ALEA, Lat. Am. J. Probab. Math. Stat. 21, 267–292 (2024)
DOI: 10.30757/ALEA.v21-11

Consistent Model Selection for the Degree Corrected Stochastic Blockmodel

Andressa Cerqueira, Sandro Gallo, Florencia Leonardi and Cristel Vera

Departamento de Estatística, Universidade Federal de São Carlos, Rod. Washington Luís km 235, São Carlos, SP, Brasil, CEP 13565-905
E-mail address: acerqueira@ufscar.br

Departamento de Estatística, Universidade Federal de São Carlos, Rod. Washington Luís km 235, São Carlos, SP, Brasil, CEP 13565-905
E-mail address: sandro.gallo@ufscar.br

Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão, 1010, São Paulo, CEP 05508-090
E-mail address: florencia@usp.br

Departamento de Estatística, Universidade Federal de São Carlos, Rod. Washington Luís km 235, São Carlos, SP, Brasil, CEP 13565-905
E-mail address: vera.cristel@gmail.com

Abstract. The Degree Corrected Stochastic Block Model (DCSBM) is a probabilistic model for random networks with community structure in which vertices of the same community are allowed to have distinct degree distributions. On the modeling side, this property makes the DCSBM more suitable for real-life complex networks. On the statistical side, it is more challenging due to a large number of parameters. In this paper, we prove that the penalized marginal likelihood estimator, when assuming prior distributions for the parameters, is strongly consistent for estimating the number of communities. We consider *dense* or *semi-sparse* random networks, and our estimator is *unbounded*, in the sense that the number of communities k considered can be as big as n, the number of nodes in the network.

1. Introduction

Many real-world phenomena can be described by the interaction of objects through a network. For example, interactions between individuals in a social network, connections between airports in

Received by the editors June 22nd, 2023; accepted December 27th, 2023.

2010 *Mathematics Subject Classification.* 62F12, 60F15.

Key words and phrases. Model selection, Stochastic Block Model, marginal likelihood estimator, strong consistency.

This work was supported by the São Paulo Research Foundation (FAPESP), grants 2013/07699-0, 2017/10555-0, 2019/23439-4 and 2023/05857-9, and by the National Council for Scientific and Technological Development (CNPq), grants 2017/10555-0 and 441884/2023-7. FL is partially supported by a CNPq's research fellowship, grant 311763/2020-0. During the realization of this work, CV was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

a country, connections between regions of the brain, etc. Most of these networks have a community structure; that is, the objects (nodes of the network) belonging to the same group tend to behave similarly. In this way, probabilistic models that aim to describe real networks need to incorporate these community structures.

The Stochastic Block Model (SBM) proposed by Holland et al. (1983) is a random network model allowing community structures. The connection between each pair of vertices is, independently of everything, distributed as a Bernoulli random variable with a parameter depending solely on the vertices' communities. The SBM, therefore, models networks where nodes in the same community have the same mean degree. This property can restrict applications to real-life networks that, not rarely, display heterogeneity (*hubs*) in the degree distributions of vertices belonging to the same community. Taking this into account, Karrer and Newman (2011) proposed the Degree-Corrected Stochastic Block Model (DCSBM), which considers the heterogeneity in the nodes' degrees within communities. In the degree-corrected model, each node has associated a non-negative real parameter, a weight, specifying its "ability" to connect to other nodes in the network. The sum of the weights in each community corresponds to the number of nodes belonging to the community, generalizing the homogeneous SBM where we can consider each node as having a weight equal to one. In the DCSBM, the Bernoulli distribution of the existence of an edge between two nodes is replaced with a Poisson distribution, which simplifies the form of the likelihood and allows multiple edges between nodes. In both models, the SBM and the DCSBM, it is a standard approach to study two different regimes, the *dense* and the *semi-sparse* regimes. In the dense regime, the parameters of the distributions governing the number of edges between each pair of nodes are fixed (do not depend on the number n of nodes in the networks), leading to linearly growing expected degrees for each node. In the semi-sparse regime, the parameters are allowed to decrease to zero at a rate ρ_n, and in this case the expected degree is of order $\rho_n n \ll n$.

