3 + w_{\text{in}} \bar{w}_{\text{out}}^2 \sum_{q \neq \ell} \pi_q^2 \pi_\ell$. Moreover, let $p_6$ denote the weight of component $G_{\text{in}} \otimes G_{\text{out}} \otimes G_{\text{out}}$, that is $p_6 = w_{\text{in}} w_{\text{out}}^2 \sum_{q \neq \ell} \pi_q^2 \pi_\ell$. Then, $(p_5 - w_{\text{in}} \bar{w}_{\text{in}}^2 \sum_{q=1}^{Q} \pi_q^3)/w_{\text{in}} = \bar{w}_{\text{out}}^2 \sum_{q \neq \ell} \pi_q^2 \pi_\ell =: u$, and $p_6/(w_{\text{in}} u) = (w_{\text{out}}/\bar{w}_{\text{out}})^2 := v$, such that $\sqrt{v}/(1+\sqrt{v}) = w_{\text{out}}$. Thus, at this state we have identified $G_0$, $G_{\text{in}}$, $G_{\text{out}}$, $w_{\text{in}}$ and $w_{\text{out}}$.

Now, to identify the proportions $\pi_q$, for some $m$, let us denote by $C_m = (X_{1,2}, X_{2,3}, \ldots, X_{m-1,m}, X_{m,1})$ the edges of a circle with $m$ nodes. As above, the distribution of this $m$–dimensional vector is a mixture distribution. One of its mixture components is $\otimes_{k=1}^{m} G_{\text{in}}$ corresponding to the case where all $m$ nodes belong to the same latent group. The associated mixing weight is $w_{\text{in}}^m \sum_q \pi_q^m$. Applying this reasoning for $1 \leq m \leq Q$ provides the values $\sum_q \pi_q^m$ for all $1 \leq m \leq Q$. Now, according to Newton's identities, the values of $\pi_q, 1 \leq q \leq Q$ can be considered as the roots of a known polynomial of degree $Q$ and thus they are identifiable up to label swapping. This concludes the proof. □

From the proof we see that the assumption that $n \geq Q$ is used to identify the group proportions $\pi_q, 1 \leq q \leq Q$, which cannot be deduced from the distribution of $K_3$ as in the proof of Theorem 2.

## Appendix B: Estimation algorithm for the NSBM

The NSBM is a latent variable model, so that an EM-type algorithm may be used to approximate the maximum likelihood estimator of the model parameter $\theta = (\pi, w, \nu_0, \nu) \in \Theta$ based on observation $X$. Such an approach is developed in this section.

### B.1. Maximum likelihood estimation

In the NSBM the complete-data likelihood function $\theta \in \Theta \mapsto \mathcal{L}(X, A, Z; \theta)$ is given by

$$\begin{aligned}\mathcal{L}(X, A, Z; \theta) &= \mathcal{L}(X \mid A, Z; \nu_0, \nu)\mathcal{L}(A \mid Z; w)\mathcal{L}(Z; \pi) \\ &= \prod_{(i,j) \in \mathcal{A}} (g_{0,\nu_0}(X_{i,j}))^{1-A_{i,j}} (g_{\nu_{Z_i,Z_j}}(X_{i,j}))^{A_{i,j}} \\ &\qquad \times \prod_{(i,j) \in \mathcal{A}} w_{Z_i,Z_j}^{A_{i,j}} (1 - w_{Z_i,Z_j})^{1-A_{i,j}} \times \prod_{i=1}^{n} \pi_{Z_i}\end{aligned}$$

$$= \prod_{\substack{(i,j)\in\mathcal{A}:\\ A_{i,j}=0}} g_{0,\nu_0}(X_{i,j}) \times \prod_{q=1}^{Q}\prod_{\ell=1}^{Q} \prod_{\substack{(i,j):A_{i,j}=1\\ Z_{i,q}Z_{j,\ell}=1}} g_{\nu_{q,\ell}}(X_{i,j})$$

$$\times \prod_{1\le q\le \ell\le Q} w_{q,\ell}^{M_{q,\ell}}(1-w_{q,\ell})^{\bar{M}_{q,\ell}} \times \prod_{q=1}^{Q} \pi_q^{\sum_{i=1}^{n} Z_{i,q}}, \quad (22)$$

where $Z_{i,q} = \mathbb{1}\{Z_i = q\}$ and

$$M_{q,\ell} = \#\{(i,j)\in\mathcal{A} : A_{i,j}=1, Z_{i,q}Z_{j,\ell} + Z_{i,\ell}Z_{j,q} > 0\},$$
$$\bar{M}_{q,\ell} = \#\{(i,j)\in\mathcal{A} : A_{i,j}=0, Z_{i,q}Z_{j,\ell} + Z_{i,\ell}Z_{j,q} > 0\}.$$

The observed likelihood function $\theta \in \Theta \mapsto \mathcal{L}(X;\theta)$ is obtained from the complete-data likelihood function by integration over all possible configurations of the latent variables $(A,Z) \in \{0,1\}^{\mathcal{A}} \times \{1,\dots,Q\}^n$. In practice, this is prohibitive for any reasonable values of $n$ and $Q$ due to the size of the latter set. As a consequence, it is not possible to compute the maximum likelihood (ML) estimator by maximizing directly the observed likelihood function.

A common estimation approach is to use an EM-type algorithm. To this end, let $\mathbb{Q}$ be any distribution of the latent variables $(A,Z)$. As usual, the observed log-likelihood can be expressed as

$$\log\mathcal{L}(X;\theta) = \mathbb{E}_{\mathbb{Q}}[\log\mathcal{L}(X,A,Z;\theta)] + \mathcal{H}(\mathbb{Q}) + KL\left(\mathbb{Q}\,\|\,P_{A,Z|X;\theta}\right)$$
$$=: J(\theta,\mathbb{Q}) + KL\left(\mathbb{Q}\,\|\,P_{A,Z|X;\theta}\right),$$

where $\mathcal{H}(\mathbb{Q})$ denotes the entropy of $\mathbb{Q}$ and $KL$ the usual Kullback-Leibler divergence. It is clear that the positivity of the Kullback-Leibler divergence implies that $J$ is a lower bound of the observed log-likelihood $\log\mathcal{L}(X;\theta)$.

Now the classical EM algorithm can be viewed as a method that maximizes the lower bound $J(\theta,\mathbb{Q})$ by alternating between maximizing $J$ with respect to $\theta$ while fixing $\mathbb{Q}$ (M-step) and maximizing $J$ with respect to $\mathbb{Q}$ while fixing $\theta$ (E-step). Indeed, the latter maximization is equivalent to minimizing the Kullback-Leibler divergence $KL\left(\mathbb{Q}\,\|\,P_{A,Z|X;\theta}\right)$ with respect to $\mathbb{Q}$. Thus, if there is no contraint on $\mathbb{Q}$, the best choice for $\mathbb{Q}$ is the conditional distribution $P_{A,Z|X;\theta}$ of the latent variables $(A,Z)$ given the data $X$. This corresponds to the E-step of the classical EM algorithm.

Now, problems come in when the conditional distribution $P_{A,Z|X;\theta}$ is not tractable. In this case, a variational approximation approach can be used.

### *B.2. E-step using variational approximation*

Let $\theta \in \Theta$ be the current value of the model parameter. As mentioned before, the E-step of the classical EM-algorithm determines the conditional distribution $P_{A,Z|X;\theta}$ of the latent variables $(A,Z)$ given the observation $X$, when $(X,A,Z) \sim P_\theta$. As in the standard SBM, we encounter the difficulty that the conditional distribution $P_{A,Z|X;\theta}$ is intractable due to the involved dependence structure of the model. We thus use a mean-field approximation, that is, an approximation by a factorized probability distribution. More precisely, we denote by $\tilde{P}_{\tau,X,\theta}$ the distribution on $\{0,1\}^{\mathcal{A}} \times \{1,\dots,Q\}^n$, depending on the observed values of $X$ and the current parameter value $\theta$, such that the corresponding likelihood is of the form

$$\mathcal{L}(A,Z;\tilde{P}_{\tau,X,\theta}) = \mathcal{L}(A \mid Z,X;\theta)\prod_{i=1}^{n}\tau_{i,Z_i}, \tag{23}$$

for a parameter $\tau = (\tau_{i,q})_{i,q}$ belonging to the set

$$\mathcal{T} = \left\{\tau = (\tau_{i,q})_{1\le i\le n,1\le q\le Q} \in [0,1]^{nQ} : \sum_{q=1}^{Q}\tau_{i,q} = 1,\ \text{for all } i \in \{1,\dots,n\}\right\}.$$

Then, the variational E-step consists in maximizing the lower bound $J(\theta,\mathbb{Q})$ with respect to $\mathbb{Q}$ taken in the restricted set of distributions of the form $\tilde{P}_{\tau,X,\theta}$. This is equivalent to searching the variational parameter $\hat{\tau}$ that gives the best approximation of the conditional distribution $P_{A,Z|X;\theta}$ of $(A,Z)$ given $X$ under $P_\theta$ by a factorized distribution $\tilde{P}_{\tau,X,\theta}$ in terms of the Kullback-Leibler divergence. More precisely, for all $\theta \in \Theta$,

$$\hat{\tau} = \hat{\tau}(\theta) = \arg\min_{\tau\in\mathcal{T}} KL\left(\tilde{P}_{\tau,X,\theta} \parallel P_{A,Z|X;\theta}\right). \tag{24}$$

The following proposition states that the optimisation problem in (24) is equivalent to solving a fixed point equation, which in practice is solved numerically by an iterative algorithm.

**Proposition 1.** *For all $\theta \in \Theta$, any solution $\hat{\tau} = (\hat{\tau}_{i,q})_{i,q} \in \mathcal{T} \cap (0,1)^{nQ}$ of (24) verifies the following fixed point equation*

$$\hat{\tau}_{i,q} = C_i\pi_q \exp\left(\sum_{\substack{j=1\\ j\ne i}}^{n}\sum_{\ell=1}^{Q}\hat{\tau}_{j,\ell}\, d_{q,\ell}^{i,j}\right),\ \ i \in \{1,\dots,n\},\ \ q \in \{1,\dots,Q\},$$

*where $C_i > 0$, $i \in \{1,\dots,n\}$ are normalization constants such that $\sum_{q=1}^{Q}\hat{\tau}_{i,q} = 1$ and*

$$\rho_{q,\ell}^{i,j} = \rho_{q,\ell}^{i,j}(\theta) = \frac{w_{q,\ell}g_{\nu_{q,\ell}}(X_{i,j})}{w_{q,\ell}g_{\nu_{q,\ell}}(X_{i,j}) + (1-w_{q,\ell})g_{0,\nu_0}(X_{i,j})}, \tag{25}$$

$$\begin{aligned} d_{q,\ell}^{i,j} &= d_{q,\ell}^{i,j}(\theta) \\ &= \rho_{q,\ell}^{i,j}\left[\log g_{\nu_{q,\ell}}(X_{i,j}) + \log w_{q,\ell} - \log(\rho_{q,\ell}^{i,j})\right] \qquad (26) \\ &\quad + (1-\rho_{q,\ell}^{i,j})\left[\log g_{\nu_0}(X_{i,j}) + \log(1-w_{q,\ell}) - \log(1-\rho_{q,\ell}^{i,j})\right]. \qquad (27) \end{aligned}$$

To prove Proposition 1 we first give an explicit expression of the lower bound $J$ in the NSBM that we restate as

$$\begin{aligned} J(\theta;\mathbb{Q}) &= J(\theta;(\tau,X,\theta')) \\ &= \tilde{E}_{\tau,X,\theta'}[\log\mathcal{L}(X,A,Z;\theta)] - \tilde{E}_{\tau,X,\theta'}[\log\mathcal{L}(A,Z;\tilde{P}_{\tau,X,\theta})], \end{aligned} \tag{28}$$

where $\tilde{E}_{\tau,X,\theta'}$ denotes the expectation when $(A,Z)$ has conditional distribution $\tilde{P}_{\tau,X,\theta'}$.

**Lemma 1.** *For any $\theta,\theta'\in\Theta$, $\tau\in\mathcal{T}$, the lower bound $J(\theta;(\tau,X,\theta'))$ has the following expression*

$$\begin{aligned} J(\theta;(\tau,X,\theta')) &= \sum_{q=1}^{Q}\sum_{i=1}^{n}\tau_{i,q}\log\frac{\pi_q}{\tau_{i,q}} \\ &+ \sum_{q=1}^{Q}\sum_{l=1}^{Q}\sum_{(i,j)\in\mathcal{A}}\rho_{q,l}^{i,j}\tau_{i,q}\tau_{j,l}\left\{\log g_{\nu_{q,l}}(X_{i,j}) + \log w_{q,l} - \log(\rho_{q,\ell}^{i,j})\right\} \\ &+ \sum_{q=1}^{Q}\sum_{l=1}^{Q}\sum_{(i,j)\in\mathcal{A}}(1-\rho_{q,l}^{i,j})\tau_{i,q}\tau_{j,l}\left\{\log g_{0,\nu_0}(X_{i,j}) + \log(1-w_{q,l}) - \log(1-\rho_{q,\ell}^{i,j})\right\}. \end{aligned}$$

*where $\rho_{q,l}^{i,j}=\rho_{q,l}^{i,j}(\theta')$ is defined by (25).*

*Proof.* We have by (23) that

$$\begin{aligned} &J(\theta;(\tau,X,\theta')) = \\ &\tilde{E}_{\tau,X,\theta'}[\log\mathcal{L}(X,A,Z;\theta)] - \tilde{E}_{\tau,X,\theta'}[\log\mathcal{L}(A\mid Z,X;\theta')] - \tilde{E}_{\tau,X,\theta'}[\log(\tilde{P}_{\tau,X,\theta'}(Z))]. \end{aligned}$$

Now, by using (22), we have

$$\begin{aligned} &\tilde{E}_{\tau,X,\theta'}[\log\mathcal{L}(X,A,Z;\theta)] \\ &= \sum_{(i,j)\in\mathcal{A}}\tilde{P}_{\tau,X,\theta'}\left(A_{i,j}=0\right)\log g_{0,\nu_0}(X_{i,j}) \\ &\quad + \sum_{q=1}^{Q}\sum_{\ell=1}^{Q}\sum_{(i,j)\in\mathcal{A}}\tilde{P}_{\tau,X,\theta'}\left(A_{i,j}=1, Z_{i,q}Z_{j,\ell}=1\right)\log g_{\nu_{q,\ell}}(X_{i,j}) \\ &\quad + \sum_{q\le\ell}\tilde{E}_{\tau,X,\theta'}\left[M_{q,\ell}\right]\log w_{q,\ell} + \tilde{E}_{\tau,X,\theta'}\left[\bar{M}_{q,\ell}\right]\log(1-w_{q,\ell}) \\ &\quad + \sum_{q=1}^{Q}\sum_{i=1}^{n}\tilde{E}_{\tau,X,\theta'}\left[Z_{i,q}\right]\log\pi_q. \end{aligned}$$

This gives

$J(\theta;(\tau,X,\theta')) =$

$$
\begin{aligned}
&-\sum_{q=1}^{Q}\sum_{\ell=1}^{Q}\sum_{(i,j)\in\mathcal{A}}\tau_{i,q}\tau_{j,\ell}\left(\rho_{q,\ell}^{i,j}\log(\rho_{q,\ell}^{i,j})+(1-\rho_{q,\ell}^{i,j})\log(1-\rho_{q,\ell}^{i,j})\right)\\
&+\sum_{q=1}^{Q}\sum_{i=1}^{n}\tau_{i,q}\log\frac{\pi_q}{\tau_{i,q}}+\sum_{(i,j)\in\mathcal{A}}\log g_{0,\nu_0}(X_{i,j})\sum_{q=1}^{Q}\sum_{\ell=1}^{Q}(1-\rho_{q,\ell}^{i,j})\tau_{i,q}\tau_{j,\ell}\\
&+\sum_{q=1}^{Q}\sum_{\ell=1}^{Q}\sum_{(i,j)\in\mathcal{A}}\rho_{q,\ell}^{i,j}\tau_{i,q}\tau_{j,\ell}\log g_{\nu_{q,\ell}}(X_{i,j})\\
&+\sum_{q=1}^{Q}\sum_{\ell=1}^{Q}\log w_{q,\ell}\sum_{(i,j)\in\mathcal{A}}\rho_{q,\ell}^{i,j}\tau_{i,q}\tau_{j,\ell}+\sum_{q=1}^{Q}\sum_{\ell=1}^{Q}\log(1-w_{q,\ell})\sum_{(i,j)\in\mathcal{A}}(1-\rho_{q,\ell}^{i,j})\tau_{i,q}\tau_{j,\ell}.
\end{aligned}
$$

Rearranging terms yields the result. □

*Proof of Proposition 1.* From Lemma 1 we see that the partial derivative of $J$ with respect to $\tau_{i,q}$ is given by

$$\frac{\partial}{\partial\tau_{i,q}}J(\theta;(\tau,X,\theta'))=-\log\tau_{i,q}+\log\pi_q-1+\sum_{j\neq i}\sum_{\ell=1}^{Q}\tau_{j,\ell}d_{q,\ell}^{i,j}.$$

And the zero of this derivate satisfies

$$\tau_{i,q}=\pi_q\exp\left(\sum_{j\neq i}\sum_{\ell=1}^{Q}\tau_{j,\ell}d_{q,\ell}^{i,j}-1\right).$$

Finally, the condition $\sum_{q=1}^{Q}\tau_{i,q}=1$ yields the result. □

### B.3. M-step

Let $\tau\in\mathcal{T}$ be the current value of the variational parameter and $\theta'\in\Theta$ the current value of the model parameter. The M-step consists in updating the value of the model parameter $\theta$ by maximizing $\theta\in\Theta\mapsto J(\theta;(\tau,X,\theta'))$.