Several works in the literature have addressed the community detection problem for SBM and DCSBM, where the goal is to estimate the k_0 latent groups of nodes in the network. For the SBM, community detection is proposed based on spectral methods (Rohe et al., 2011; Lei and Rinaldo, 2015; Sarkar and Bickel, 2015), modularity (Newman and Girvan, 2004), likelihood methods (Bickel and Chen, 2009; Celisse et al., 2012; Amini et al., 2013) and under a Bayesian perspective (Decelle et al., 2011; Latouche et al., 2012; van der Pas and van der Vaart, 2018). For the DCSBM, Zhao et al. (2012) study consistency of modularity-based and likelihood-based methods, Qin and Rohe (2013) proposed a regularized spectral clustering algorithm, and Jin (2015) proposed an approach based on the entry-wise ratios between eigenvectors of the adjacency matrix. In order to select the best model between the SBM and DCSBM to fit the data, Yan et al. (2014) proposed an approach based on the likelihood ratio test computed approximately using belief propagation. All these methods assume the number of communities k_0 is known, something that rarely occurs in practice. Estimating the number of communities can be considered a model selection problem.

The literature on estimating the number of communities is more recent and not that extensive, at least from the theoretical point of view. In the case of standard SBM, some approaches include sequential hypothesis tests (Lei, 2016), cross-validation (Chen and Lei, 2018), spectral methods (Le and Levina, 2022), penalized likelihood criteria (Wang and Bickel, 2017; Hu et al., 2020) and Bayesian approaches, as penalized marginal likelihood estimators (Daudin et al., 2008; Biernacki et al., 2010; Latouche et al., 2012; Cerqueira and Leonardi, 2020). Specifically for the DCSBM with n nodes and unknown weights and under the semi-sparse regime with $n^{1/2}\rho_n/\log n \to \infty$, Wang and Bickel (2017) proved the consistency of the penalized likelihood estimator with a penalty function of order $k^2 n \log n$ where k is the number of communities of the candidate model. More recently, Ma et al. (2021) proposed a likelihood ratio test to estimate the number of communities and proved its consistency for the semi-sparse regime where $n\rho_n/\log n$ is sufficiently large. Their approach is based on spectral algorithms and so they assume many further hypotheses to correctly detect the groups. Both approaches assume the number of communities is bounded from above by a known constant.

The present paper considers a penalized marginal likelihood estimator for the number of communities under a DCSBM with unknown weights. The marginal likelihood, also known as *model evidence*, is obtained by marginalizing over the parameters under suitable *a priori* distributions. Then the marginal likelihood is penalized, or regularized, with a penalty that depends on both the number of communities k and the sample size n, as it is done by the general principle known as Bayesian Information Criterion. The obtained estimator can also be seen as a *minimum-description length* principle and is known as *Krichevsky-Trofimov* estimator in the information theory community.

We prove that our estimator equals the correct number of communities k_0 asymptotically almost surely (*i.e.* for a sufficiently large number of vertices n with probability one) in the following context:

The model: we consider the general degree-corrected model as considered by Karrer and Newman (2011) (Poisson number of edges and degree-corrected vertices).

Sparsity: our result holds under the same semi-sparse regime of Ma et al. (2021), where $n\rho_n/\log n$ is sufficiently large. This rate is the phase transition for the exact recovery of the communities, see Abbe (2017).

No upper bound is assumed for k_0: the optimization is made over all possible numbers of communities between 1 and n.

The paper is organized as follows. We define the DCSBM and its associated likelihood function (for known parameters) in Section 2. In Section 3, we introduce the *a priori* distributions for the parameters, define the penalized marginal likelihood estimator, and state our main theorem, the consistency result. Finally, in Section 4 we present the proof of the main result. Technical proofs and other auxiliary results are deferred to the appendix.

2. The Degree Corrected Stochastic Block Model: Definition and Likelihood

For any $n \in \mathbb{Z}_+$, let $X = (X_{ij})_{i,j\in[n]}$ ($[n] := \{1, \ldots, n\}$) denote the symmetric adjacency matrix of a random network on n vertices, with $X_{ij} \in \mathbb{Z}_+$. For each pair $i, j \in [n]$, with $i \neq j$ the variable X_{ij} represents the number of non-oriented edges (or alternatively, the strength of connection) between vertices i and j. For convenience, we define X_{ii} as two times the number of self-loops at vertex i.