**Proposition 2** (M-step)**.** *The optimisation problem*

$$\arg\max_{\theta\in\Theta}J(\theta;(\tau,X,\theta')),\tag{29}$$

*splits into three independent problems. The solutions for $\pi$ and $w$ are given by*

$$\hat\pi_q=\frac{1}{n}\sum_{i=1}^{n}\tau_{i,q},\quad q\in\{1,\dots,Q\}\tag{30}$$

$$\hat w_{q,\ell}=\frac{\sum_{(i,j)\in\mathcal{A}}\kappa_{q,\ell}^{i,j}}{\sum_{(i,j)\in\mathcal{A}}(\tau_{i,q}\tau_{j,\ell}+\tau_{i,\ell}\tau_{j,q})},\quad q\neq\ell,\tag{31}$$

$$\hat{w}_{q,q} = \frac{\sum_{(i,j)\in\mathcal{A}} \kappa_{q,q}^{i,j}}{\sum_{(i,j)\in\mathcal{A}} \tau_{i,q}\tau_{j,q}}, \quad q \in \{1,\dots,Q\}, \tag{32}$$

*where*

$$\kappa_{q,\ell}^{i,j} = \begin{cases} (\tau_{i,q}\tau_{j,\ell} + \tau_{i,\ell}\tau_{j,q})\rho_{q,\ell}^{i,j} & \text{if } q \neq \ell; \\ \tau_{i,q}\tau_{j,q}\rho_{q,q}^{i,j} & \text{if } q = \ell, \end{cases} \tag{33}$$

*and*

$$\bar{\kappa}_{q,\ell}^{i,j} = \begin{cases} (\tau_{i,q}\tau_{j,\ell} + \tau_{i,\ell}\tau_{j,q})(1-\rho_{q,\ell}^{i,j}) & \text{if } q \neq \ell; \\ \tau_{i,q}\tau_{j,q}(1-\rho_{q,q}^{i,j}) & \text{if } q = \ell. \end{cases} \tag{34}$$

*with* $\rho_{q,\ell}^{i,j} = \rho_{q,\ell}^{i,j}(\theta')$ *defined by* (25)*. In addition, the solution of* (29) *in* $(\nu_0,\nu)$ *is given by, for* $1 \le q \le \ell \le Q$,

$$\arg\max_{\nu_0\in\mathcal{T}_0} \sum_{(i,j)\in\mathcal{A}} \log g_{0,\nu_0}(X_{i,j}) \sum_{q\le\ell} \bar{\kappa}_{q,\ell}^{i,j}, \quad \arg\max_{\nu_{q,\ell}\in\mathcal{T}} \sum_{(i,j)\in\mathcal{A}} \kappa_{q,\ell}^{i,j} \log(g_{\nu_{q,\ell}}(X_{i,j})). \tag{35}$$

*Proof.* First note that

$$\arg\max_{\theta\in\Theta} J(\theta;(\tau,X,\theta')) = \arg\max_{\theta\in\Theta} \tilde{E}_{\tau,X,\theta'}[\log\mathcal{L}(X,A,Z;\theta)]$$

and

$$\begin{aligned}\max_{\theta} \tilde{E}_{\tau,X,\theta'}\left[\log\mathcal{L}(X,A,Z;\theta)\right] &= \max_{\nu_0,\nu} \tilde{E}_{\tau,X,\theta'}\left[\log\mathcal{L}(X \mid A,Z;\nu_0,\nu)\right] \\ &\quad + \max_{w} \tilde{E}_{\tau,X,\theta'}\left[\log\mathcal{L}(A \mid Z;w)\right] + \max_{\pi} \tilde{E}_{\tau,X,\theta'}\left[\log\mathcal{L}(Z;\pi)\right].\end{aligned}$$

For the term in $\pi$ we have

$$\tilde{E}_{\tau,X,\theta'}\left[\log\mathcal{L}(Z;\pi)\right] = \sum_{q=1}^{Q} \log\pi_q \sum_{i=1}^{n} \tilde{E}_{\tau,X,\theta'}[Z_{i,q}] = \sum_{q=1}^{Q} \log\pi_q \sum_{i=1}^{n} \tau_{i,q}.$$

Taking into account the condition $\sum_{q=1}^{Q} \pi_q = 1$, we obtain that the maximum is attained at $\hat{\pi}_q$ given by (30). Concerning the optimization with respect to $w$, we have

$$\begin{aligned}&\tilde{E}_{\tau,X,\theta'}\left[\log\mathcal{L}(A \mid Z;w)\right] \\ &= \sum_{q\le\ell} \tilde{E}_{\tau,X,\theta'}\left[M_{q,\ell}\right]\log(w_{q,\ell}) + \tilde{E}_{\tau,X,\theta'}\left[\bar{M}_{q,\ell}\right]\log(1-w_{q,\ell}),\end{aligned}$$

which is maximal at

$$\hat{w}_{q,\ell} = \frac{\tilde{E}_{\tau,X,\theta'}\left[M_{q,\ell}\right]}{\tilde{E}_{\tau,X,\theta'}\left[M_{q,\ell} + \bar{M}_{q,\ell}\right]}.$$

Now, for all $q \neq \ell$,

$$\begin{aligned}
&\tilde{E}_{\tau,X,\theta'}\left[M_{q,\ell}\right] \\
&= \sum_{(i,j)\in\mathcal{A}} \tilde{P}_{\tau,X,\theta'}\left(A_{i,j}=1, Z_{i,q}Z_{j,\ell}=1\right) + \tilde{P}_{\tau,X,\theta'}\left(A_{i,j}=1, Z_{i,\ell}Z_{j,q}=1\right) \\
&= \sum_{(i,j)\in\mathcal{A}} \mathbb{P}_{\theta'}\left(A_{i,j}=1 | Z_{i,q}Z_{j,\ell}=1, X\right)\tau_{i,q}\tau_{j,\ell} \\
&\quad + \mathbb{P}_{\theta'}\left(A_{i,j}=1 | Z_{i,\ell}Z_{j,q}=1, X\right)\tau_{i,\ell}\tau_{j,q} \\
&= \sum_{(i,j)\in\mathcal{A}} \rho^{i,j}_{q,\ell}(\tau_{i,q}\tau_{j,\ell}+\tau_{i,\ell}\tau_{j,q}) = \sum_{(i,j)\in\mathcal{A}} \kappa^{i,j}_{q,\ell}.
\end{aligned}$$

Since $M_{q,\ell} + \bar{M}_{q,\ell} = \#\{(i,j)\in\mathcal{A} : Z_{i,q}Z_{j,\ell} + Z_{i,\ell}Z_{j,q} > 0\}$, we obtain

$$\begin{aligned}
\tilde{E}_{\tau,X,\theta'}\left[M_{q,\ell}+\bar{M}_{q,\ell}\right] &= \sum_{(i,j)\in\mathcal{A}} \tilde{P}_{\tau,X,\theta'}\left(Z_{i,q}Z_{j,\ell}=1\right) + \tilde{P}_{\tau,X,\theta'}\left(Z_{i,\ell}Z_{j,q}=1\right) \\
&= \sum_{(i,j)\in\mathcal{A}} \tau_{i,q}\tau_{j,\ell} + \tau_{i,\ell}\tau_{j,q}.
\end{aligned}$$

This yields the solutions given in (31)–(32).

As for $(\nu_0,\nu)$, we have

$$\begin{aligned}
&\tilde{E}_{\tau,X,\theta'}[\log \mathcal{L}(X \mid A, Z; \nu_0, \nu)] \\
&= \tilde{E}_{\tau,X,\theta'}\left[\sum_{\substack{(i,j)\in\mathcal{A}\\ A_{i,j}=0}} \log g_{0,\nu_0}(X_{i,j})\right] + \tilde{E}_{\tau,X,\theta'}\left[\sum_{q=1}^{Q}\sum_{\ell=1}^{Q}\sum_{\substack{(i,j):A_{i,j}=1\\ Z_{i,q}Z_{j,\ell}=1}} \log(g_{\nu_{q,\ell}}(X_{i,j}))\right] \\
&= \sum_{(i,j)\in\mathcal{A}} \log g_{0,\nu_0}(X_{i,j}) \sum_{q\leq\ell} \bar{\kappa}^{i,j}_{q,\ell} + \sum_{q\leq\ell}\sum_{(i,j)\in\mathcal{A}} \kappa^{i,j}_{q,\ell} \log(g_{\nu_{q,\ell}}(X_{i,j})),
\end{aligned}$$

which yields the result. □

Concerning the maximization in $(\nu_0,\nu)$, we see that the terms to maximize in (35) have the form of weighted likelihood functions. This implies that the solutions have the form of the traditional ML estimates where sample means are replaced with weighted means. For instance, in the Gaussian NSBM defined in (1), the solution of (29) in $\nu_0 = \sigma_0^2$ and $\nu_{q,\ell} = (\mu_{q,\ell}, \sigma^2_{q,\ell})$ is given by

$$\begin{aligned}
&\hat{\mu}_{q,\ell} = \frac{\sum_{(i,j)\in\mathcal{A}} \kappa^{i,j}_{q,\ell} X_{i,j}}{\sum_{(i,j)\in\mathcal{A}} \kappa^{i,j}_{q,\ell}}, \quad \hat{\sigma}^2_{q,\ell} = \frac{\sum_{(i,j)\in\mathcal{A}} \kappa^{i,j}_{q,\ell} (X_{i,j}-\hat{\mu}_{q,\ell})^2}{\sum_{(i,j)\in\mathcal{A}} \kappa^{i,j}_{q,\ell}}, \quad \forall q \leq \ell, \\
&\hat{\sigma}_0^2 = \frac{\sum_{q\leq\ell}\sum_{(i,j)\in\mathcal{A}} \bar{\kappa}^{i,j}_{q,\ell} X^2_{i,j}}{\sum_{q\leq\ell}\sum_{(i,j)\in\mathcal{A}} \bar{\kappa}^{i,j}_{q,\ell}},
\end{aligned}$$

where $\kappa^{i,j}_{q,\ell}$ and $\bar{\kappa}^{i,j}_{q,\ell}$ are given by (33) and (34).

---

**Algorithm 3:** General VEM algorithm for the NSBM

---

**Input**: Observation $X$, number $Q$ of latent groups.
**Output**: Estimator $\widehat{\theta}$, clustering $\widehat{Z}$, variational parameters $\tau$ and probabilities $\rho_{q,\ell}^{i,j}$.
Initialization of $\theta$ and $\tau$;
**while** *not converged* **do**
  VE-step: update probabilities $\rho_{q,\ell}^{i,j}(\theta)$ and update $\tau = \tau(\theta)$ by solving the fix point equation in Proposition 1;
  M-step: update $\theta$ according to Proposition 2 ;
**end**
Set $\widehat{\theta} = \theta$;
Set $\widehat{Z}_i = \arg\max_{q\in\{1,\dots,Q\}}\{\tau_{i,q}(\theta)\}$, $i \in \{1,\dots,n\}$.

---

Overall, the VEM algorithm has the form given in Algorithm 3. Its principal outputs are an estimate of the model parameter denoted $\widehat{\theta}$ and a clustering of the nodes denoted $\widehat{Z}$. In addition, the final variational parameters $\tau$ can be used as a soft clustering of the nodes. Finally, the conditional probabilities $\rho_{q,\ell}^{i,j}$ are estimates of the probabilities of the presence of an edge between $i$ and $j$ in the graph $A$ given that the interacting nodes belong to the latent blocks $q$ and $\ell$, resp.

**Remark 1** (Variational approximation does not affect $\ell$-values)**.** By definition, the conditional distribution of $A$ given $(Z, X)$ is unchanged after applying the variational approximation, see (27), because the variational approximation only acts on $(Z, X)$. In particular, computing the $\ell$-values according to $\tilde{P}_{\tau,X,\theta}$ or $\mathbb{P}_\theta$ are identical, and some of the quantities introduced in the VEM approach can be expressed in terms of the likelihood ratio functional $\boldsymbol{\ell}(x, q, \ell, \theta)$, defined by (6). More precisely, we easily check that $\rho_{q,\ell}^{i,j}(\theta) = 1 - \boldsymbol{\ell}(X_{i,j}, q, \ell, \theta)$, see (25).

### *B.4. Model selection*

In practice the number of latent blocks $Q$ is generally unknown and has to be estimated from the data. Here we use the classical integrated classification likelihood (ICL) approach [8], which can be interpreted as a penalized observed likelihood criterion, where the penalty is the sum of the traditional BIC penalty and of the entropy of the latent variable distribution. The entropy is large when the uncertainty of the underlying clustering is high, so that the quality of the obtained clustering is taken into account in the model selection procedure. More precisely, in the NSBM the ICL criterion is given by

$$\mathrm{ICL}(Q) = \tilde{\mathbb{E}}_{\hat{\tau}^{[Q]},X,\hat{\theta}^{[Q]}}[\log \mathcal{L}(X, A, Z; \hat{\theta}^{[Q]})] + \mathrm{pen}_{\mathrm{BIC}}(Q), \tag{36}$$

where $\hat{\theta}^{[Q]}$ and $\tau^{[Q]}$ are the output of the VEM algorithm with $Q$ groups, and $\mathrm{pen}_{\mathrm{BIC}}(Q)$ denotes the BIC penalty, which is (roughly) the number of model parameters multiplied with the logarithm of the number of observations. In the NSBM, the parameter $\theta$ splits into two parts: for the group proportion vector $\pi$, there are $n$ observations corresponding to the nodes, while for the other parameters $w, \nu_0, \nu$ there are $m$ observations corresponding to the observed edges, which leads to

$$\text{pen}_{\text{BIC}}(Q) = -(Q-1)\log n - \left((1+d_1)\frac{Q(Q+1)}{2} + d_0\right)\log m.$$

Now, for some given maximal number $Q_{\max}$ of groups, the number of latent groups $\widehat{Q}$ chosen by the ICL criterion is given by

$$\widehat{Q} = \arg\max_{1\le Q\le Q_{\max}} \text{ICL}(Q).$$

### B.5. R package

The VEM algorithm as presented here has been implemented in R. The R code is available on CRAN via the package `noisySBM`. Several variants of the Gaussian NSBM with varying constraints on the parameter space of the Gaussian parameters are available. Also the Gamma NSBM has been implemented.

## Appendix C: Proofs for Section 5

In this section, we prove Theorem 7 and Theorem 8. For this, we first present in Section C.1 more general non-asymptotical results, namely Theorems 10 and 11, and we prove Theorems 7 and 8 as easy corollaries. Section C.2 then presents a proof for Theorems 10 and 11, while Section C.3 introduces appropriate lemmas. Technical tools are given in Section C.4.

### C.1. More general results

To states our non-asymptotic results, we need some additional notation. Recall the assumptions of Section 5.1, relying on $\theta^* \in \Theta$, $\alpha_c = \alpha_c(\theta^*)$ given by (17) and $\alpha \in (\alpha_c, e_0)$, where $e_0 = e_0(\theta^*)$ is given by (14).

First, let for $\theta^* = (\pi^*, w^*, \nu_0^*, \nu^*)$,

$$\kappa(\theta^*,\alpha) = \sum_{q,\ell} \pi_q^*\pi_\ell^*[(1-w_{q,\ell}^*)\boldsymbol{L}_0(T_{\theta^*}((\alpha_c+\alpha)/2), q,\ell;\theta^*,\theta^*) + w_{q,\ell}^*\boldsymbol{L}_1(T_{\theta^*}((\alpha_c+\alpha)/2), q,\ell;\theta^*,\theta^*)] \in (0,1]. \tag{37}$$

Second, for convenience, we formalize a bit further Assumption (38) by defining the set of subsets of $\mathbb{R}$ containing the union of at most $M$ non-empty open intervals of $\mathbb{R}$ by

$$\mathcal{I}_M = \left\{ \bigcup_{k=1}^{M} (a_k,b_k), -\infty \le a_k \le b_k \le +\infty \text{ for } 1\le k\le M, \right.$$

$$\text{and } b_k \le a_{k+1} \text{ for } 1 \le k \le M-1 \Big\}. \qquad (38)$$

For instance, the set $(-\infty,-1) \cup (5,7)$ is in $\mathcal{I}_2$, is also in $\mathcal{I}_3$ (because empty intervals are allowed), but is not in $\mathcal{I}_1$.

Third, we avoid to rely on Assumption 5 by defining the following continuity moduli: for all $u \in (0,1)$, and for $\mathcal{K} \subset [(\alpha_c+\alpha)/2, (\alpha+e_0)/2]$ a compact interval,

$$\mathcal{W}_{\theta^*,\boldsymbol{L}}(u) = \sup_{q,\ell} \sup_{t \in T_{\theta^*}(\mathcal{K})} \sup_{\delta \in \{0,1\}} \sup \{ |\boldsymbol{L}_\delta(t,q,\ell;\theta',\theta) - \boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta^*)| \; : \; \theta,\theta' \in \Theta, \|\theta-\theta^*\|_\infty \le u, \|\theta'-\theta^*\|_\infty \le u \} ; \quad (39)$$

$$\mathcal{W}_{T,\boldsymbol{L}_1}(u) = \sup_{q,\ell} \sup \{ |\boldsymbol{L}_1(t,q,\ell;\theta^*,\theta^*) - \boldsymbol{L}_1(T_{\theta^*}(\alpha),q,\ell;\theta^*,\theta^*)| \, , \; t \in T_{\theta^*}(\mathcal{K}), |t - T_{\theta^*}(\alpha)| \le u \} ; \quad (40)$$

$$\mathcal{W}_{\alpha,T}(u) = \sup\{ |T_{\theta^*}(y) - T_{\theta^*}(\alpha)| \; : \; y \in \mathcal{K}, |y-\alpha| \le u \}. \quad (41)$$

Above, we implicitly used the generic notation "$\mathcal{W}_{x,f}$" for the modulus of the function "$f$" in the point "$x$". Note that these functions also depend on $\alpha, \theta^*$ and $\mathcal{K}$.

The first result provides the non-asymptotical behavior of the FDR of the procedure $\varphi^{\text{New}}$.

**Theorem 10.** *Let $n \ge 2$. There exist universal constants $c_1, c_1', c_2, c_2' > 0$ such that the following holds. Let Assumptions 3 and 4 (with some constant $M$) be true and let $\theta^* = (\pi^*, w^*, \nu_0^*, \nu^*) \in \Theta$. Consider $\alpha_c = \alpha_c(\theta^*)$ given by (17), $\alpha \in (\alpha_c, e_0)$, $\kappa = \kappa(\theta^*,\alpha)$ given by (37), $\mathcal{K} \subset [(\alpha_c+\alpha)/2, (\alpha+e_0)/2]$ a compact interval, and the modulus $\mathcal{W}_{\theta^*,\boldsymbol{L}}$ defined by (39). Let $\pi^*_{min} = \min_q\{\pi^*_q\}$ and $w^*_{max} = \max_{q,\ell}\{w^*_{q,\ell}\}$. Consider the procedure $\varphi^{New}$ of Algorithm 2 for some estimator $\widehat{\theta}$ and clustering $\widehat{Z}$ having for risk probability $\eta(\theta^*,\cdot)$ defined by (18). Then there exists $\epsilon = \epsilon(\theta^*,\alpha,\mathcal{K}) \in (0,1)$ such that for all $\varepsilon \in (0,\epsilon)$, for all $x > 0$ with $x < (\pi^*_{min})^2 \wedge (1 - w^*_{max})$,*

$$\mathrm{FDR}_{\theta^*}(\varphi^{New}) \le \alpha + x + 16\kappa^{-1}(\mathcal{W}_{\theta^*,\boldsymbol{L}}(\varepsilon) + 3Q^2\varepsilon) + \eta(\theta^*,\varepsilon) + c_1 Q^2 \mathrm{e}^{-c_1' n\kappa^2 x^2/Q^4} + c_2 Q^2 \mathrm{e}^{-c_2' n^2 (\pi^*_{min})^2 (1-w^*_{max})\kappa^2 x^2/M^2}.$$

The second result provides the non-asymptotical behavior of the TDR of the procedure $\varphi^{\text{New}}$.