The vertices are randomly divided into $k_0 \geq 1$ communities and this community attribution is represented by the vector $Z = (Z_1, \ldots, Z_n)$ of $[k_0]$-valued random variables (i.i.d. with marginal distribution π). We will often use the notation $i \in [a]$ ($i \in [n], a \in [k_0]$) to mean that $Z_i = a$.

In the homogeneous SBM, the expected number of edges between vertices i and j does not depend on the specific vertices but only on the communities. Assuming the number of edges between communities a, b has a Poisson distribution with parameter λ_{ab} we have that

$$\mathbb{E}_\lambda(X_{ij}|Z_i = z_i, Z_j = z_j) = \lambda_{z_i z_j}$$

for any $i \neq j$.

The fact that, within each community, the vertices behave identically, is a disadvantage of the SBM when modeling real-world complex networks. In order to allow different vertices to behave differently inside each community, the degree-corrected SBM (DCSBM) incorporates a weight w_i for each vertex $i \in [n]$ which influences the capacity of the vertex to connect to other vertices. In this case, the expected number of edges between vertices $i \neq j$ is given by

$$\mathbb{E}_{\lambda,w}(X_{ij}|Z_i = z_i, Z_j = z_j) = w_i w_j \lambda_{z_i z_j}$$

and may be different for different nodes in the same community. For this reason such networks are sometimes called *inhomogeneous*.

Let the symmetric matrix $\lambda = (\lambda_{ab})_{a,b\in[k_0]}$ have all entries greater than zero. In the *dense* regime, the matrix λ is fixed (does not depend on n) and has all its entries bounded from below by a positive constant. In this case each node has an expected degree that grows linearly on n, which makes the network over-connected. For this reason, it is interesting to consider a *semi-sparse* regime, where λ is allowed to decrease to zero as a function of n. We take this approach here and we assume that for each n, the distribution of the network on n nodes has parameter $\lambda = \rho_n\tilde{\lambda}$, with $\rho_n \to 0$ as $n \to \infty$ and $\tilde{\lambda}$ a constant symmetric matrix with entries bounded from below by a positive constant. We will give conditions on ρ_n on our main results. For identifiability of the order k_0, i.e the number of communities of the model, we assume that no column in $\tilde{\lambda}$ is proportional to any other column. This is usually assumed in the literature, see for example Ma et al. (2021).

Consider now the DCSBM with $k \in [n]$ communities. In order to compute the joint distribution of (Z, X) we need to define the following counters. For any $z \in [k]^n$ and $a \in [k]$, let $n_a(z)$ be the number of vertices in the network that belong to community a, that is

$$n_a(z) = \sum_{i=1}^{n} \mathbb{1}\{z_i = a\}\,.$$

Following Karrer and Newman (2011) we assume that the vector of weights inside each community satisfies

$$\sum_{i\colon z_i=a} w_i \;=\; n_a(z)$$

for all $a \in [k]$. This implies in particular that the total weight in the network is n, the number of vertices, and putting $w_i \equiv 1$ we retrieve the classical SBM.

For any symmetric matrix $x \in \mathbb{Z}_+^{n\times n}$ (that is, any realization of the network) define also the counter $o_{ab}(x, z)$ as the number of edges between nodes of communities a and b, that is

$$o_{ab}(x,z) = \begin{cases} \sum_{1\le i,j\le n} x_{ij}\mathbb{1}\{z_i = a, z_j = b\} & \text{for } a \neq b\,; \\ \frac{1}{2}\sum_{1\le i,j\le n} x_{ij}\mathbb{1}\{z_i = a, z_j = a\} & \text{for } a = b \end{cases} \tag{2.1}$$

and the degree of node i by

$$d_i(x) = \sum_{1\le j\le n} x_{ij}. \tag{2.2}$$

The total degree on community a is denoted by $d_a^t(x, z)$, and is given by

$$d_a^t(x,z) = \sum_{i\colon z_i=a} d_i(x) = \sum_{1\le i\le n} d_i(x)\mathbb{1}\{z_i = a\} = \sum_{1\le i,j\le n} x_{ij}\mathbb{1}\{z_i = a\}\,.$$

Observe that we have

$$d_a^t(x,z) = \sum_{b\neq a} o_{ab}(x,z) + 2o_{aa}(x,z)$$

and that the number of pairs of nodes in communities a and b, denoted by $n_{ab}(z)$, is given by

$$n_{ab}(z) = \begin{cases} n_a(z)n_b(z) & \text{for } a \neq b\,; \\ \frac{1}{2}n_a(z)^2 & \text{for } a = b. \end{cases} \tag{2.3}$$

We can now write down the joint distribution of (X, Z) for the model with k communities when we are given all the other parameters by

$$\mathbb{P}_{\pi,\lambda,w}(X = x, Z = z) =: \mathbb{P}_{\pi,\lambda,w}(x,z) = \mathbb{P}_{\lambda,w}(x|z)\mathbb{P}_{\pi}(z) \tag{2.4}$$

for $x \in \mathbb{Z}_+^{n\times n}$ and $z \in [k]^n$, where

$$\mathbb{P}_{\lambda,w}(x|z) = \frac{1}{c(x)}\left[\prod_{1\le i\le n} w_i^{\,d_i(x)}\right]\left[\prod_{1\le a\le b\le k} \lambda_{ab}^{o_{ab}(x,z)} \exp\{-n_{ab}(z)\lambda_{ab}\}\right] \tag{2.5}$$

with

$$c(x) := \left[\prod_{i<j} x_{ij}!\right]\left[\prod_{i} 2^{x_{ii}/2}(x_{ii}/2)!\right] \tag{2.6}$$

and

$$\mathbb{P}_\pi(z) = \prod_{1\le a\le k} \pi_a^{n_a(z)}\,. \tag{2.7}$$

The factorization in (2.4) follows by the fact that Z is an i.i.d vector of community attributions with distribution π, and given Z, the matrix X has independent entries with Poisson distribution with means given by λ and w, and depending on Z.

3. Model Selection for the DCSBM

To define the estimator, we introduce convenient *a priori* distributions for the parameters (π, λ, w). Then the hierarchical model distribution of the DCSBM is given by

$$\begin{aligned}
\pi &\sim \text{Dirichlet}(\overbrace{1/2,\dots,1/2}^{k}) \\
\lambda_{ab} &\sim \text{Gamma}(1/2,1)\,, \quad \text{for } a,b\in[k],\ a\le b \\
z_i \,|\, \pi &\sim \pi\,, \quad i\in[n] \\
(w_i)_{i\in[a]} \,|\, z &\sim n_a(z)\,\text{Dirichlet}(\overbrace{1/2,\dots,1/2}^{n_a(z)})\,, \quad a\in[k] \\
x_{ij} \,|\, z_i, z_j, w_i, w_j, \lambda &\sim (2-\mathbb{1}\{i\neq j\})\,\text{Poisson}(w_i w_j \lambda_{z_i,z_j})\,, \quad \text{for } i,j\in[n],\ i\le j\,.
\end{aligned}$$

Denote by Θ_k the space where the hyperparameters $\theta = (\pi, \lambda, w)$ take values and by $\nu_k(\theta)$ the *a priori* distribution over Θ_k.

For any $x \in \mathbb{Z}_+^{n\times n}$, the marginal likelihood $p_k(x)$ is given by the integral

$$p_k(x) = \sum_{z\in[k]^n} \int_{\Theta_k} \mathbb{P}_\theta(x,z)\nu_k(\theta)d\theta\,. \tag{3.1}$$

We can now define the estimator for the number of communities as

$$\widehat{k}_n(x) := \underset{1\le k\le n}{\arg\max}\left\{\, \log p_k(x) - (k^3+3kn)\log(n+1)\,\right\}\,. \tag{3.2}$$

Our estimator shares some features with the estimator proposed by Cerqueira and Leonardi (2020) for the “Bernoulli and homogeneous” SBM. This is mainly because both arise as penalized marginal likelihood estimators. However, there are some substantial differences when considering the more general “Poisson and degree corrected” SBM. First, the marginal likelihood defined by (3.1) is based on a Poisson distribution that incorporates the weights and the degrees corresponding to each node in the network. This allows us to generate networks with an unbounded number of edges between each pair of nodes. Secondly, our estimator (3.2) has an extra penalty term of order $n\log n$, which is needed to compensate for the addition of n parameters w_i, $i = 1,\dots,n$ (the degree corrections).