**Theorem 11.** *Let $n \ge 2$. There exist universal constants $c_1, c_1', c_2, c_2' > 0$ such that the following holds. Consider the setting of Theorem 10 and additionally let $w^*_{min} = \min_{q,\ell}\{w^*_{q,\ell}\}$ and $e_1 = 1 - e_0$. Consider the functions $\mathcal{W}_{T,\boldsymbol{L}_1}$, $\mathcal{W}_{\alpha,T}$ given respectively by (40),(41) and the optimal procedure $\varphi^*$ defined by (10). Then there exists $\epsilon = \epsilon(\theta^*,\alpha,\mathcal{K}) \in (0,1)$ such that for all $\varepsilon \in (0,\epsilon)$, for all $x > 0$ with $x < (\pi^*_{min})^2 \wedge w^*_{min}$,*

$$e_1 \mathrm{TDR}_{\theta^*}(\varphi^{New}) \ge e_1 \mathrm{TDR}_{\theta^*}(\varphi^*) - x - e_1 \eta(\theta^*,\varepsilon)$$

$$- 2\mathcal{W}_{\theta^*,\boldsymbol{L}}(\varepsilon) - 6Q^2\varepsilon - \mathcal{W}_{T,\boldsymbol{L}_1} \circ \mathcal{W}_{\alpha,T}\left(8\kappa^{-1}(\mathcal{W}_{\theta^*,\boldsymbol{L}}(\varepsilon) + 3Q^2\varepsilon)\right)$$
$$- c_1 e_1 Q^2 \mathrm{e}^{-c_1' n x^2/Q^4} - c_2 e_1 Q^2 \mathrm{e}^{-c_2' n^2 (\pi^*_{min})^2 w^*_{min} x^2/M^2}.$$

*Proof of Theorem 7.* By using Theorems 10 and 11, and since $\widehat{\theta}$, $\widehat{Z}$ are consistent, we have for all $\varepsilon \in (0, \epsilon)$, and $x \in (0, (\pi^*_{min})^2 \wedge (1 - w^*_{max}))$,

$$\limsup_n \mathrm{FDR}_{\theta^*}(\varphi^{\mathrm{New}}) \leq \alpha + x + 16\kappa^{-1}(\mathcal{W}_{\theta^*,\boldsymbol{L}}(\varepsilon) + 3Q^2\varepsilon)$$

and

$$e_1 \liminf_n \{\mathrm{TDR}_{\theta^*}(\varphi^{\mathrm{New}}) - \mathrm{TDR}_{\theta^*}(\varphi^*)\}$$
$$\geq -x - 2\mathcal{W}_{\theta^*,\boldsymbol{L}}(\varepsilon) - 6Q^2\varepsilon - \mathcal{W}_{T,\boldsymbol{L}_1} \circ \mathcal{W}_{\alpha,T}\left(8\kappa^{-1}(\mathcal{W}_{\theta^*,\boldsymbol{L}}(\varepsilon) + 3Q^2\varepsilon)\right).$$

Now, by Assumption 3, the functions $\mathcal{W}_{\theta^*,\boldsymbol{L}}$, $\mathcal{W}_{T,\boldsymbol{L}_1}$ have both a zero limit in zero. In addition, $\mathcal{W}_{\alpha,T}$ also has a zero limit in zero (because $y \in \mathcal{K} \mapsto T_{\theta^*}(y)$ is continuous by Lemma 2). Hence, taking $\varepsilon$ and $x$ tending to 0 gives the result. □

*Proof of Theorem 8.* By Assumption 5 (i), we have $\mathcal{W}_{\theta^*,\boldsymbol{L}}(\varepsilon) \leq C_2\varepsilon$. Hence, (19) easily derives from Theorem 10. Relation (20) is proved similarly from Assumption 5 (ii)-(iii) and Theorem 11. □

### *C.2. Proving Theorems 10 and 11*

First, note that by (15), the optimal procedure $\varphi^*$ and our procedure $\varphi^{\mathrm{New}}$ can be equivalently written as $\ell$-value thresholding procedures, that is, for $(i,j) \in \mathcal{A}$,

$$\varphi^*_{i,j} = \mathbb{1}\{\ell_{i,j}(X, Z, \theta^*) \leq T_{\theta^*}(\alpha)\} = \mathbb{1}\{\mathrm{MFDR}_{\theta^*}(\theta^*, \ell_{i,j}(X, Z, \theta^*)) \leq \alpha\};$$
$$\varphi^{\mathrm{New}}_{i,j} = \mathbb{1}\{\ell_{i,j}(X, \widehat{Z}, \widehat{\theta}) \leq T_{\widehat{\theta}}(\alpha)\} = \mathbb{1}\{\mathrm{MFDR}_{\widehat{\theta}}(\widehat{\theta}, \ell_{i,j}(X, \widehat{Z}, \widehat{\theta})) \leq \alpha\},$$

for which we recall that $\theta^*$ is the true value of the parameter.

Next, recall that on the event

$$\mathcal{E} = \mathcal{E}(\theta^*, \varepsilon) = \left\{\|\widehat{\theta} - \theta^*\| \leq \varepsilon \text{ and } \widehat{Z} = Z \text{ up to label switching}\right\}, \quad (42)$$

there exists some permutation $\sigma$ of $\{1, \ldots, Q\}$ such that both $\sigma(Z) = \widehat{Z}$ and $\|\widehat{\theta}^\sigma - \theta^*\|_\infty \leq \varepsilon$, where $\widehat{\theta}^\sigma = (\hat{\pi}^\sigma, \hat{w}^\sigma, \hat{\nu}_0, \hat{\nu}^\sigma)$ for $\hat{\pi}^\sigma = (\hat{\pi}_{\sigma(q)})_{1\leq q\leq Q}$, $\hat{w}^\sigma = (\hat{w}_{\sigma(q),\sigma(\ell)})_{1\leq q,\ell\leq Q}$, $\hat{\nu}^\sigma = (\hat{\nu}_{\sigma(q),\sigma(\ell)})_{1\leq q,\ell\leq Q}$. Hence, we have $\ell_{i,j}(X, \widehat{Z}; \widehat{\theta}) = \ell_{i,j}(X, \sigma(Z); \widehat{\theta}) = \ell_{i,j}(X, Z; \widehat{\theta}^\sigma)$ and $\mathrm{MFDR}_{\hat{\theta}^\sigma}(\widehat{\theta}^\sigma, t) = \mathrm{MFDR}_{\hat{\theta}}(\widehat{\theta}, t)$. This gives $\varphi^{\mathrm{New}} = \varphi^Z$, where $\varphi^Z$ denotes the procedure defined by

$$\varphi^Z_{i,j} = \mathbb{1}\{\mathrm{MFDR}_{\hat{\theta}^\sigma}(\widehat{\theta}^\sigma, \ell_{i,j}(X, Z, \widehat{\theta}^\sigma)) \leq \alpha\} = \mathbb{1}\{\ell_{i,j}(X, Z, \widehat{\theta}^\sigma) \leq T_{\hat{\theta}^\sigma}(\alpha)\}. \quad (43)$$

The latter is easier to study than $\varphi^{\mathrm{New}}$ because $\|\widehat{\theta}^\sigma - \theta^*\|_\infty \leq \varepsilon$ and $Z$ is the true clustering. In the sequel, for any $\theta$, we denote the $\ell$-values $\ell_{i,j}(X, Z; \theta)$ (5) by $\ell_{i,j}(\theta)$ for short.

Third, let us introduce additional useful notation. Define for any $t \in [0,1]$, $\theta \in \Theta$,

$$\widehat{F}_0(\theta,t) = m^{-1} \sum_{(i,j)\in\mathcal{A}} (1 - A_{i,j}) \mathbb{1}\{\ell_{i,j}(\theta) \leq t\}; \tag{44}$$

$$\widehat{F}_1(\theta,t) = m^{-1} \sum_{(i,j)\in\mathcal{A}} A_{i,j} \mathbb{1}\{\ell_{i,j}(\theta) \leq t\}; \tag{45}$$

$$\widehat{F}(\theta,t) = \widehat{F}_0(\theta,t) + \widehat{F}_1(\theta,t) = m^{-1} \sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{\ell_{i,j}(\theta) \leq t\}; \tag{46}$$

$$\mathrm{FDP}(\theta,t) = \frac{\widehat{F}_0(\theta,t)}{\widehat{F}(\theta,t)}. \tag{47}$$

Also let for $\theta' = (\pi', w', \nu_0', \nu') \in \Theta$,

$$F_{0,\theta'}(\theta,t) = \mathbb{E}_{\theta'}[\widehat{F}_0(\theta,t)] = \sum_{q,\ell} \pi_q' \pi_\ell' (1 - w_{q,\ell}')\, \boldsymbol{L}_0(t,q,\ell;\theta',\theta); \tag{48}$$

$$F_{1,\theta'}(\theta,t) = \mathbb{E}_{\theta'}[\widehat{F}_1(\theta,t)] = \sum_{q,\ell} \pi_q' \pi_\ell' w_{q,\ell}'\, \boldsymbol{L}_1(t,q,\ell;\theta',\theta); \tag{49}$$

$$F_{\theta'}(\theta,t) = F_{0,\theta'}(\theta,t) + F_{1,\theta'}(\theta,t). \tag{50}$$

Note that $\mathrm{MFDR}_{\theta^*}(\theta,t)$ defined in (7) is thus such that

$$\mathrm{MFDR}_{\theta^*}(\theta,t) = \frac{F_{0,\theta^*}(\theta,t)}{F_{\theta^*}(\theta,t)}. \tag{51}$$

Also, note that $\kappa(\theta^*,\alpha)$ defined in (37) is thus equal to $F_{\theta^*}(\theta^*, T_{\theta^*}((\alpha_c+\alpha)/2))$.

*Proof of Theorem 10.* First observe that by using $\varphi^{\text{New}} = \varphi^{z}$ and (43), we have

$$\mathrm{FDR}_{\theta^*}(\varphi^{\text{New}}) \leq \mathbb{E}_{\theta^*}\left[\mathrm{FDP}(\hat{\theta}^\sigma, T_{\hat{\theta}^\sigma}(\alpha))\mathbb{1}_{\mathcal{E}}\right] + \eta(\theta^*,\varepsilon),$$

where $\sigma$ denotes any (data dependent) permutation of $\{1,\dots,Q\}$ such that both $\sigma(Z) = \widehat{Z}$ and $\|\widehat{\theta}^\sigma - \theta^*\|_\infty \leq \varepsilon$. This permutation exists on the event $\mathcal{E}$ by definition (42). Now, let $x > 0$ with $x < (\pi^*_{min})^2 \wedge (1 - w^*_{max})$ and $y > 0$ and consider the event

$$\Omega = \Omega(\theta^*, x, y) = \left\{ \sup_{\substack{\theta\in\Theta, t\in[0,1]\\ F_{\theta^*}(\theta,t)\geq y}} |\mathrm{FDP}(\theta,t) - \mathrm{MFDR}_{\theta^*}(\theta,t)| \leq x \right\}. \tag{52}$$

We have

$$\mathbb{E}_{\theta^*}\left[\mathrm{FDP}(\hat{\theta}^\sigma, T_{\hat{\theta}^\sigma}(\alpha))\mathbb{1}_{\mathcal{E}}\right] \leq \mathbb{E}_{\theta^*}\left[\mathrm{FDP}(\hat{\theta}^\sigma, T_{\hat{\theta}^\sigma}(\alpha))\mathbb{1}_{\Omega}\mathbb{1}_{\mathcal{E}}\right] + \mathbb{P}_{\theta^*}(\Omega^c).$$

Now, applying Lemma 3 (56) (with the definition of $v(\varepsilon)$ therein), there exists $\epsilon = \epsilon(\theta^*,\alpha,\mathcal{K}) \in (0,1)$ such that for all $\varepsilon \leq \epsilon$, if $\|\hat{\theta}^\sigma - \theta^*\|_\infty \leq \varepsilon$ then

$T_{\theta^*}(\min \mathcal{K}) \leq T_{\theta^*}(\alpha - \upsilon(\varepsilon)) \leq T_{\hat{\theta}^\sigma}(\alpha) \leq T_{\theta^*}(\alpha + \upsilon(\varepsilon)) \leq T_{\theta^*}(\max \mathcal{K})$. In particular, $F_{\theta^*}(\hat{\theta}^\sigma, T_{\hat{\theta}^\sigma}(\alpha)) \geq F_{\theta^*}(\hat{\theta}^\sigma, T_{\theta^*}(\min \mathcal{K})) \geq \kappa(1 - \upsilon(\varepsilon)/4)$, by applying Lemma 3 (54). Hence, choosing $y = \kappa/2$ so that $y \leq \kappa(1-\upsilon(\varepsilon)/4)$ (which holds by choosing $\epsilon$ small enough), we get by definition of $\Omega$,

$$
\begin{aligned}
&\mathbb{E}_{\theta^*}\left[\mathrm{FDP}(\hat{\theta}^\sigma, T_{\hat{\theta}^\sigma}(\alpha))\mathbb{1}_\Omega \mathbb{1}_\mathcal{E}\right] \\
&\leq x + \mathbb{E}_{\theta^*}\left[\mathrm{MFDR}_{\theta^*}(\hat{\theta}^\sigma, T_{\hat{\theta}^\sigma}(\alpha))\mathbb{1}_\mathcal{E}\right] \\
&\leq x + \mathbb{E}_{\theta^*}\left[\mathrm{MFDR}_{\theta^*}(\hat{\theta}^\sigma, T_{\theta^*}(\alpha + \upsilon(\varepsilon)))\mathbb{1}_\mathcal{E}\right] \\
&\leq x + \mathrm{MFDR}_{\theta^*}(\theta^*, T_{\theta^*}(\alpha + \upsilon(\varepsilon))) + \upsilon(\varepsilon) = x + \alpha + 2\upsilon(\varepsilon),
\end{aligned}
$$

by applying (55). This gives

$$\mathrm{FDR}_{\theta^*}(\varphi^{\mathrm{New}}) \leq \alpha + x + 2\upsilon(\varepsilon) + \mathbb{P}_{\theta^*}(\Omega^c) + \eta(\theta^*, \varepsilon).$$

We conclude by upper bounding $\mathbb{P}_{\theta^*}(\Omega^c)$ according to Lemma 4. □

*Proof of Theorem 11.* First observe that, similarly to the proof of Theorem 10 (and using the same notation for the permutation $\sigma$), we have

$$e_1 \mathrm{TDR}_{\theta^*}(\varphi^{\mathrm{New}}) \geq \mathbb{E}_{\theta^*}\left[\widehat{F}_1(\widehat{\theta}^\sigma, T_{\hat{\theta}^\sigma}(\alpha))\mathbb{1}_\mathcal{E}\right].$$

Also observe that for the procedure $\varphi^*$ given by (10), we have

$$e_1 \mathrm{TDR}_{\theta^*}(\varphi^*) = \mathbb{E}_{\theta^*}\left[\widehat{F}_1(\theta^*, T_{\theta^*}(\alpha))\right] = F_{1,\theta^*}(\theta^*, T_{\theta^*}(\alpha)).$$

For all $x > 0$ with $x < (\pi^*_{min})^2 \wedge w^*_{min}$, consider the event

$$\Omega_1 = \Omega_1(\theta^*, x) = \left\{ \sup_{\theta \in \Theta, t \in [0,1]} \left| \widehat{F}_1(\theta, t) - F_{1,\theta^*}(\theta, t) \right| \leq x \right\}.$$

We obviously have

$$
\begin{aligned}
\mathbb{E}_{\theta^*}\left[\widehat{F}_1(\widehat{\theta}^\sigma, T_{\hat{\theta}^\sigma}(\alpha))\mathbb{1}_\mathcal{E}\right] &\geq \mathbb{E}_{\theta^*}\left[\widehat{F}_1(\widehat{\theta}^\sigma, T_{\hat{\theta}^\sigma}(\alpha))\mathbb{1}_{\Omega_1}\mathbb{1}_\mathcal{E}\right] \\
&\geq \mathbb{E}_{\theta^*}\left[F_{1,\theta^*}(\widehat{\theta}^\sigma, T_{\hat{\theta}^\sigma}(\alpha))\mathbb{1}_{\Omega_1}\mathbb{1}_\mathcal{E}\right] - x.
\end{aligned}
$$

Now, by applying Lemma 3 (with the definition of $\upsilon(\varepsilon)$), there exists $\epsilon = \epsilon(\theta^*, \alpha, \mathcal{K}) \in (0,1)$ such that for all $\varepsilon \leq \epsilon$, if $\|\hat{\theta}^\sigma - \theta^*\|_\infty \leq \varepsilon$ then $T_{\theta^*}(\min \mathcal{K}) \leq T_{\theta^*}(\alpha - \upsilon(\varepsilon)) \leq T_{\hat{\theta}^\sigma}(\alpha) \leq T_{\theta^*}(\alpha + \upsilon(\varepsilon)) \leq T_{\theta^*}(\max \mathcal{K})$ and

$$
\begin{aligned}
&\mathbb{E}_{\theta^*}\left[F_{1,\theta^*}(\widehat{\theta}^\sigma, T_{\hat{\theta}^\sigma}(\alpha))\mathbb{1}_{\Omega_1}\mathbb{1}_\mathcal{E}\right] \\
&\geq \mathbb{E}_{\theta^*}\left[(F_{1,\theta^*}(\theta^*, T_{\hat{\theta}^\sigma}(\alpha)) - \kappa\upsilon(\varepsilon)/4)\mathbb{1}_{\Omega_1}\mathbb{1}_\mathcal{E}\right] \\
&\geq F_{1,\theta^*}(\theta^*, T_{\theta^*}(\alpha - \upsilon(\varepsilon))) - \kappa\upsilon(\varepsilon)/4 - e_1\mathbb{P}_{\theta^*}(\Omega_1^c) - e_1\eta(\theta^*, \varepsilon),
\end{aligned}
$$

because $F_{1,\theta^*}(\theta^*, T_{\hat{\theta}^\sigma}(\alpha)) \leq e_1$ pointwise. Now using the functions $\mathcal{W}_{T,\boldsymbol{L}_1}$ and $\mathcal{W}_{\alpha,T}$ defined by (40) and (41), respectively, we have by (57),

$$\begin{aligned} F_{1,\theta^*}(\theta^*, T_{\theta^*}(\alpha - \upsilon(\varepsilon))) &\geq F_{1,\theta^*}(\theta^*, T_{\theta^*}(\alpha)) - \mathcal{W}_{T,\boldsymbol{L}_1}\left(T_{\theta^*}(\alpha) - T_{\theta^*}(\alpha - \upsilon(\varepsilon))\right) \\ &\geq F_{1,\theta^*}(\theta^*, T_{\theta^*}(\alpha)) - \mathcal{W}_{T,\boldsymbol{L}_1} \circ \mathcal{W}_{\alpha,T}\left(\upsilon(\varepsilon)\right). \end{aligned}$$

Using Lemma 5 to upper-bound $\mathbb{P}_{\theta^*}(\Omega_1^c)$ concludes the proof. □

### *C.3. Main lemmas for Section 5*

**Lemma 2.** *Let Assumption 3 be true and consider any $\theta \in \Theta$ with the corresponding quantities $t_{1,q,\ell}(\theta), t_{2,q,\ell}(\theta)$, $q, \ell \in \{1,\ldots,Q\}^2$, $t_1(\theta) = \min_{q,\ell} t_{1,q,\ell}(\theta)$ and $t_2(\theta) = \max_{q,\ell} t_{2,q,\ell}(\theta)$. Then the function $t \mapsto \mathrm{MFDR}_\theta(\theta,t)$ is increasing on $[t_1(\theta), t_2(\theta)]$, continuous on $(t_1(\theta), 1]$, satisfies $\mathrm{MFDR}_\theta(\theta,t) = 0$ for $t \in [0, t_1(\theta)]$, $\mathrm{MFDR}_\theta(\theta,t) = e_0$ for $t \in [t_2(\theta), 1]$ and $\mathrm{MFDR}_\theta(\theta,t) < t$ for $t \in (t_1(\theta), 1]$.*

*Proof.* First note that the following relation holds (coming from (5), (7) and Fubini's theorem): for all $\theta \in \Theta$, $t \in [0,1]$,

$$\mathrm{MFDR}_\theta(\theta,t) = \frac{\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \ell_{i,j}(\theta)\mathbb{1}\{\ell_{i,j}(\theta) \leq t\}\right]}{\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{\ell_{i,j}(\theta) \leq t\}\right]}. \tag{53}$$

In the sequel, denote respectively $\mathrm{MFDR}_\theta(\theta,t)$ by $f(t)$ and $\ell_{i,j}(\theta)$ by $\ell_{i,j}$ for short. Note that by Assumption 3, we have $f(t) = 0$ for $t \in [0,t_1]$, $f(t) = e_0$ for $t \in [t_2, 1]$. Notice that the $\ell_{i,j}$ are continuous random variables because their c.d.f.'s are continuous by Assumption 3. Thus, (53) yields

$$f(t) = \frac{F_{0,\theta}(\theta,t)}{F_\theta(\theta,t)} = \frac{\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \ell_{i,j}\mathbb{1}\{\ell_{i,j} < t\}\right]}{\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{\ell_{i,j} < t\}\right]}, \text{ for } t \in [0,1].$$

As a result, by the dominated convergence theorem, $f$ is both left-continuous and right-continuous in any $t$ such that $F_\theta(\theta,t) > 0$.

Now prove that $f$ is increasing on $[t_1, t_2]$. For this, let $t_1 \leq t < t' \leq t_2$ and prove $f(t') > f(t)$. If $F_\theta(\theta,t) = 0$, then $f(t) = 0$. Since $F_{0,\theta}(\theta,t') > 0$ and $F_\theta(\theta,t') > 0$, we have $f(t') > 0 = f(t)$. Now assume $F_\theta(\theta,t) > 0$, so that $F_\theta(\theta,t') > 0$ also holds. We let

$$\delta = \frac{F_\theta(\theta,t')}{F_\theta(\theta,t)} - 1 = \frac{\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{t < \ell_{i,j} \leq t'\}\right]}{\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{\ell_{i,j} \leq t\}\right]}.$$

Now, we have

$$\left(f(t') - f(t)\right) F_\theta(\theta,t')$$

$$
\begin{aligned}
&= \mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \ell_{i,j}\mathbb{1}\{\ell_{i,j}\le t'\}\right] - (1+\delta)\,\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \ell_{i,j}\mathbb{1}\{\ell_{i,j}\le t\}\right]\\
&= \mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \ell_{i,j}\mathbb{1}\{t<\ell_{i,j}\le t'\}\right] - \delta\,\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \ell_{i,j}\mathbb{1}\{\ell_{i,j}\le t\}\right]\\
&\ge t\,\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{t<\ell_{i,j}\le t'\}\right] - \delta\,\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \ell_{i,j}\mathbb{1}\{\ell_{i,j}\le t\}\right]\\
&\ge t\left(\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{t<\ell_{i,j}\le t'\}\right] - \delta\,\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{\ell_{i,j}\le t\})\right]\right) = 0.
\end{aligned}
$$

Now, since $F_\theta(\theta,t')>0$, this entails $f(t')\ge f(t)$. Also, if $f(t')=f(t)$, the inequalities above are all equalities and we have

$$
\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \ell_{i,j}\mathbb{1}\{t<\ell_{i,j}\le t'\}\right] = t\,\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{t<\ell_{i,j}\le t'\}\right]
$$

and thus

$$
\mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} (\ell_{i,j}-t)\mathbb{1}\{t<\ell_{i,j}\le t'\}\right] = 0
$$

which gives $(\ell_{i,j}-t)\mathbb{1}\{t<\ell_{i,j}\le t'\}=0$ $\mathbb{P}_\theta$-a.s. for all $(i,j)$, which is impossible because $\ell_{i,j}$ is a continuous random variable with an increasing c.d.f. on $[t,t']$ (Assumption 3). Hence, $f(t')>f(t)$ and the increasingness of $f$ is proved.

Finally, let $t\in(t_1,1]$ and prove that $t>f(t)$.

$$
(t-f(t))F_\theta(\theta,t) = \mathbb{E}_\theta\left[\sum_{(i,j)\in\mathcal{A}} (t-\ell_{i,j})\mathbb{1}\{\ell_{i,j}<t\}\right] \ge 0
$$

and thus $t\ge f(t)$. Moreover, $t=f(t)$ entails $(t-\ell_{i,j})\mathbb{1}\{\ell_{i,j}\le t\}=0$ $\mathbb{P}_\theta$-a.s. for all $(i,j)$, which is impossible because $\ell_{i,j}$ is a continuous random variable with an increasing c.d.f. on $[t_1,t]$ (Assumption 3). Hence $t>f(t)$. □

**Lemma 3.** *Let Assumption 3 be true, $\theta^*\in\Theta$, $\alpha_c=\alpha_c(\theta^*)$ given by (17), $\alpha\in(\alpha_c,e_0)$, $\kappa=\kappa(\theta^*,\alpha)$ given by (37), $\mathcal{K}\subset[(\alpha_c+\alpha)/2,(\alpha+e_0)/2]$ a compact interval, the modulus $\mathcal{W}_{\theta^*,\boldsymbol{L}}$ defined by (39) and the modulus $\mathcal{W}_{T,\boldsymbol{L}_1}$ defined by (40). Then there exists $\epsilon=\epsilon(\theta^*,\alpha,\mathcal{K})\in(0,1)$ such that for all $\varepsilon\le\epsilon$, letting $\upsilon(\varepsilon)=8\kappa^{-1}(\mathcal{W}_{\theta^*,\boldsymbol{L}}(\varepsilon)+3Q^2\varepsilon)$, we have $\min\mathcal{K}\le\alpha-\upsilon(\varepsilon)$ and $\alpha+\upsilon(\varepsilon)\le\max\mathcal{K}$ and for any $\theta,\theta'\in\Theta$ with $\|\theta-\theta^*\|_\infty\le\varepsilon$ and $\|\theta'-\theta^*\|_\infty\le\varepsilon$, we have*

$$
\begin{aligned}
\sup_{t\in T_{\theta^*}(\mathcal{K})} &|F_{0,\theta'}(\theta,t)-F_{0,\theta^*}(\theta^*,t)| \vee |F_{1,\theta'}(\theta,t)-F_{1,\theta^*}(\theta^*,t)|\\
&\vee |F_{\theta'}(\theta,t)-F_{\theta^*}(\theta^*,t)| \le 2\mathcal{W}_{\theta^*,\boldsymbol{L}}(\varepsilon)+6Q^2\varepsilon=\kappa\upsilon(\varepsilon)/4; \quad (54)
\end{aligned}
$$

$$\sup_{t\in T_{\theta^*}(\mathcal{K})} |\mathrm{MFDR}_{\theta'}(\theta,t) - \mathrm{MFDR}_{\theta^*}(\theta^*,t)| \le \upsilon(\varepsilon); \tag{55}$$

$$T_{\theta^*}(\min\mathcal{K}) \le T_{\theta^*}(\alpha-\upsilon(\varepsilon)) \le T_\theta(\alpha) \le T_{\theta^*}(\alpha+\upsilon(\varepsilon)) \le T_{\theta^*}(\max\mathcal{K}). \tag{56}$$

*Moreover, for* $t\in T_{\theta^*}(\mathcal{K})$ *with* $|t-T_{\theta^*}(\alpha)|\le\varepsilon$, *we have*

$$|F_{1,\theta^*}(\theta^*,t) - F_{1,\theta^*}(\theta^*,T_{\theta^*}(\alpha))| \le \mathcal{W}_{T,\boldsymbol{L}_1}(\varepsilon). \tag{57}$$

*Proof.* To prove (54), we have by (39) and (49), for all $t\in T_{\theta^*}(\mathcal{K})$, $\theta\in\Theta$, $\theta^*=(\pi^*,w^*,\nu_0^*,\nu^*)$, $\theta'=(\pi',w',\nu_0',\nu')$,

$$\begin{aligned}
&|F_{1,\theta'}(\theta,t) - F_{1,\theta^*}(\theta^*,t)|\\
&= \left|\sum_{q,\ell}\pi'_q\pi'_\ell w'_{q,\ell}\,\boldsymbol{L}_1(t,q,\ell;\theta',\theta) - \sum_{q,\ell}\pi^*_q\pi^*_\ell w^*_{q,\ell}\,\boldsymbol{L}_1(t,q,\ell;\theta^*,\theta^*)\right|\\
&\le \sum_{q,\ell}\pi'_q\pi'_\ell w'_{q,\ell}\left|\boldsymbol{L}_1(t,q,\ell;\theta',\theta) - \frac{\pi^*_q\pi^*_\ell w^*_{q,\ell}}{\pi'_q\pi'_\ell w'_{q,\ell}}\,\boldsymbol{L}_1(t,q,\ell;\theta^*,\theta^*)\right|\\
&\le \sup_{q,\ell}|\boldsymbol{L}_1(t,q,\ell;\theta',\theta) - \boldsymbol{L}_1(t,q,\ell;\theta^*,\theta^*)| + \sum_{q,\ell}\left|\pi'_q\pi'_\ell w'_{q,\ell} - \pi^*_q\pi^*_\ell w^*_{q,\ell}\right|\\
&\le \mathcal{W}_{\theta^*,\boldsymbol{L}}(\varepsilon) + 3Q^2\varepsilon.
\end{aligned}$$

Similarly, the latter bound is also valid for $|F_{0,\theta'}(\theta,t) - F_{0,\theta^*}(\theta^*,t)|$. Since we have $F_{\theta'}(\theta,t) = F_{0,\theta'}(\theta,t)+F_{1,\theta'}(\theta,t)$, this proves (54). The proof of (57) follows similarly.

Let us now establish (55). First, by (54), and since $F_{\theta^*}(\theta^*,T_{\theta^*}(\min\mathcal{K}))\ge\kappa>0$, we have for all $t\in T_{\theta^*}(\mathcal{K})$, $F_{\theta'}(\theta,t)\ge\kappa/2$ by choosing $\epsilon=\epsilon(\theta^*,\alpha)$ small enough. Hence,

$$\begin{aligned}
&|\mathrm{MFDR}_{\theta'}(\theta,t) - \mathrm{MFDR}_{\theta^*}(\theta^*,t)|\\
&= \left|\frac{F_{0,\theta'}(\theta,t)}{F_{\theta'}(\theta,t)} - \frac{F_{0,\theta^*}(\theta^*,t)}{F_{\theta^*}(\theta^*,t)}\right|\\
&\le \left|\frac{F_{0,\theta'}(\theta,t)}{F_{\theta'}(\theta,t)} - \frac{F_{0,\theta^*}(\theta^*,t)}{F_{\theta'}(\theta,t)}\right| + F_{0,\theta^*}(\theta^*,t)\left|\frac{1}{F_{\theta'}(\theta,t)} - \frac{1}{F_{\theta^*}(\theta^*,t)}\right|\\
&\le \frac{|F_{0,\theta'}(\theta,t) - F_{0,\theta^*}(\theta^*,t)|}{\kappa/2} + \frac{F_{0,\theta^*}(\theta^*,t)}{F_{\theta^*}(\theta^*,t)}\frac{|F_{\theta'}(\theta,t) - F_{\theta^*}(\theta^*,t)|}{\kappa/2}\\
&\le 4\kappa^{-1}(2\mathcal{W}_{\theta^*,\boldsymbol{L}}(\varepsilon) + 6Q^2\varepsilon),
\end{aligned}$$

which proves (55) by using again (54).

Let us finally prove (56). The relation (55), used with $\theta=\theta'$ gives for all $t\in T_{\theta^*}(\mathcal{K})$,

$$\mathrm{MFDR}_{\theta^*}(\theta^*,t) - \upsilon(\varepsilon) \le \mathrm{MFDR}_\theta(\theta,t) \le \mathrm{MFDR}_{\theta^*}(\theta^*,t) + \upsilon(\varepsilon). \tag{58}$$

Furthermore, applying (58) for $t=T_{\theta^*}(\alpha+\upsilon(\varepsilon))$ and $t=T_{\theta^*}(\alpha-\upsilon(\varepsilon))$ yields

$$\alpha \le \mathrm{MFDR}_\theta(\theta,T_{\theta^*}(\alpha+\upsilon(\varepsilon))) \text{ and } \mathrm{MFDR}_\theta(\theta,T_{\theta^*}(\alpha-\upsilon(\varepsilon))) \le \alpha,$$

which gives the result by definition of the pseudo-inverse (15). □

**Lemma 4** (Concentration of the FDP process)**.** *There exists universal constants $c_1, c_1', c_2, c_2' > 0$ such that the following holds. Let Assumption 4 be true for some integer $M \geq 1$. Let $\theta^* = (\pi^*, w^*, \nu_0^*, \nu^*) \in \Theta$, $\pi^*_{min} = \min_q\{\pi^*_q\}$, $w^*_{min} = \min_{q,\ell}\{w^*_{q,\ell}\}$ and $w^*_{max} = \max_{q,\ell}\{w^*_{q,\ell}\}$. Then for all $x > 0$ with $x < (\pi^*_{min})^2 \wedge (1 - w^*_{max})$ and $y > 0$,*

$$\mathbb{P}_{\theta^*}\left( \sup_{\substack{\theta\in\Theta, t\in[0,1]\\ F_{\theta^*}(\theta,t)\geq y}} |FDP(\theta,t) - \mathrm{MFDR}_{\theta^*}(\theta,t)| > x \right) \\ \leq c_1 Q^2 \mathrm{e}^{-c_1' n y^2 x^2/Q^4} + c_2 M Q^2 \mathrm{e}^{-c_2' n^2 (\pi^*_{min})^2 (1-w^*_{max}) y^2 x^2/M^2}.$$

*Proof.* We have

$$\begin{aligned} &\left| \frac{\widehat{F}_0(\theta,t)}{\widehat{F}(\theta,t)} - \frac{F_{0,\theta^*}(\theta,t)}{F_{\theta^*}(\theta,t)} \right| \\ &\leq \left| \frac{\widehat{F}_0(\theta,t) - F_{0,\theta^*}(\theta,t)}{\widehat{F}(\theta,t)} \right| + \frac{F_{0,\theta^*}(\theta,t)}{\widehat{F}(\theta,t)F_{\theta^*}(\theta,t)} |\widehat{F}(\theta,t) - F_{\theta^*}(\theta,t)| \\ &\leq \left| \frac{\widehat{F}_0(\theta,t) - F_{0,\theta^*}(\theta,t)}{\widehat{F}(\theta,t)} \right| + \frac{1}{\widehat{F}(\theta,t)} |\widehat{F}(\theta,t) - F_{\theta^*}(\theta,t)| \\ &\leq \frac{2}{F_{\theta^*}(\theta,t)} \left( \left|\widehat{F}_0(\theta,t) - F_{0,\theta^*}(\theta,t)\right| + \left|\widehat{F}(\theta,t) - F_{\theta^*}(\theta,t)\right| \right) \\ &\leq \frac{2}{y} \left( \left|\widehat{F}_0(\theta,t) - F_{0,\theta^*}(\theta,t)\right| + \left|\widehat{F}(\theta,t) - F_{\theta^*}(\theta,t)\right| \right), \end{aligned}$$

which holds provided that $F_{\theta^*}(\theta,t) \geq y$ and by assuming $\widehat{F}(\theta,t) \geq F_{\theta^*}(\theta,t)/2$. Now consider the event

$$\left\{ \sup_{\substack{\theta\in\Theta, t\in[0,1]\\ F_{\theta^*}(\theta,t)\geq y}} |\widehat{F}_0(\theta,t) - F_{0,\theta^*}(\theta,t)| \leq xy/4 \right\} \\ \cap \left\{ \sup_{\substack{\theta\in\Theta, t\in[0,1]\\ F_{\theta^*}(\theta,t)\geq y}} |\widehat{F}(\theta,t) - F_{\theta^*}(\theta,t)| \leq xy/4 \right\}.$$