As in the case of the SBM, the choice of prior distributions does not influence the estimator defined in (3.2) since the parameters are integrated out in calculating the marginal likelihood $p_k(x)$. In fact, the particular choice for the prior was made to simplify the calculations. Also as in the case of the SBM, the computation of the marginal likelihood $p_k(x)$ is infeasible because the sum of the community's assignments grows exponentially with the number of vertices. For the SBM, Latouche et al. (2012) proposed an approximation of the marginal likelihood based on a variational Bayes EM algorithm. The same approximation could be used to approximate the marginal likelihood of the DCSBM, but this goes beyond the scope of this work, which primarily focuses on theoretical aspects of the proposed estimator under sparse regimes.

We can now state our consistency theorem. In words, our estimator (3.2) is strongly consistent for the number of communities of the DCSBM, as the number of vertices grows, considering the sparse regime where $n\rho_n/\log n$ is sufficiently large.

Theorem 3.1. *For the DCSBM with k_0 communities and $\rho_n \geq C\frac{\log n}{n}, n \geq 1$, where C is a sufficiently large constant not depending on n, the estimator defined in (3.2) satisfies*

$$\mathbb{P}_\theta(\widehat{k}_n = k_0) = 1 \tag{3.3}$$

for all sufficiently large n.

The proof of the theorem is given in the next section and several auxiliary results (lemmas or technical calculations) are deferred to the appendix. Since our estimator is inspired by Cerqueira and Leonardi (2020), it is expected that our proofs follow more or less the same path. As they did, we prove successively that our estimator does not overestimate nor underestimate the true parameter asymptotically almost-surely. We also do this by invoking a key proposition (Proposition 4.1 below) which roughly says that the marginal likelihood (3.1) is close to the maximum likelihood. However, the more general framework that we consider forces us to make some substantial modifications along this path. First, since we have an unbounded number of edges in the network, we define a set $\Omega_n, n \geq 1$ of "good" networks on which we can work, leaving aside the bad networks, which are proved to be negligible in Lemma 5.6. Using this idea, the proof of the non-overestimation works quite similarly to their proofs. The proof of the non-underestimation, however, is much more involved due to the non-homogeneity over the vertices of the network. To control the asymptotic behavior of our estimator, we will need explicit concentration bounds for some empirical quantities (observed number of edges and degrees), holding uniformly over the community allocations. This is done in Lemma 5.7. Then, the identifiability hypothesis, namely that $\tilde{\lambda}$ has no two proportional columns, allows us to obtain Lemma 5.9 that implies the non-underestimation of the number of communities.

4. Proof of Theorem 3.1

The proof of Theorem 3.1 is a direct consequence of Propositions 4.2 (non-overestimation) and 4.3 (non-underestimation) which will be stated and proved in Subsection 4.1 and 4.2 respectively. The proofs of these propositions are based on the *key* Proposition 4.1 (see below) relating the marginal likelihood $p_k(x)$ with the maximum likelihood $\sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x)$.

Before stating the key proposition, we define a set of "good" networks (which will be proved to hold with high probability), given by

$$\Omega_n := \{x\colon x_{ij} \leq \log n, \text{ for all } i,j \in [n]\}\,. \tag{4.1}$$

Proposition 4.1. *For all $k \geq 1$, all $n \geq \max(k,3)$ and all $x \in \Omega_n$ we have that*

$$0 \;\leq\; \log\left(\frac{\sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x)}{p_k(x)}\right) \;\leq\; k(k+2)\log(n+1) + 3n\log n\,.$$

The proof of this proposition will be given in the appendix. The remaining of the section is dedicated to prove non-overestimation and non-underestimation.