On this event, we indeed have $\widehat{F}(\theta,t) \geq F_{\theta^*}(\theta,t) - xy/4 \geq F_{\theta^*}(\theta,t) - y/4 \geq F_{\theta^*}(\theta,t)/2$ (because $x \leq 1$ and, again, $y \leq F_{\theta^*}(\theta,t)$). Hence, the display above entails

$$\sup_{\substack{\theta\in\Theta, t\in[0,1]\\ F_{\theta^*}(\theta,t)\geq y}} \left| \frac{\widehat{F}_0(\theta,t)}{\widehat{F}(\theta,t)} - \frac{F_{0,\theta^*}(\theta,t)}{F_{\theta^*}(\theta,t)} \right| \leq \frac{2}{y}(xy/4 + xy/4) = x.$$

Applying Lemma 5 with $x$ in place of $xy/4$ finishes the proof. ∎

### *C.4. Auxiliary results*

**Lemma 5.** *Consider the setting of Lemma 4 and let $m = n(n-1)/2$. Then for all $x > 0$ with $x < (\pi^*_{min})^2 \wedge (1 - w^*_{max})$,*

$$\mathbb{P}_{\theta^*}\left(\sup_{\theta\in\Theta, t\in[0,1]}\left|\widehat{F}_0(\theta,t) - F_{0,\theta^*}(\theta,t)\right| > x\right) \leq 2Q^2\mathrm{e}^{-2\lfloor n/2\rfloor x^2/(9Q^4)} + 3Q^2\mathrm{e}^{-m(\pi^*_{min})^2(1-w^*_{max})x^2/(72M^2)}.$$

*For all $x > 0$ with $x < (\pi^*_{min})^2 \wedge w^*_{min}$,*

$$\mathbb{P}_{\theta^*}\left(\sup_{\theta\in\Theta, t\in[0,1]}\left|\widehat{F}_1(\theta,t) - F_{1,\theta^*}(\theta,t)\right| > x\right) \leq 2Q^2\mathrm{e}^{-2\lfloor n/2\rfloor x^2/(9Q^4)} + 3Q^2\mathrm{e}^{-m(\pi^*_{min})^2 w^*_{min}x^2/(72M^2)}.$$

*Finally, for all $x > 0$ with $x < (\pi^*_{min})^2$,*

$$\mathbb{P}_{\theta^*}\left(\sup_{\theta\in\Theta, t\in[0,1]}\left|\widehat{F}(\theta,t) - F_{\theta^*}(\theta,t)\right| > x\right) \leq 2Q^2\mathrm{e}^{-2\lfloor n/2\rfloor x^2/(9Q^4)} + 2Q^2\mathrm{e}^{-m(\pi^*_{min})^2x^2/36}.$$

*Proof.* For short, we denote in this proof $\pi^*$, $w^*$, $\pi^*_{min}$, $w^*_{min}$, $w^*_{max}$ by $\pi$, $w$, $\pi_{min}$, $w_{min}$, $w_{max}$, respectively. Let us denote

$$m_{q,\ell}(Z) = \sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{Z_i = q, Z_j = \ell\}$$
$$m_{0,q,\ell}(A,Z) = \sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{Z_i = q, Z_j = \ell\}(1 - A_{i,j}),$$

for $q, \ell \in \{1,\dots,Q\}$. For all $\theta \in \Theta$ and $t \in [0,1]$, we have

$$\widehat{F}_0(\theta,t) = m^{-1}\sum_{(i,j)\in\mathcal{A}}(1 - A_{i,j})\mathbb{1}\{\ell_{i,j}(\theta) \leq t\} = \sum_{1\leq q,\ell\leq Q}\frac{m_{0,q,\ell}(A,Z)}{m}\widehat{F}_{0,q,\ell}(\theta,t),$$

where we let

$$\widehat{F}_{0,q,\ell}(\theta,t) = (m_{0,q,\ell}(A,Z))^{-1}\sum_{(i,j)\in\mathcal{A}}\mathbb{1}\{Z_i = q, Z_j = \ell\}(1 - A_{i,j})\mathbb{1}\{\boldsymbol{\ell}(X_{i,j}, q, \ell;\theta) \leq t\}.$$

Note that

$$\mathbb{E}_{\theta^*}(\widehat{F}_{0,q,\ell}(\theta,t) \mid A, Z) = \boldsymbol{L}_0(t,q,\ell;\theta^*,\theta),$$

as defined by (8). As a consequence, we have

$$|\widehat{F}_0(\theta,t) - F_{0,\theta^*}(\theta,t)|$$

$$
\begin{aligned}
&= \left| \sum_{1\le q,\ell\le Q} \frac{m_{0,q,\ell}(A,Z)}{m} \widehat{F}_{0,q,\ell}(\theta,t) - \sum_{1\le q,\ell\le Q} \pi_q\pi_\ell(1-w_{q,\ell})\boldsymbol{L}_0(t,q,\ell;\theta^*,\theta) \right| \\
&\le \sum_{1\le q,\ell\le Q} \left| \frac{m_{0,q,\ell}(A,Z)}{m} \widehat{F}_{0,q,\ell}(\theta,t) - \pi_q\pi_\ell(1-w_{q,\ell})\boldsymbol{L}_0(t,q,\ell;\theta^*,\theta) \right| \\
&\le \sum_{1\le q,\ell\le Q} \left| \frac{m_{0,q,\ell}(A,Z)}{m} - \pi_q\pi_\ell(1-w_{q,\ell}) \right| \\
&\quad + \sum_{1\le q,\ell\le Q} \pi_q\pi_\ell(1-w_{q,\ell}) \left| \widehat{F}_{0,q,\ell}(\theta,t) - \boldsymbol{L}_0(t,q,\ell;\theta^*,\theta) \right| \\
&\le \sum_{1\le q,\ell\le Q} \frac{m_{q,\ell}(Z)}{m} \left| \frac{m_{0,q,\ell}(A,Z)}{m_{q,\ell}(Z)} - (1-w_{q,\ell}) \right| \\
&\quad + \sum_{1\le q,\ell\le Q} (1-w_{q,\ell}) \left| \frac{m_{q,\ell}(Z)}{m} - \pi_q\pi_\ell \right| \\
&\quad + \sum_{1\le q,\ell\le Q} \pi_q\pi_\ell(1-w_{q,\ell}) \left| \widehat{F}_{0,q,\ell}(\theta,t) - \boldsymbol{L}_0(t,q,\ell;\theta^*,\theta) \right| .
\end{aligned}
$$

The latter is smaller than or equal to $x$ on the event

$$
\begin{aligned}
\Omega = \Bigg\{ &\forall q,\ell \in \{1,\dots,Q\} \ : \ \left| \frac{m_{0,q,\ell}(A,Z)}{m_{q,\ell}(Z)} - (1-w_{q,\ell}) \right| \le x/3, \\
&\left| \frac{m_{q,\ell}(Z)}{m} - \pi_q\pi_\ell \right| \le x/(3Q^2), \sup_{\theta\in\Theta,t\in[0,1]} \left| \widehat{F}_{0,q,\ell}(\theta,t) - \boldsymbol{L}_0(t,q,\ell;\theta^*,\theta) \right| \le x/3 \Bigg\} .
\end{aligned}
$$

Let us now provide an upper-bound for $\mathbb{P}_{\theta^*}(\Omega^c)$. We have

$$
\begin{aligned}
&\mathbb{P}_{\theta^*}(\Omega^c) \\
&\le \sum_{q,\ell=1}^{Q} \mathbb{P}_{\theta^*}\left( \left| \frac{m_{0,q,\ell}(A,Z)}{m_{q,\ell}(Z)} - (1-w_{q,\ell}) \right| > x/3, \left| \frac{m_{q,\ell}(Z)}{m} - \pi_q\pi_\ell \right| \le x/(3Q^2) \right) \\
&\quad + \sum_{q,\ell=1}^{Q} \mathbb{P}_{\theta^*}\left( \left| \frac{m_{q,\ell}(Z)}{m} - \pi_q\pi_\ell \right| > x/(3Q^2) \right) \\
&\quad + \sum_{q,\ell=1}^{Q} \mathbb{P}_{\theta^*}\Bigg( \sup_{\theta\in\Theta,t\in[0,1]} \left| \widehat{F}_{0,q,\ell}(\theta,t) - \boldsymbol{L}_0(t,q,\ell;\theta^*,\theta) \right| > x/3, \\
&\qquad\qquad \left| \frac{m_{0,q,\ell}(A,Z)}{m_{q,\ell}(Z)} - (1-w_{q,\ell}) \right| \le x/3, \left| \frac{m_{q,\ell}(Z)}{m} - \pi_q\pi_\ell \right| \le x/(3Q^2) \Bigg) . \\
&= (I) + (II) + (III).
\end{aligned}
$$

To bound (I), we note that, conditionally on $Z$, $m_{0,q,\ell}(A,Z)$ is the sum of $m_{q,\ell}(Z)$ i.i.d. $\mathcal{B}(1-w_{q,\ell})$, which gives by applying (i) of Lemma 7 ($p = 1-w_{q,\ell}$,

$n = m_{q,\ell}(Z)$), that

$$(I) \le \sum_{q,\ell=1}^{Q} \mathbb{E}_{\theta^*} \left( 2\mathrm{e}^{-2m_{q,\ell}(Z)(x/3)^2} \mathbb{1}\left\{ \left| \frac{m_{q,\ell}(Z)}{m} - \pi_q \pi_\ell \right| \le x/(3Q^2) \right\} \right)$$
$$\le 2Q^2 \mathrm{e}^{-m(2/9)\left(\pi_q\pi_\ell - x/(3Q^2)\right)_+ x^2}.$$

For bounding (II), we use readily (ii) of Lemma 7 to obtain

$$(II) \le 2Q^2 \mathrm{e}^{-2\lfloor n/2 \rfloor x^2/(9Q^4)}.$$

For bounding (III), note that

$$\widehat{F}_{0,q,\ell}(\theta, t)$$
$$= (m_{0,q,\ell}(A,Z))^{-1} \sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{Z_i = q, Z_j = \ell\}(1 - A_{i,j})\mathbb{1}\{\boldsymbol{\ell}(X_{i,j}, q, \ell; \theta) \le t\}$$
$$= (m_{0,q,\ell}(A,Z))^{-1} \sum_{(i,j)\in\mathcal{A}} \mathbb{1}\{Z_i = q, Z_j = \ell\}(1 - A_{i,j})\mathbb{1}\{X_{i,j} \in I(q,\ell,\theta,t)\},$$

for some $I(q,\ell,\theta,t) \in \mathcal{I}_M$, by using Assumption 4. Note that, conditionally on $A, Z$, the variables of $(X_{i,j}, (i,j) \in \mathcal{A}, A_{i,j} = 0, Z_j = q, Z_j = \ell)$ are i.i.d. Hence, we can apply Lemma 6 ($n = m_{0,q,\ell}(A,Z)$), to get for all $x > 0$,

$$(III) \le Q^2 \mathbb{E}_{\theta^*} \Big( 2\mathrm{e}^{-m_{0,q,\ell}(A,Z)x^2/(18M^2)}$$
$$\mathbb{1}\left\{ m_{0,q,\ell}(A,Z) \ge m(\pi_q\pi_\ell - x/(3Q^2))((1 - w_{q,\ell}) - x/3) \right\} \Big)$$
$$\le 2Q^2 \mathrm{e}^{-m(\pi_q\pi_\ell - x/(3Q^2))_+((1-w_{q,\ell}) - x/3)_+ x^2/(18M^2)}.$$

Finally, note that $\pi_q\pi_\ell - x/3 \ge \pi_{min}^2/2$ provided that $x \le 3\pi_{min}^2/2$ and $(1 - w_{q,\ell}) - x/3 \ge (1 - w_{max})/2$, provided that $x \le 3(1 - w_{max})/2$, so that $\mathbb{P}_{\theta^*}(\Omega^c)$ is smaller than

$$2Q^2 \mathrm{e}^{-m\pi_{min}^2 x^2/9} + 2Q^2 \mathrm{e}^{-2\lfloor n/2 \rfloor x^2/(9Q^4)} + 2Q^2 \mathrm{e}^{-m\pi_{min}^2(1-w_{max})x^2/(72M^2)}.$$

This concludes the proof of the first inequality. The second inequality is similar, by replacing $1 - w_{q,\ell}$ (resp. $1 - A_{i,j}$) by $w_{q,\ell}$ (resp. $A_{i,j}$). The third inequality is obtained similarly. □

**Lemma 6.** *Let $U_1, \dots, U_n$ be $n$ i.i.d. continuous real random variables and $M$ a positive integer. Then we have for all $x > 0$,*

$$\mathbb{P}\left( \sup_{I \in \mathcal{I}_M} \left| n^{-1} \sum_{i=1}^{n} \mathbb{1}\{U_i \in I\} - \mathbb{P}(U_1 \in I) \right| \ge x \right) \le 2\mathrm{e}^{-nx^2/(2M^2)},$$

*where $\mathcal{I}_M$ is defined by (38).*

*Proof.* Denote

$$\mathcal{C} = \{(a_k, b_k)_{1\le k\le M} \text{ such that } -\infty \le a_k \le b_k \le +\infty \text{ for } 1 \le k \le M, \\ \text{and } b_k \le a_{k+1} \text{ for } 1 \le k \le M-1\}.$$

First, note that, almost surely, $\forall i \in \{1, \dots, n\}$, $\forall t \in \mathbb{Q}$, $U_i \neq t$, so that we have almost surely,

$$\begin{aligned}\sup_{t\in\mathbb{R}} \left| n^{-1} \sum_{i=1}^n \mathbb{1}\{U_i < t\} - \mathbb{P}(U_1 < t) \right| &= \sup_{t\in\mathbb{Q}} \left| n^{-1} \sum_{i=1}^n \mathbb{1}\{U_i < t\} - \mathbb{P}(U_1 < t) \right| \\ &= \sup_{t\in\mathbb{Q}} \left| n^{-1} \sum_{i=1}^n \mathbb{1}\{U_i \le t\} - \mathbb{P}(U_1 \le t) \right| \\ &= \sup_{t\in\mathbb{R}} \left| n^{-1} \sum_{i=1}^n \mathbb{1}\{U_i \le t\} - \mathbb{P}(U_1 \le t) \right|.\end{aligned}$$

Hence, almost surely,

$$\begin{aligned}&\sup_{I\in\mathcal{I}_M} \left| n^{-1} \sum_{i=1}^n \mathbb{1}\{U_i \in I\} - \mathbb{P}(U_1 \in I) \right| \\ &\le 2 \sup_{(a_k,b_k)_k\in\mathcal{C}} \sum_{k=1}^M \left| n^{-1} \sum_{i=1}^n \mathbb{1}\{U_i \le a_k\} - \mathbb{P}(U_1 \le a_k) \right| \\ &\le 2M \sup_{t\in\mathbb{R}} \left| n^{-1} \sum_{i=1}^n \mathbb{1}\{U_i \le t\} - \mathbb{P}(U_1 \le t) \right|,\end{aligned}$$

because $\mathbb{1}\{a_k < U_i < b_k\} = \mathbb{1}\{U_i < b_k\} - \mathbb{1}\{U_i \le a_k\}$. The proof is finished by using DKW inequality with Massart's constant, see [31]. □

**Lemma 7.** *Let $n \ge 2$ be an integer. Then*

*(i) For $Y \sim \mathcal{B}(n, p)$, $p \in (0, 1)$, we have for all $x > 0$,*

$$\mathbb{P}\left(|Y/n - p| \ge x\right) \le 2\mathrm{e}^{-2nx^2}.$$

*(ii) For $Z_i$ for $1 \le i \le n$ i.i.d. where $\pi_q = \mathbb{P}(Z_1 = q) \in (0, 1)$, $q = 1, \dots, Q$, $\sum_{q=1}^Q \pi_q = 1$, we have for all $x > 0$,*

$$\mathbb{P}\left( \left| m^{-1} \sum_{1\le i<j\le n} \mathbb{1}\{Z_i = q, Z_j = \ell\} - \pi_q\pi_\ell \right| \ge x \right) \le 2\mathrm{e}^{-2\lfloor n/2\rfloor x^2}.$$

*Proof.* Both inequalities are applications of versions of Hoeffding inequalities: (i) is the classical version, while (ii) is the one devoted to $U$-statistics, see e.g. [41]. □

### C.5. Proof of Theorem 6

We follow an argument inspired from the proof of Theorem 1 in [14]. Denote $\ell_{i,j}(\theta^*)$ by $\ell_{i,j}$ for short. First, note that by Lemma 2, $\mathrm{MFDR}_{\theta^*}(\varphi^*) = \mathrm{MFDR}_{\theta^*}(\theta^*, T_{\theta^*}(\alpha)) = \alpha$ by definition of $T_{\theta^*}(\alpha) \in (t_1(\theta^*), t_2(\theta^*))$, see (15). This gives

$$\mathbb{E}_{\theta^*}\left[\sum_{(i,j)\in\mathcal{A}} (\ell_{i,j}-\alpha)\varphi^*_{i,j}\right] = \mathbb{E}_{\theta^*}\left[\sum_{(i,j)\in\mathcal{A}} \ell_{i,j}\varphi^*_{i,j}\right] - \alpha\mathbb{E}_{\theta^*}\left[\sum_{(i,j)\in\mathcal{A}} \varphi^*_{i,j}\right] = 0,$$

by using the MFDR expression given by (53). Also, since $\Psi : x \in [0,1) \mapsto (x-\alpha)/(1-x)$ is continuous increasing and defines a one to one map from $[0,1)$ to $[-\alpha,+\infty)$, we have, almost surely,

$$\varphi^*_{i,j} = \mathbb{1}\{\Psi(\ell_{i,j}) \le \Psi(T_{\theta^*}(\alpha))\} = \mathbb{1}\left\{\ell_{i,j}-\alpha \le \frac{T_{\theta^*}(\alpha)-\alpha}{1-T_{\theta^*}(\alpha)}(1-\ell_{i,j})\right\}.$$

As a result, it can then be checked that for any procedure $\varphi$,

$$(\ell_{i,j}-\alpha)(\varphi^*_{i,j}-\varphi_{i,j}) \le \frac{T_{\theta^*}(\alpha)-\alpha}{1-T_{\theta^*}(\alpha)}(1-\ell_{i,j})(\varphi^*_{i,j}-\varphi_{i,j}). \tag{59}$$

Indeed, this is true if $\varphi^*_{i,j}=1$ and $\varphi_{i,j}=0$. If $\varphi^*_{i,j}=\varphi_{i,j}$ this obviously holds. If $\varphi^*_{i,j}=0$ and $\varphi_{i,j}=1$, then $\ell_{i,j}-\alpha \ge \frac{T_{\theta^*}(\alpha)-\alpha}{1-T_{\theta^*}(\alpha)}(1-\ell_{i,j})$ and the relation is also true. Hence, provided that $\mathrm{MFDR}_{\theta^*}(\varphi) \le \alpha$, we have

$$\mathbb{E}_{\theta^*}\left[\sum_{(i,j)\in\mathcal{A}} (\ell_{i,j}-\alpha)\varphi_{i,j}\right] \le 0 = \mathbb{E}_{\theta^*}\left[\sum_{(i,j)\in\mathcal{A}} (\ell_{i,j}-\alpha)\varphi^*_{i,j}\right].$$

This implies

$$0 \le \mathbb{E}_{\theta^*}\left[\sum_{(i,j)\in\mathcal{A}} (\ell_{i,j}-\alpha)(\varphi^*_{i,j}-\varphi_{i,j})\right].$$

Hence, by (59), we get

$$0 \le \frac{T_{\theta^*}(\alpha)-\alpha}{1-T_{\theta^*}(\alpha)}\,\mathbb{E}_{\theta^*}\left[\sum_{(i,j)\in\mathcal{A}} (1-\ell_{i,j})(\varphi^*_{i,j}-\varphi_{i,j})\right],$$

which in turn gives $\mathrm{TDR}_{\theta^*}(\varphi^*) \ge \mathrm{TDR}_{\theta^*}(\varphi)$, because $T_{\theta^*}(\alpha) \in (\alpha, 1)$.