4.1. *Non-overestimation.* The objective of the present subsection is to prove the following proposition.

Proposition 4.2. *For the DCSBM with k_0 communities, the estimator $\widehat{k}_n$ defined in (3.2) satisfies*

$$\mathbb{P}_\theta(\widehat{k}_n \leq k_0) = 1$$

for all sufficiently large n.

Observe that there is no assumption on ρ_n for this proposition.

Proof: By the Borel-Cantelli Lemma, it is enough to prove that the following series converges

$$\begin{aligned}\sum_{n=1}^{\infty}\sum_{k=k_0+1}^{n}\mathbb{P}_\theta(\widehat{k}_n = k) &= \sum_{n=1}^{\infty}\sum_{k=k_0+1}^{n}\sum_{x\in\Omega_n^c}\mathbb{P}_\theta(x)\mathbb{1}\{\widehat{k}_n(x)=k\}\\ &\quad+\sum_{n=1}^{\infty}\sum_{k=k_0+1}^{n}\sum_{x\in\Omega_n}\mathbb{P}_\theta(x)\mathbb{1}\{\widehat{k}_n(x)=k\}.\end{aligned} \tag{4.2}$$

Let us start by the first term in the right-hand side, and observe that for each n

$$\begin{aligned}\sum_{k=k_0+1}^{n}\sum_{x\in\Omega_n^c}\mathbb{P}_\theta(x)\mathbb{1}\{\widehat{k}_n(x)=k\} &= \sum_{x\in\Omega_n^c}\mathbb{P}_\theta(x)\sum_{k=k_0}^{n}\mathbb{1}\{\widehat{k}_n(x)=k\}\\ &\leq \sum_{x\in\Omega_n^c}\mathbb{P}_\theta(x) = \mathbb{P}_\theta(\Omega_n^c).\end{aligned}$$

An upper bound for $\mathbb{P}_\theta(\Omega_n^c)$ which is summable in n is given in the proof of Lemma 5.6, see (5.30). So we conclude that the first term of (4.2) is indeed summable in n. We will now prove that the second term of (4.2) is also summable in n. First observe that for some fixed $k = k_0+1,\ldots,n$ we have that

$$\begin{aligned}&\sum_{x\in\Omega_n}\mathbb{P}_\theta(x)\mathbb{1}\{\widehat{k}_n(x)=k\}\\ &\qquad= \sum_{x\in\Omega_n}\mathbb{P}_\theta(x)\mathbb{1}\{\arg\max_{\ell}(\log p_\ell(x)-(\ell^3+3\ell n)\log(n+1))=k\}.\end{aligned} \tag{4.3}$$

By the definition of $\widehat{k}_n(x)$, we have, when $k>k_0$, that

$$\begin{aligned}&\Big\{\arg\max_{\ell}[\log\,\log p_\ell(x)-(\ell^3+3\ell n)\log(n+1)]=k\Big\}\\ &\subset\Big\{\log p_k(x)-(k^3+3kn)\log(n+1)\geq\log p_{k_0}(x)-(k_0^3+3k_0 n)\log(n+1)\Big\}.\end{aligned}$$

So (4.3) can be bounded above by

$$\sum_{x\in\Omega_n}\mathbb{P}_\theta(x)\mathbb{1}\big\{\log p_k(x)-(k^3+3kn)\log(n+1)\geq\log p_{k_0}(x)-(k_0^3+3k_0 n)\log(n+1)\big\}$$

and we get

$$\begin{aligned}&\sum_{x\in\Omega_n}\mathbb{P}_\theta(x)\mathbb{1}\{\widehat{k}_n(x)=k\}\\ &\qquad\leq\sum_{x\in\Omega_n}\mathbb{P}_\theta(x)\mathbb{1}\big\{p_{k_0}(x)\leq p_k(x)\exp[(-k^3-3kn+k_0^3+3k_0 n)\log(n+1)]\big\}.\end{aligned} \tag{4.4}$$