Finally, let us prove $\limsup_n \{\mathrm{FDR}_{\theta^*}(\varphi^*)\} \le \alpha$. First observe that by using (47) and (52), we have for $x>0$ with $x < (\pi^*_{min})^2 \wedge (1-w^*_{max})$ and $y = F_{\theta^*}(\theta^*, T_{\theta^*}(\alpha))$,

$$\mathrm{FDR}_{\theta^*}(\varphi^*) \le \mathbb{E}_{\theta^*}\left[\mathrm{FDP}(\theta^*, T_{\theta^*}(\alpha))\mathbb{1}_\Omega\right] + \mathbb{P}_{\theta^*}(\Omega^c)$$

Hence, noting that $\mathrm{MFDR}_{\theta^*}(\theta^*, T_{\theta^*}(\alpha)) = \alpha$ and upper bounding $\mathbb{P}_{\theta^*}(\Omega^c)$ according to Lemma 4 (that uses Assumption 4), we get

$$\mathrm{FDR}_{\theta^*}(\varphi^*) \leq x + \alpha + c_1 Q^2 \mathrm{e}^{-c_1' \lfloor n/2 \rfloor y^2 x^2 / Q^4} + c_2 M Q^2 \mathrm{e}^{-c_2' m (\pi^*_{min})^2 (1-w^*_{max}) y^2 x^2 / M^2}.$$

Making $n$ goes to infinity and $x$ goes to zero in the last display gives the result.

## Appendix D: Computations in the Gaussian model

### D.1. Checking assumptions in a Gaussian NSBM

In this section, we consider the Gaussian NSBM (1) with parameter set such that

$$\text{for all } q, \ell \in \{1, \dots, Q\},\ (\mu_{q,\ell}, \sigma_{q,\ell}) \neq (0, \sigma_0). \tag{60}$$

The next sections present explicit calculations showing parameter configurations $\theta = (\pi, w, \sigma_0, \mu, \sigma)$ for which Assumptions 3, 4 and 5 hold true. The following propositions gather the obtained results.

**Proposition 3.** *In the Gaussian NSBM satisfying* (60)*, Assumption 3 hold, with, for all $q, \ell$, $t_{1,q,\ell}(\theta) = 0$ if and only if $\sigma_{q,\ell} \geq \sigma_0$, and $t_{2,q,\ell}(\theta) = 1$ if and only if $\sigma_{q,\ell} \leq \sigma_0$. In addition, Assumption 4 hold for $M = 2$.*

Proposition 3 is proved in Sections D.2 and D.3. Hence, the function $t \mapsto \mathrm{MFDR}_\theta(\theta, t)$ enjoys the properties given in Lemma 2. Nevertheless, it might jump in $t_1(\theta)$ (see case 4), so is not necessarily continuous on $[0, 1]$. Also, this function might have infinite derivative at the boundary points $\{t_{1,q,\ell}(\theta)\}_{q,\ell}$, $\{t_{2,q,\ell}(\theta)\}_{q,\ell}$ (see cases 2-3-4).

**Proposition 4.** *In the Gaussian NSBM satisfying* (60)*, for any parameter $\theta = (\pi, w, \sigma_0, \mu, \sigma) \in \Theta$, we have $\alpha_c(\theta) = 0$ if and only if there exists $q, \ell \in \{1, \dots, Q\}$, such that $\sigma_{q,\ell} \geq \sigma_0$.*

Proposition 4 is proved in Section D.4. For instance, in the context of Figure 9, $\alpha_c = 0$ in cases 1-2-3 while $\alpha_c > 0$ in case 4.

Avoiding the non regular behavior of our functionals at the boundary points $\{t_{1,q,\ell}(\theta)\}_{q,\ell} \cup \{t_{2,q,\ell}(\theta)\}_{q,\ell}$, we can provide Assumption 5 (i)-(ii)-(iii).

Let us consider $\theta^*$ the true value of the parameter and let

$$A(\theta^*) = \mathrm{MFDR}_{\theta^*}(\theta^*, t_2'(\theta^*)) \in (\alpha_c, e_0] \tag{61}$$

$$t_2'(\theta^*) = \min\{t_{1,q,\ell}(\theta^*), t_{2,q,\ell}(\theta^*),\ \text{for } q, \ell \text{ s.t. } t_{1,q,\ell}(\theta^*) \neq t_1(\theta^*)\} \in (t_1(\theta^*), 1]. \tag{62}$$

The following proposition is proved in Section D.5.

**Proposition 5.** *In the Gaussian NSBM satisfying* (60)*, assume $\alpha \in (\alpha_c, A(\theta^*))$. Then Assumption 5 (ii) holds with any compact interval $\mathcal{K} \subset [(\alpha_c + \alpha)/2, (\alpha + A(\theta^*))/2]$.*

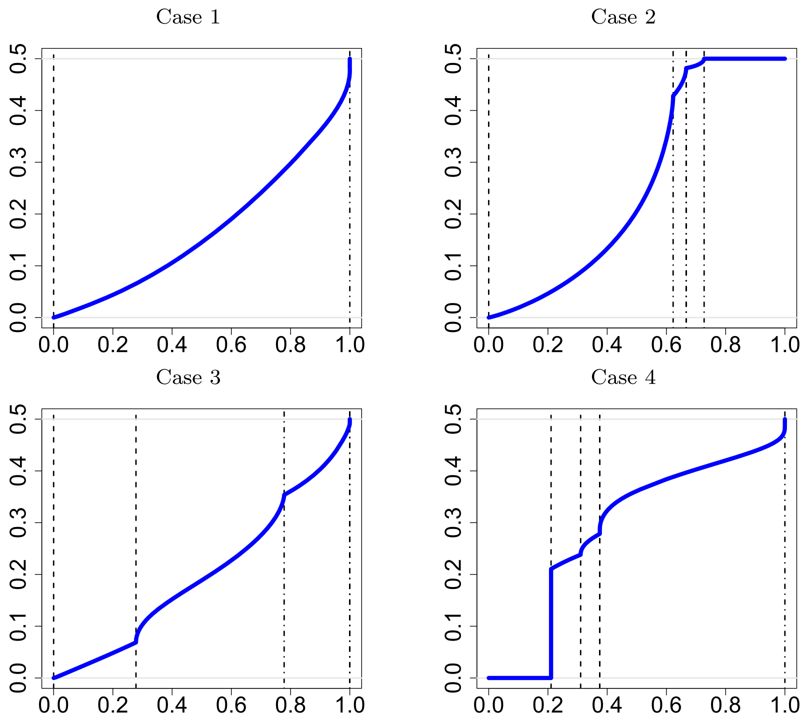



FIG 9. *Plot of* $t \mapsto \mathrm{MFDR}_\theta(\theta, t)$ *defined by* (7) *in the Gaussian case, for* 2 *different values of the parameter* $\theta$. *In each case, the vertical dashed (resp. dashed-dotted) lines correspond to* $\{t_{1,q,\ell}\}_{q,\ell}$ *(resp.* $\{t_{2,q,\ell}\}_{q,\ell}$*). In all cases, we have* $\pi_{q,\ell} = 0.5$ *for all* $q, \ell$*;* $Q = 2$*;* $w_{1,1} = 0.4$*,* $w_{1,2} = 0.5$*,* $w_{2,2} = 0.6$*;* $\sigma_0 = 1$*. For case* 1*:* $\mu_{1,1} = 1$*,* $\mu_{1,2} = -2$*,* $\mu_{2,2} = 4$*,* $\sigma_{q,\ell} = \sigma_0$ *for all* $q, \ell$*. For case* 2*:* $\mu_{q,\ell} = 0$ *for all* $q, \ell$*,* $\sigma_{1,1} = 1.1$*,* $\sigma_{1,2} = 2$*,* $\sigma_{2,2} = 4$*. For case* 3*:* $\mu_{1,1} = 1$*,* $\mu_{1,2} = -2$*,* $\mu_{2,2} = 4$*,* $\sigma_{1,1} = 0.5$*,* $\sigma_{1,2} = 1.1$*,* $\sigma_{2,2} = 3$*. For case* 4*:* $\mu_{q,\ell} = 0$ *for all* $q, \ell$*,* $\sigma_{1,1} = 0.3$*,* $\sigma_{1,2} = 0.9$*,* $\sigma_{2,2} = 0.4$*.*

For instance, in Figure 9, $t'_2(\theta^*)$ is equal to 1, $\approx 0.62$, $\approx 0.27$ and $\approx 0.31$ in case 1-2-3-4, respectively. As a matter of fact, $t'_2(\theta^*)$ is often fairly away from zero. For instance, we establish in Section D.6, that $t'_2(\theta^*) \geq 1/2$ when $\theta^* = (\pi^*, w^*, \sigma_0^*, \mu^*, \sigma^*)$ is such that $w^*_{q,\ell} \leq 1/2$ and $\sigma^*_{q,\ell} \geq \sigma^*_0$ for all $q, \ell$.

**Proposition 6.** *In the Gaussian NSBM satisfying* (60)*, assume* $\alpha \in (\alpha_c, A(\theta^*))$*. Then Assumption 5 (i) holds with any compact interval* $\mathcal{K} \subset [(\alpha_c + \alpha)/2, (\alpha + A(\theta^*))/2]$ *provided that the parameter set* $\Theta$ *satisfies*

$$\mathcal{C}(\theta) = \{(q, \ell) \in \{1, \dots, Q\}^2 \,:\, \sigma_{q,\ell} = \sigma_0\} \tag{63}$$

*does not depend on* $\theta = (\pi, w, \sigma_0, \mu, \sigma) \in \Theta$.

Proposition 6 is proved in Section D.7. The main idea is that (63) ensures that, in a neighborhood of $\theta^*$, the formulas given in (i)-(ii)-(iii) of Section D.2 are active, without indicator, which means that the involved functionals are regular enough.

Equivalently, (63) means that there exists some $\mathcal{C} \subset \{1, \ldots, Q\}^2$ such that

$$\Theta \subset \{\theta = (\pi, w, \sigma_0, \mu, \sigma) \in \Theta \,:\, \forall (q, \ell) \in \mathcal{C}, \sigma_{q,\ell} = \sigma_0,\ \forall (q, \ell) \notin \mathcal{C}, \sigma_{q,\ell} \neq \sigma_0\}.$$

For instance, it is satisfied in one of the two following parameter sets (mentioned in the main manuscript):

- Equal variances: $\sigma_{q,\ell} = \sigma_0$ (and thus also $\mu_{q,\ell} \neq 0$), for all $1 \leq q, \ell \leq Q$. This corresponds to $\mathcal{C}(\theta) = \{1, \ldots, Q\}^2$;
- Larger variances under alternatives: $\sigma_{q,\ell} > \sigma_0$ for all $1 \leq q, \ell \leq Q$. This corresponds to $\mathcal{C}(\theta) = \emptyset$.

**Proposition 7.** *In the Gaussian NSBM satisfying (60), there exists some measurable set $\Lambda \subset [0, 1]$ of Lebesgue measure 0 such that if $\alpha \in (\alpha_c, A(\theta^*))\backslash\Lambda$, Assumption 5 (iii) holds with any compact interval $\mathcal{K} \subset [(\alpha_c + \alpha)/2, (\alpha + A(\theta^*))/2]$.*

The proof of Proposition 7 is straightforward and does not rely on the Gaussian distribution; since $T_{\theta^*}$ is continuous increasing on $(\alpha_c, e_0)$, it is almost everywhere differentiable on $(\alpha_c, A(\theta^*))$. It is thus differentiable in $\alpha$, up to remove a subset of Lebesgue measure equal to zero.

### D.2. Computing $\boldsymbol{\ell}(\cdot)$ and $\boldsymbol{L}_\delta(\cdot)$

The functional $\boldsymbol{\ell}$ (6) is clearly given by

$$\boldsymbol{\ell}(x, q, \ell; \theta) = \frac{(1 - w_{q,\ell})\phi(x/\sigma_0)/\sigma_0}{(1 - w_{q,\ell})\phi(x/\sigma_0)/\sigma_0 + w_{q,\ell}\phi((x - \mu_{q,\ell})/\sigma_{q,\ell})/\sigma_{q,\ell}}. \tag{64}$$

Now, let us fix $q, \ell \in \{1, \ldots, Q\}$, $\theta', \theta \in \Theta$, $\delta \in \{0, 1\}$ and let us compute the functional $\boldsymbol{L}_\delta$. We have

$$\boldsymbol{L}_\delta(t, q, \ell; \theta', \theta) = \mathbb{P}_{\theta'}(\boldsymbol{\ell}(X_{i,j}, q, \ell; \theta) \leq t \mid Z_i = q, Z_j = \ell, A_{i,j} = \delta).$$

First, for $t = 0$, $\boldsymbol{L}_\delta(t, q, \ell; \theta', \theta) = 0$. Second, for all $t \in (0, 1]$, if $\theta = (\pi, w, \sigma_0, \mu, \sigma)$, observe that

$$\begin{aligned}
&\{x \in \mathbb{R} \,:\, \boldsymbol{\ell}(x, q, \ell; \theta) \leq t\} \\
&= \left\{x \in \mathbb{R} \,:\, \frac{(1 - w_{q,\ell})\phi(x/\sigma_0)/\sigma_0}{(1 - w_{q,\ell})\phi(x/\sigma_0)/\sigma_0 + w_{q,\ell}\phi((x - \mu_{q,\ell})/\sigma_{q,\ell})/\sigma_{q,\ell}} \leq t\right\} \\
&= \left\{x \in \mathbb{R} \,:\, \frac{\phi((x - \mu_{q,\ell})/\sigma_{q,\ell})}{\phi(x/\sigma_0)} \geq (\sigma_{q,\ell}/\sigma_0)(1/w_{q,\ell} - 1)(1/t - 1)\right\}.
\end{aligned}$$

Now, since $-2\log\left(\frac{\phi((x-\mu_{q,\ell})/\sigma_{q,\ell})}{\phi(x/\sigma_0)}\right) = [\sigma_{q,\ell}^{-2} - \sigma_0^{-2}]x^2 - 2\mu_{q,\ell}\sigma_{q,\ell}^{-2}x + \mu_{q,\ell}^2\sigma_{q,\ell}^{-2}$, we have

$$\{\boldsymbol{\ell}(X_{i,j}, q, \ell; \theta) \leq t\} = \left\{aX_{i,j}^2 + bX_{i,j} + c \leq 0\right\}, \tag{65}$$

for the values $a = a(q,\ell,\theta)$, $b = b(q,\ell,\theta)$, $c = c(q,\ell,\theta)$ given by

$$\begin{cases} a &= \sigma_{q,\ell}^{-2} - \sigma_0^{-2}; \\ b &= -2\mu_{q,\ell}\sigma_{q,\ell}^{-2}; \\ c &= \mu_{q,\ell}^2\sigma_{q,\ell}^{-2} + 2\log\left((\sigma_{q,\ell}/\sigma_0)(1/w_{q,\ell}-1)(1/t-1)\right). \end{cases} \tag{66}$$

As a result, we have

$$\boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta) = \mathbb{P}_{U\sim\mathcal{N}(\mu(\delta),\sigma(\delta)^2)}(aU^2+bU+c\le 0), \tag{67}$$

for $\mu(\delta) = \mu(\delta,q,\ell,\theta^*)$, $\sigma(\delta) = \sigma(\delta,q,\ell,\theta^*)$ given by

$$\mu(0) = 0, \sigma(0) = \sigma_0^*, \text{ and } \mu(1) = \mu_{q,\ell}^*, \sigma(1) = \sigma_{q,\ell}^*. \tag{68}$$

In addition, expression (67) can be made explicit by an elementary inversion of $aU^2+bU+c<0$ in $U$. More precisely, denoting $\Phi_\delta$ the cumulative distribution function of the distribution $\mathcal{N}(\mu(\delta),\sigma(\delta)^2)$, we obtain

(i) if $a<0$ (that is, $\sigma_{q,\ell} > \sigma_0$),

$$\boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta) = \mathbb{1}\{b^2<4ac\} + \left[\Phi_\delta\left(\frac{b-(b^2-4ac)_+^{1/2}}{2|a|}\right) + 1 - \Phi_\delta\left(\frac{b+(b^2-4ac)_+^{1/2}}{2|a|}\right)\right]\mathbb{1}\{b^2\ge 4ac\};$$

(ii) if $a>0$ (that is, $\sigma_{q,\ell} < \sigma_0$),

$$\boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta) = \left[\Phi_\delta\left(\frac{-b+(b^2-4ac)_+^{1/2}}{2a}\right) - \Phi_\delta\left(\frac{-b-(b^2-4ac)_+^{1/2}}{2a}\right)\right]\mathbb{1}\{b^2>4ac\};$$

(iii) if $a=0$ and $b\neq 0$ (that is, $\sigma_{q,\ell}=\sigma_0$ and $\mu_{q,\ell}\neq 0$),

$$\boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta) = (1-\Phi_\delta(-c/b))\mathbb{1}\{b<0\} + \Phi_\delta(-c/b)\mathbb{1}\{b>0\}.$$

Also, both equations in (i) and (ii) can be merged as follows: if $a\neq 0$,

$$\boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta) = \mathbb{1}\{a<0\} + \left[\Phi_\delta\left(\frac{-b+(b^2-4ac)_+^{1/2}}{2a}\right) - \Phi_\delta\left(\frac{-b-(b^2-4ac)_+^{1/2}}{2a}\right)\right]\mathbb{1}\{b^2>4ac\}.$$

### *D.3. Proof of Proposition 3*

By Lemma 9, and expression (67), we get that, for all $q,\ell\in\{1,\dots,Q\}$, the function

$$(t,\theta,\theta^*)\in[0,1]\times\Theta^2 \mapsto \boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta)$$

is continuous on $[0,1]\times\Theta^2$, which proves part (i) of Assumption 3.

Now, let us define $t_{1,q,\ell}(\theta)$ and $t_{2,q,\ell}(\theta)$ as follows:

$$(t_{1,q,\ell}(\theta), t_{2,q,\ell}(\theta)) = \begin{cases} (0,t_0) & \text{if } \sigma_{q,\ell} > \sigma_0; \\ (t_0,1) & \text{if } \sigma_{q,\ell} < \sigma_0; \\ (0,1) & \text{if } \sigma_{q,\ell} = \sigma_0; \end{cases} \tag{69}$$

$$t_0 = \left(1 + \frac{w_{q,\ell}}{1-w_{q,\ell}}\frac{\sigma_0}{\sigma_{q,\ell}} \exp\left(\frac{\mu_{q,\ell}^2}{2(\sigma_0^2-\sigma_{q,\ell}^2)}\right)\right)^{-1}.$$

Then, we easily show that the function $t\in[0,1]\mapsto \boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta)$ is constant equal to 0 on $[0,t_{1,q,\ell}(\theta)]$, continuous increasing on $[t_{1,q,\ell}(\theta), t_{2,q,\ell}(\theta)]$ from $t = t_{1,q,\ell}(\theta)$ (value 0) to $t = t_{2,q,\ell}(\theta)$ (value 1) and then is constant equal to 1 on $[t_{2,q,\ell}(\theta),1]$.

Indeed, if $a = 0$ the result is obvious. If $a\neq 0$, $t_0$ is the only value of $t$ such that $b^2 = 4ac$; if $a<0$, the quantity $ac$, as a function of $t$, is continuous increasing with limits $-\infty$ and $+\infty$ in $t = 0^+$ and $t = 1^-$. If $a>0$, the quantity $ac$, as a function of $t$, is decreasing with limits $+\infty$ and $-\infty$ in $t = 0^+$ and $t = 1^-$, respectively. This proves part (ii) of Assumption 3.

Finally, note that (65) shows that Assumption 4 holds with $M = 2$.

### *D.4. Proof of Proposition 4*

Let $\theta = (\pi, w, \sigma_0, \mu, \sigma)\in\Theta$. Recall

$$\alpha_c(\theta) = \lim_{t\to t_1(\theta)^+}\{\mathrm{MFDR}_\theta(\theta,t)\} = \lim_{t\to t_1(\theta)^+}\left\{\frac{1}{1+\frac{\sum_{q,\ell}\pi_q\pi_\ell w_{q,\ell}\boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\sum_{q,\ell}\pi_q\pi_\ell(1-w_{q,\ell})\boldsymbol{L}_0(t,q,\ell;\theta,\theta)}}\right\}.$$

In this section, we compute $\alpha_c$ in different parameter configurations, which will prove Proposition 4. First note that for all $t\in[0,1]$, we have

$$\begin{aligned}\min_{q,\ell}\left\{\frac{w_{q,\ell}}{1-w_{q,\ell}}\;\frac{\boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\boldsymbol{L}_0(t,q,\ell;\theta,\theta)}\right\} &\le \frac{\sum_{q,\ell}\pi_q\pi_\ell w_{q,\ell}\boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\sum_{q,\ell}\pi_q\pi_\ell(1-w_{q,\ell})\boldsymbol{L}_0(t,q,\ell;\theta,\theta)} \\ &\le \max_{q,\ell}\left\{\frac{w_{q,\ell}}{1-w_{q,\ell}}\;\frac{\boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\boldsymbol{L}_0(t,q,\ell;\theta,\theta)}\right\}.\end{aligned}$$

We distinguish among the three following cases:

- if $\theta$ is such that for all $q,\ell$, we have $\sigma_{q,\ell}<\sigma_0$. Then $t_1(\theta) = \min_{q,\ell} t_{1,q,\ell}(\theta) > 0$. In that case,

$$\alpha_c(\theta) = \frac{1}{1+\frac{\sum_{q,\ell}\pi_q\pi_\ell w_{q,\ell}\boldsymbol{L}_1(t_1(\theta)^+,q,\ell;\theta,\theta)}{\sum_{q,\ell}\pi_q\pi_\ell(1-w_{q,\ell})\boldsymbol{L}_0(t_1(\theta)^+,q,\ell;\theta,\theta)}} > 0,$$

  because in the sums, $\boldsymbol{L}_0(t_1(\theta)^+,q,\ell;\theta,\theta)$ and $\boldsymbol{L}_1(t_1(\theta)^+,q,\ell;\theta,\theta)$ are non-zero for $q,\ell$ such that $t_{1,q,\ell}(\theta) = t_1(\theta)$ (and are zero otherwise).

- if $\theta$ is such that for all $q, \ell$, we have $\sigma_{q,\ell} \leq \sigma_0$ and there exists $q, \ell$ such that $\sigma_{q,\ell} = \sigma_0$. Then $t_1(\theta) = \min_{q,\ell} t_{1,q,\ell}(\theta) = 0$. Also, for $t$ close enough to 0, we have

$$\frac{\sum_{q,\ell} \pi_q \pi_\ell w_{q,\ell} \boldsymbol{L}_1(t, q, \ell; \theta, \theta)}{\sum_{q,\ell} \pi_q \pi_\ell (1 - w_{q,\ell}) \boldsymbol{L}_0(t, q, \ell; \theta, \theta)} = \frac{\sum_{q,\ell:\sigma_{q,\ell}=\sigma_0} \pi_q \pi_\ell w_{q,\ell} \boldsymbol{L}_1(t, q, \ell; \theta, \theta)}{\sum_{q,\ell:\sigma_{q,\ell}=\sigma_0} \pi_q \pi_\ell (1 - w_{q,\ell}) \boldsymbol{L}_0(t, q, \ell; \theta, \theta)},$$

because by Section D.3, $\boldsymbol{L}_\delta(t, q, \ell; \theta, \theta) = 0$ for $\sigma_{q,\ell} < \sigma_0$ when $t$ is close enough to 0. Next, by the computations of Section D.2, for all $\delta \in \{0, 1\}$ (see case $a = 0$ therein), for all $q, \ell$, with $\sigma_{q,\ell} = \sigma_0$, we have

$$\begin{aligned}
&\boldsymbol{L}_\delta(t, q, \ell; \theta, \theta) \\
&= \left(1 - \Phi_\delta\left(\frac{\mu_{q,\ell}^2 \sigma_{q,\ell}^{-2} + 2\log\left((\sigma_{q,\ell}/\sigma_0)(1/w_{q,\ell} - 1)(1/t - 1)\right)}{2\mu_{q,\ell}\sigma_{q,\ell}^{-2}}\right)\right) \\
&\times \mathbb{1}\{\mu_{q,\ell} > 0\} \\
&+ \Phi_\delta\left(\frac{\mu_{q,\ell}^2 \sigma_{q,\ell}^{-2} + 2\log\left((\sigma_{q,\ell}/\sigma_0)(1/w_{q,\ell} - 1)(1/t - 1)\right)}{2\mu_{q,\ell}\sigma_{q,\ell}^{-2}}\right) \mathbb{1}\{\mu_{q,\ell} < 0\} \\
&= \left(1 - \Phi_\delta\left(\frac{\mu_{q,\ell}}{2} + \sigma_0^2 \frac{\log\left((1/w_{q,\ell} - 1)(1/t - 1)\right)}{\mu_{q,\ell}}\right)\right) \mathbb{1}\{\mu_{q,\ell} > 0\} \\
&+ \Phi_\delta\left(\frac{\mu_{q,\ell}}{2} + \sigma_0^2 \frac{\log\left((1/w_{q,\ell} - 1)(1/t - 1)\right)}{\mu_{q,\ell}}\right) \mathbb{1}\{\mu_{q,\ell} < 0\}.
\end{aligned}$$

Recall that $1 - \Phi(x) \sim \phi(x)/x$ when $x \to \infty$. Hence, for all $q, \ell$, when $t \to 0^+$,

$$\begin{aligned}
&\frac{\boldsymbol{L}_1(t, q, \ell; \theta, \theta)}{\boldsymbol{L}_0(t, q, \ell; \theta, \theta)} \\
&\sim \frac{\frac{|\mu_{q,\ell}|}{2\sigma_0} + \sigma_0 \frac{\log((1/w_{q,\ell}-1)(1/t-1))}{|\mu_{q,\ell}|}}{\frac{-|\mu_{q,\ell}|}{2\sigma_0} + \sigma_0 \frac{\log((1/w_{q,\ell}-1)(1/t-1))}{|\mu_{q,\ell}|}} \frac{\phi\left(\frac{-|\mu_{q,\ell}|}{2\sigma_0} + \sigma_0 \frac{\log((1/w_{q,\ell}-1)(1/t-1))}{|\mu_{q,\ell}|}\right)}{\phi\left(\frac{|\mu_{q,\ell}|}{2\sigma_0} + \sigma_0 \frac{\log((1/w_{q,\ell}-1)(1/t-1))}{|\mu_{q,\ell}|}\right)} \\
&\sim \exp\left\{\log\left((1/w_{q,\ell} - 1)(1/t - 1)\right)\right\} = \frac{1 - w_{q,\ell}}{w_{q,\ell}}(1/t - 1).
\end{aligned}$$

because $\phi(x - y)/\phi(x + y) = e^{2xy}$ for all $x, y \in \mathbb{R}$. Hence, in that case, when $t \to 0^+$,

$$\frac{\sum_{q,\ell} \pi_q \pi_\ell w_{q,\ell} \boldsymbol{L}_1(t, q, \ell; \theta, \theta)}{\sum_{q,\ell} \pi_q \pi_\ell (1 - w_{q,\ell}) \boldsymbol{L}_0(t, q, \ell; \theta, \theta)} \sim 1/t.$$

and $\alpha_c(\theta) = 0$.

- if $\theta$ is such that there exist $q,\ell$ with $\sigma_{q,\ell} > \sigma_0$, then $t_1(\theta) = \min_{q,\ell} t_{1,q,\ell}(\theta) = 0$. Also, for $t$ close enough to 0, we have

$$\begin{aligned}&\frac{\sum_{q,\ell} \pi_q \pi_\ell w_{q,\ell} \boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\sum_{q,\ell} \pi_q \pi_\ell (1-w_{q,\ell}) \boldsymbol{L}_0(t,q,\ell;\theta,\theta)}\\ &= \frac{\sum_{q,\ell:\sigma_{q,\ell}\geq\sigma_0} \pi_q \pi_\ell w_{q,\ell} \boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\sum_{q,\ell:\sigma_{q,\ell}\geq\sigma_0} \pi_q \pi_\ell (1-w_{q,\ell}) \boldsymbol{L}_0(t,q,\ell;\theta,\theta)},\end{aligned}$$

because by Section D.3, $\boldsymbol{L}_\delta(t,q,\ell;\theta,\theta) = 0$ for $\sigma_{q,\ell} < \sigma_0$ when $t$ is close enough to 0. Hence, we have for $t$ close enough to 0,

$$\begin{aligned}&\min_{q,\ell:\sigma_{q,\ell}\geq\sigma_0} \left\{ \frac{w_{q,\ell}}{1-w_{q,\ell}} \, \frac{\boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\boldsymbol{L}_0(t,q,\ell;\theta,\theta)} \right\}\\ &\leq \frac{\sum_{q,\ell} \pi_q \pi_\ell w_{q,\ell} \boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\sum_{q,\ell} \pi_q \pi_\ell (1-w_{q,\ell}) \boldsymbol{L}_0(t,q,\ell;\theta,\theta)}\\ &\leq \max_{q,\ell:\sigma_{q,\ell}\geq\sigma_0} \left\{ \frac{w_{q,\ell}}{1-w_{q,\ell}} \, \frac{\boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\boldsymbol{L}_0(t,q,\ell;\theta,\theta)} \right\}.\end{aligned}$$

We know by above that when $t \to 0^+$,

$$\begin{aligned}&\min_{q,\ell:\sigma_{q,\ell}=\sigma_0} \left\{ \frac{w_{q,\ell}}{1-w_{q,\ell}} \, \frac{\boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\boldsymbol{L}_0(t,q,\ell;\theta,\theta)} \right\}\\ &\sim \max_{q,\ell:\sigma_{q,\ell}=\sigma_0} \left\{ \frac{w_{q,\ell}}{1-w_{q,\ell}} \frac{\boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\boldsymbol{L}_0(t,q,\ell;\theta,\theta)} \right\} \sim 1/t.\end{aligned}$$

Now, for $q,\ell$ such that $\sigma_{q,\ell} > \sigma_0$, we have for $t$ small enough ($a < 0$, $c > 0$)

$$\begin{aligned}&\boldsymbol{L}_\delta(t,q,\ell;\theta,\theta)\\ &= \Phi_\delta\left(\frac{b-(b^2-4ac)_+^{1/2}}{2|a|}\right) + 1 - \Phi_\delta\left(\frac{b+(b^2-4ac)_+^{1/2}}{2|a|}\right)\\ &= 1 - \Phi\left(\frac{-b+2|a|\mu(\delta)+(b^2-4ac)_+^{1/2}}{2|a|\sigma(\delta)}\right)\\ &\quad + 1 - \Phi\left(\frac{b-2|a|\mu(\delta)+(b^2-4ac)_+^{1/2}}{2|a|\sigma(\delta)}\right)\\ &= 1 - \Phi\left(\frac{-|(b-2|a|\mu(\delta))|+(b^2-4ac)_+^{1/2}}{2|a|\sigma(\delta)}\right)\\ &\quad + 1 - \Phi\left(\frac{|(b-2|a|\mu(\delta))|+(b^2-4ac)_+^{1/2}}{2|a|\sigma(\delta)}\right)\end{aligned}$$

Now use for all $z > 0$, for $x \to \infty$,

$$\frac{1-\Phi(-z+x)}{1-\Phi(z+x)} \sim \frac{\phi(-z+x)}{\phi(z+x)} = e^{2xz} \to \infty$$

so that $1-\Phi(-z+x)+1-\Phi(z+x)\sim 1-\Phi(-z+x)$. As a result, since $|(b-2|a|\mu(1))|=2|\mu_{q,\ell}||\sigma_{q,\ell}^{-2}+|\sigma_{q,\ell}^{-2}-\sigma_0^{-2}||=2|\mu_{q,\ell}|\sigma_0^{-2}$, when $t\to 0^+$,

$$\begin{aligned}\boldsymbol{L}_0(t,q,\ell;\theta,\theta)&\sim 1-\Phi\left(\frac{-2|\mu_{q,\ell}|\sigma_{q,\ell}^{-2}+(b^2-4ac)_+^{1/2}}{2|a|\sigma_0}\right)\\ \boldsymbol{L}_1(t,q,\ell;\theta,\theta)&\sim 1-\Phi\left(\frac{-2|\mu_{q,\ell}|\sigma_0^{-2}+(b^2-4ac)_+^{1/2}}{2|a|\sigma_{q,\ell}}\right)\\ &=1-\Phi\left(\frac{-|\mu_{q,\ell}|}{\sigma_{q,\ell}}+\frac{\sigma_0}{\sigma_{q,\ell}}\frac{-2|\mu_{q,\ell}|\sigma_{q,\ell}^{-2}+(b^2-4ac)_+^{1/2}}{2|a|\sigma_0}\right).\end{aligned}$$

Now use for all $z>0$, $u\in(0,1)$, for $x\to\infty$,

$$\frac{1-\Phi(-z+ux)}{1-\Phi(x)}\sim u^{-1}\frac{\phi(-z+ux)}{\phi(x)}=u^{-1}e^{((1-u^2)x^2+2uzx-z^2)/2}\to\infty,$$

to conclude that

$$\min_{q,\ell:\sigma_{q,\ell}\geq\sigma_0}\left\{\frac{w_{q,\ell}}{1-w_{q,\ell}}\,\frac{\boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\boldsymbol{L}_0(t,q,\ell;\theta,\theta)}\right\},\ \max_{q,\ell:\sigma_{q,\ell}\geq\sigma_0}\left\{\frac{w_{q,\ell}}{1-w_{q,\ell}}\,\frac{\boldsymbol{L}_1(t,q,\ell;\theta,\theta)}{\boldsymbol{L}_0(t,q,\ell;\theta,\theta)}\right\}$$

both tends to infinity when $t\to 0^+$. Hence, $\alpha_c(\theta)=0$ in that case.

### *D.5. Proof of Proposition 5*

It follows from the following result.

**Lemma 8.** *The function $t\in[0,1]\mapsto\boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta)$ is infinitely differentiable on $(t_{1,q,\ell}(\theta),t_{2,q,\ell}(\theta))$, but not differentiable in $t_{1,q,\ell}(\theta)$ when $t_{1,q,\ell}(\theta)>0$ and in $t_{2,q,\ell}(\theta)$ when $t_{2,q,\ell}(\theta)<1$.*

The derivative in $t\notin\{t_{1,q,\ell}(\theta),t_{2,q,\ell}(\theta)\}$ is equal to, when $a\neq 0$,

$$\frac{2\,\mathbb{1}\{b^2>4ac\}}{t(1-t)(b^2-4ac)^{1/2}}\left(\phi_\delta\left(\frac{-b+(b^2-4ac)_+^{1/2}}{2a}\right)+\phi_\delta\left(\frac{-b-(b^2-4ac)_+^{1/2}}{2a}\right)\right),$$

and when $a=0$ (and thus $b\neq 0$),

$$\frac{2}{t(1-t)|b|}\phi_\delta\left(\frac{-c}{b}\right),$$

where $\phi_\delta$ denotes the density of the distribution $\mathcal{N}(\mu(\delta),\sigma(\delta)^2)$. This entails Lemma 8.

In particular, $t\mapsto\boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta^*)$ is differentiable in $t=T_{\theta^*}(\alpha)$ when $T_{\theta^*}(\alpha)\notin\{t_{1,q,\ell}(\theta),t_{2,q,\ell}(\theta)\}$, which proves Proposition 5.

### D.6. Studying $t_2'(\theta)$

In the case where $\theta = (\pi, w, \sigma_0, \mu, \sigma)$ is such that $\forall q, \ell,\ \sigma_{q,\ell} \geq \sigma_0$, we have by Section D.3 that $t_{1,q,\ell}(\theta) = 0$ for all $q, \ell$ and thus

$$\begin{aligned} t_2'(\theta) &= \min\{t_{1,q,\ell}(\theta), t_{2,q,\ell}(\theta),\ \text{for } q,\ell \text{ s.t. } t_{1,q,\ell}(\theta) \neq t_1(\theta)\} \\ &= \min_{q,\ell}\{t_{2,q,\ell}(\theta)\} \end{aligned}$$

If $\forall q, \ell,\ \sigma_{q,\ell} = \sigma_0$, the latter is simply $t_2'(\theta) = 1$. Otherwise, we have

$$\begin{aligned} t_2'(\theta) &= \min\left\{ \left(1 + \frac{w_{q,\ell}}{1 - w_{q,\ell}} \frac{\sigma_0}{\sigma_{q,\ell}} \exp\left(\frac{\mu_{q,\ell}^2}{2(\sigma_0^2 - \sigma_{q,\ell}^2)}\right)\right)^{-1},\ \text{for } q,\ell \text{ s.t. } \sigma_{q,\ell} > \sigma_0 \right\} \\ &\geq \min_{q,\ell} \left(1 + \frac{w_{q,\ell}}{1 - w_{q,\ell}}\right)^{-1} = 1 - \max_{q,\ell}\{w_{q,\ell}\}. \end{aligned}$$

### D.7. Proof of Proposition 6

Denote $\mathcal{C}$ the set (63). Recall that by (39), we have

$$\begin{aligned} \mathcal{W}_{\theta^*, \boldsymbol{L}}(u) = \sup_{q,\ell} \sup_{t \in T_{\theta^*}(\mathcal{K})} \sup_{\delta \in \{0,1\}} \sup \{ &|\boldsymbol{L}_\delta(t, q, \ell; \theta', \theta) - \boldsymbol{L}_\delta(t, q, \ell; \theta^*, \theta^*)| \ : \\ &\theta, \theta' \in \Theta, \|\theta - \theta^*\|_\infty \leq u, \|\theta' - \theta^*\|_\infty \leq u \}. \end{aligned}$$

For short, denote $d(q, \ell, \theta, t) = b^2(q, \ell, \theta) - 4a(q, \ell, \theta)c(q, \ell, \theta, t)$ for any $\theta \in \Theta$ (and $a$, $b$ and $c$ being the quantities defined by (66)), and consider

$$\begin{aligned} \mathcal{B}(\theta^*, \alpha) = \big\{ \theta \in \Theta \ : &\forall (q, \ell) \notin \mathcal{C},\ |a(q, \ell, \theta) - a(q, \ell, \theta^*)| \leq |a(q, \ell, \theta^*)|/2, \\ &\forall t \in T_{\theta^*}(\mathcal{K}),\ |d(q, \ell, \theta, t) - d(q, \ell, \theta^*, t)| \leq |d(q, \ell, \theta^*, t)|/2, \\ &\forall (q, \ell) \in \mathcal{C},\ |b(q, \ell, \theta) - b(q, \ell, \theta^*)| \leq |b(q, \ell, \theta^*)|/2 \big\}. \end{aligned}$$

We check that there exists $\upsilon(\theta^*, \alpha)$ such that for all $\varepsilon \leq \upsilon(\theta^*, \alpha)$, we have

$$\{\theta \in \Theta \ : \ \|\theta - \theta^*\|_\infty \leq \varepsilon\} \subset \mathcal{B}(\theta^*, \alpha). \tag{70}$$

To see this, let $\theta \in \Theta$ with $\|\theta - \theta^*\|_\infty \leq \varepsilon$. For $(q, \ell) \notin \mathcal{C}$, we have $a(q, \ell, \theta^*) \neq 0$ so that, since $\theta \in \Theta \mapsto a(q, \ell, \theta)$ is continuous, for $\varepsilon$ smaller than some positive number $\upsilon_1(q, \ell, \theta^*)$, we have $|a(q, \ell, \theta) - a(q, \ell, \theta^*)| \leq |a(q, \ell, \theta^*)|/2$. In addition, by definition of $T_{\theta^*}(\mathcal{K})$, we have $t_{1,q,\ell}(\theta^*), t_{2,q,\ell}(\theta^*) \notin T_{\theta^*}(\mathcal{K})$, see (69), and thus the sign of $d(q, \ell, \theta, t)$ does not depend on $t \in T_{\theta^*}(\mathcal{K})$. Assume without loss of generality that it is positive, so that $\inf_{t \in T_{\theta^*}(\mathcal{K})} d(q, \ell, \theta^*, t) > 0$. Now, since $T_{\theta^*}(\mathcal{K})$ is a compact set and $(\theta, t) \in \Theta \times T_{\theta^*}(\mathcal{K}) \mapsto d(q, \ell, \theta, t)$ is continuous, we have that $\lim_{\theta \to \theta^*} \sup_{t \in T_{\theta^*}(\mathcal{K})} |d(q, \ell, \theta, t) - d(q, \ell, \theta^*, t)| = 0$ (otherwise, there exists $c > 0, \theta_n \to \theta^*$ and $t_n$ such that for $n$ large, $|d(q, \ell, \theta_n, t_n) - d(q, \ell, \theta^*, t_n)| >$

$c$ and we obtain a contradiction by considering any limit $t_0$ of $t_n$). As a result, there is $\upsilon_2(q,\ell,\theta^*,\alpha)$ such that for $\varepsilon \le \upsilon_2(q,\ell,\theta^*,\alpha)$,

$$\sup_{t\in T_{\theta^*}(\mathcal{K})} |d(q,\ell,\theta,t) - d(q,\ell,\theta^*,t)| \le \inf_{t\in T_{\theta^*}(\mathcal{K})} d(q,\ell,\theta^*,t)/2,$$

that is, $\forall t \in T_{\theta^*}(\mathcal{K})$, $|d(q,\ell,\theta,t)-d(q,\ell,\theta^*,t)| \le d(q,\ell,\theta^*,t)/2$. Finally, consider $(q,\ell) \in \mathcal{C}$. In that case, $a(q,\ell,\theta^*) = 0$ and thus $b(q,\ell,\theta^*) \neq 0$. Since $\theta \in \Theta \mapsto b(q,\ell,\theta)$ is continuous, for $\varepsilon$ smaller than some positive number $\upsilon_3(q,\ell,\theta^*)$, we have $|b(q,\ell,\theta) - b(q,\ell,\theta^*)| \le |b(q,\ell,\theta^*)|/2$. Summing up, we obtain (70) for $\varepsilon \le \upsilon(\theta^*,\alpha) = \min_{(q,\ell)\notin\mathcal{C}}\{\upsilon_1(q,\ell,\theta^*) \wedge \upsilon_2(q,\ell,\theta^*,\alpha)\} \wedge \min_{(q,\ell)\in\mathcal{C}} \upsilon_3(q,\ell,\theta^*)$.

We have for all $\delta \in \{0,1\}$, $\theta,\theta' \in \Theta$ with $\|\theta-\theta^*\|_\infty \le \varepsilon$ and $\|\theta'-\theta^*\|_\infty \le \varepsilon$, for all $q,\ell \in \{1,\dots,Q\}$, for all $t \in T_{\theta^*}(\mathcal{K})$, when $\varepsilon \le \upsilon(\theta^*,\alpha)$,

$$\begin{aligned}&\frac{|\boldsymbol{L}_\delta(t,q,\ell;\theta',\theta) - \boldsymbol{L}_\delta(t,q,\ell;\theta^*,\theta^*)|}{\|\theta-\theta^*\|_\infty \vee \|\theta'-\theta^*\|_\infty}\\ &\le \sup_{q,\ell} \sup_{t\in T_{\theta^*}(\mathcal{K})} \sup_{(\theta',\theta)\in\mathcal{B}(\theta^*,\alpha)^2} \|\nabla_{(\theta',\theta)}\boldsymbol{L}_\delta(t,q,\ell;\cdot,\cdot)\|_\infty,\end{aligned}$$

where $\nabla_{(\theta',\theta)}\boldsymbol{L}_\delta(t,q,\ell;\cdot,\cdot)$ denotes the gradient of the function $(\theta',\theta) \in \mathcal{B}(\theta^*,\alpha)^2 \mapsto \boldsymbol{L}_\delta(t,q,\ell;\theta',\theta)$. Now, thanks to the definition of $\mathcal{B}(\theta^*,\alpha)$, the formulas given in (i)-(ii)-(iii) of Section D.2 are active, without indicator, which means that $(t,\theta',\theta) \in T_{\theta^*}(\mathcal{K}) \times \mathcal{B}(\theta^*,\alpha)^2 \mapsto \nabla_{(\theta',\theta)}\boldsymbol{L}_\delta(t,q,\ell;\cdot,\cdot)$ is continuous and thus for all $\delta \in \{0,1\}$ and $\varepsilon \le \upsilon(\theta^*,\alpha)$, the quantities

$$\sup_{t\in T_{\theta^*}(\mathcal{K})} \sup_{(\theta',\theta)\in\mathcal{B}(\theta^*,\alpha)^2} \|\nabla_{(\theta',\theta)}\boldsymbol{L}_\delta(t,q,\ell;\cdot,\cdot)\|_\infty$$

for $(q,\ell) \in \{1,\dots,Q\}^2$, are below some constant that depends only $\theta^*$ and $\alpha$. This proves Proposition 6.

### *D.8. A useful lemma*

**Lemma 9.** *Let* $\mathcal{D} = \{(a,b,c,\mu,\sigma^2) \in \mathbb{R}^2\times\mathbb{R}\cup\{+\infty\}\times\mathbb{R}\times(0,\infty) : a^2+b^2 \neq 0\}$. *Then the function*

$$(a,b,c,\mu,\sigma^2) \in \mathcal{D} \mapsto \mathbb{P}_{U\sim\mathcal{N}(\mu,\sigma^2)}(aU^2+bU+c<0). \tag{71}$$

*is continuous on* $\mathcal{D}$.

*Proof.* Consider $(a_0,b_0) \in \mathbb{R}^2\backslash\{0\}$, $c_0 \in \mathbb{R}$, $\mu_0 \in \mathbb{R}$, $\sigma_0 > 0$ and some sequence $(a_n,b_n,c_n,\mu_n,\sigma_n)$ with $(a_n,b_n,c_n,\mu_n,\sigma_n) \to (a_0,b_0,c_0,\mu_0,\sigma_0)$ as $n$ tends to infinity. Consider $V \sim \mathcal{N}(0,1)$, $U_n = \mu_n + \sigma_n V$ and $U = \mu_0 + \sigma_0 V$ so that $U_n \sim \mathcal{N}(\mu_n,\sigma_n^2)$ and $U \sim \mathcal{N}(\mu_0,\sigma_0^2)$. Then $a_nU_n^2 + b_nU_n + c_n$ converges to $a_0U^2+b_0U+c_0$ almost surely and thus also in distribution. Since the distribution of $a_0U^2+b_0U+c_0$ is continuous (because $a_0$ and $b_0$ are not both zero), we have by the Portmanteau Lemma that $\mathbb{P}(a_nU_n^2+b_nU_n+c_n<0)$ converges to $\mathbb{P}(a_0U^2+b_0U+c_0<0)$. Finally, a similar reasoning can be applied when $c_0 = +\infty$, with a limit equal to 0. The continuity follows. ☐

## Acknowledgments

We would like to thank the two anonymous referees for their helpful comments and valuable remarks. We are also grateful to Stéphane Robin for interesting discussions.

## References

[1] Christopher Aicher, Abigail Z. Jacobs, and Aaron Clauset. Learning latent block structure in weighted networks. *Journal of Complex Networks*, 3(2):221–248, 06 2014. MR3365464
[2] Elizabeth S. Allman, Catherine Matias, and John A. Rhodes. Parameter identifiability in a class of random graph mixture models. *Journal of Statistical Planning and Inference*, 141(5):1719–1736, 2011. MR2763202
[3] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. MR2091634
[4] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300, 1995. MR1325392
[5] Yoav Benjamini, Abba M. Krieger, and Daniel Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507, 2006. MR2261438
[6] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188, 2001. MR1869245
[7] Peter Bickel, David Choi, Xiangyu Chang, and Hai Zhang. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.*, 41(4):1922–1943, 08 2013. MR3127853
[8] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):719–725, 2000.
[9] Gilles Blanchard and Etienne Roquain. Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.*, 10:2837–2871, 2009. MR2579914
[10] Vincent Brault, Christine Keribin, and Mahendra Mariadassou. Consistency and asymptotic normality of latent blocks model estimators. *Electronic Journal of Statistics*, 14(1):1234–1268, 2020. MR4079457
[11] T. Tony Cai. R codes for the adaptive testing procedure. http://stat.wharton.upenn.edu/ tcai/paper/html/FDR.html.
[12] T. Tony Cai and Weidong Liu. Large-scale multiple testing of correlations. *J. Amer. Statist. Assoc.*, 111(513):229–240, 2016. MR3494655
[13] T. Tony Cai and Wenguang Sun. Simultaneous testing of grouped hypotheses: finding needles in multiple haystacks. *J. Amer. Statist. Assoc.*, 104(488):1467–1481, 2009. MR2597000

[14] T. Tony Cai, Wenguang Sun, and Weinan Wang. Covariate-assisted ranking and screening for large-scale two-sample inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(2):187–234, 2019. MR3928141

[15] Ismaël Castillo and Étienne Roquain. On spike and slab empirical Bayes multiple testing. *Ann. Statist.*, 48(5):2548–2574, 2020. MR4152112

[16] A. Celisse, J.-J. Daudin, and L. Pierre. Consistency of maximum-likelihood and variational estimators in the Stochastic Block Model. *Electron. J. Statist.*, 6:1847–1899, 2012. MR2988467

[17] Jinyuan Chang, E. Kolaczyk, and Q. Yao. Estimation of subgraph densities in noisy networks. *Journal of the American Statistical Association*, pages 1–14, 2020. MR4399091

[18] Zhiyi Chi. On the performance of FDR control: constraints and a partial solution. *Ann. Statist.*, 35(4):1409–1431, 2007. MR2351091

[19] J.-J. Daudin, F. Picard, and S. Robin. A mixture model for random graphs. *Statistics and Computing*, 18(2):173–183, 2008. MR2390817

[20] Mathias Drton, Michael D. Perlman, et al. Multiple testing and error control in gaussian graphical model selection. *Statistical Science*, 22(3):430–449, 2007. MR2416818

[21] Nathan Eagle and Alex (Sandy) Pentland. Reality mining: Sensing complex social systems. *Journal of Personal and Ubiquitous Computing*, 10(4):255–268, March 2006.

[22] Bradley Efron. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Amer. Statist. Assoc.*, 99(465):96–104, 2004. MR2054289

[23] Bradley Efron, Robert Tibshirani, John D. Storey, and Virginia Tusher. Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.*, 96(456):1151–1160, 2001. MR1946571

[24] P.W. Holland, K.B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983. MR0718088

[25] Can M. Le, Keith Levin, and Elizaveta Levina. Estimating a network from multiple noisy realizations. *Electronic Journal of Statistics*, 12(2):4697–4740, 2018. MR3894068

[26] Jie Liu, Chunming Zhang, and David Page. Multiple testing under dependence via graphical models. *Ann. Appl. Stat.*, 10(3):1699–1724, 2016. MR3553241

[27] Weidong Liu. Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.*, 41(6):2948–2978, 2013. MR3161453

[28] David Lusseau. Evidence for social role in a dolphin social network. *Evolutionary Ecology*, 21:357–366, 2007.

[29] Mahendra Mariadassou, Stéphane Robin, and Corinne Vacher. Uncovering latent structure in valued graphs: a variational approach. *Ann. Appl. Stat.*, 4(2):715–742, 2010. MR2758646

[30] T. Martin, B. Ball, and M. Newman. Structural inference for uncertain networks. *Physical Review E*, 93(1):012306, 2016.

[31] P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.*, 18(3):1269–1283, 1990. MR1062069

[32] C. Matias and S. Robin. Modeling heterogeneity in random graphs through latent space models: a selective review. *Esaim Proc. & Surveys*, 47:55–74, 2014. MR3419385

[33] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018. MR3876880

[34] United Nations. https://www.un.org/en/development/desa/population/migration/data, 2019.

[35] M. Newman. Network structure from rich but noisy data. *Nature Physics*, 14:542–545, 2017.

[36] M. Newman. Estimating network structure from unreliable measurements. *Physical Review E*, 98, 2018.

[37] Krzysztof Nowicki and Tom A. B Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001. MR1947255

[38] Tiago P. Peixoto. Nonparametric weighted stochastic block models. *Physical review. E*, 97(1-1):012306, 2018.

[39] Tiago P. Peixoto. Reconstructing networks with unknown and heterogeneous errors. *Phys. Rev. X*, 8:041011, Oct 2018.

[40] Franck Picard, Vincent Miele, Jean-Jacques Daudin, Ludovic Cottret, and Stéphane Robin. Deciphering the connectivity structure of biological networks using mixnet. *BMC Bioinformatics*, 10(6), 2009.

[41] Yannik Pitcan. A Note on Concentration Inequalities for U-Statistics. *arXiv e-prints*, page arXiv:1712.06160, Dec 2017.

[42] T. Schweder and E. Spjøtvoll. Plots of P-values to evaluate many tests simultaneously. *Biometrika*, 69(3):493–502, 1982.

[43] John D. Storey. A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(3):479–498, 2002. MR1924302

[44] John D. Storey. The positive false discovery rate: a Bayesian interpretation and the $q$-value. *Ann. Statist.*, 31(6):2013–2035, 2003. MR2036398

[45] Wenguang Sun and T. Tony Cai. Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.*, 102(479):901–912, 2007. MR2411657

[46] Wenguang Sun and T. Tony Cai. Large-scale multiple testing under dependence. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71(2):393–424, 2009. MR2649603

[47] Henry Teicher. Identifiability of mixtures. *The Annals of Mathematical Statistics*, 32(1):244–248, 1961. MR0120677

[48] Henry Teicher. Identifiability of finite mixtures. *Ann. Math. Statist.*, 34(4):1265–1269, 12 1963. MR0155376

[49] Jichun Xie, T. Tony Cai, John Maris, and Hongzhe Li. Optimal false discovery rate control for dependent data. *Statistics and Its Interface*, 4(4):417, 2011. MR2868825

[50] Jean-Gabriel Young, George T. Cantwell, and M. Newman. Robust bayesian inference of network structure from unreliable data. *J. Complex Networks*, 8, 2021. MR4228651