Using Proposition 4.1, we have for $x \in \Omega_n$ that

$$\begin{aligned}\log \mathbb{P}_\theta(x) &\leq \log \sup_{\theta\in\Theta^{k_0}} \mathbb{P}_\theta(x) \\ &\leq \log p_{k_0}(x) + k_0(k_0+2)\log(n+1) + 3n\log n, \end{aligned} \tag{4.5}$$

giving that

$$\mathbb{P}_\theta(x) \leq p_{k_0}(x) e^{k_0(k_0+2)\log(n+1)+3n\log n}. \tag{4.6}$$

Then, for $x \in \left\{p_{k_0}(x) \leq p_k(x)\exp[(k_0^3 + 3k_0 n - k^3 - 3kn)\log(n+1)]\right\}$, we have that

$$\begin{aligned}\mathbb{P}_\theta(x) &\leq p_{k_0}(x) e^{k_0(k_0+2)\log(n+1)+3n\log n} \\ &\leq p_k(x) e^{k_0(k_0+2)\log(n+1)+3n\log n + (k_0^3+3k_0 n - k^3 - 3kn)\log(n+1)}. \end{aligned} \tag{4.7}$$

Observe that as $k \geq k_0+1$, the exponent in (4.7) can be upper bounded by

$$(-2k_0^2 - k_0 - 1)\log(n+1) \leq -4\log n\,.$$

Substituting now (4.7) in (4.4) and summing in $k = k_0+1,\ldots,n$ gives that

$$\begin{aligned}\sum_{k=1}^{n}\sum_{x\in\Omega_n} \mathbb{P}_\theta(x)\mathbb{1}\{\widehat{k}_n(x)=k\} &\leq n\, n^{-4}\sum_{x\in\Omega_n} p_k(x) \\ &\leq n^{-3}\end{aligned} \tag{4.8}$$

that is summable in n. This concludes the proof of Proposition 4.2. □

4.2. *Non underestimation.* We conclude the proof of Theorem 3.1 by proving that $\widehat{k}_n(x)$ does not underestimate k_0, the true number of communities. In the proof, we exploit the relationship of the integrated likelihod function $p_k(x)$ and the maximum likelihood function, as stated in Proposition 4.1. First observe that the maximum likelihood function for the joint distribution of (X,Z) satisfies

$$\sup_{\theta\in\Theta_k} \mathbb{P}_\theta(x) = \sup_{\theta\in\Theta_k}\sum_z \mathbb{P}_\theta(x,z) \leq \sum_z \sup_\pi \mathbb{P}_\pi(z) \sup_{\lambda,w} \mathbb{P}_{\lambda,w}(x|z)\,.$$

While it is not possible to derive an explicit expression for the supremum in the left hand side of the above inequality, we do have closed forms for each of the supremum inside the sum in the right hand side. We first observe that for any $x \in \mathbb{Z}_+^{n\times n}, z\in[k]^n, i\in[n]$ and $a,b\in[k]$, the maximum likelihood estimators of π, λ and w are given by

$$\widehat{\pi}_a(z) = \frac{n_a(z)}{n}\,,\quad \widehat{\lambda}_{ab}(x,z) = \frac{o_{ab}(x,z)}{n_{ab}(z)}\,,\quad \widehat{w}_i(x,z) = \sum_{1\leq a\leq k}\mathbb{1}\{z_i=a\}\frac{n_a(z)d_i(x)}{d_a^t(x,z)}\,,$$

and thus

$$\sup_\pi \mathbb{P}_\pi(z) = \prod_{1\leq a\leq k}\left(\frac{n_a(z)}{n}\right)^{n_a(z)} \tag{4.9}$$

and

$$\begin{aligned}\sup_{\lambda,w} \mathbb{P}_{\lambda,w}(x|z) = \frac{1}{c(x)} &\prod_{1\leq a\leq b\leq k}\left(\frac{o_{ab}(x,z)}{n_{ab}(z)}\right)^{o_{ab}(x,z)} e^{-o_{ab}(x,z)} \\ &\times \prod_{a\in[k],i\in[n]:z_i=a}\left(\frac{n_a(z)d_i(x)}{d_a^t(x,z)}\right)^{d_i(x)}.\end{aligned} \tag{4.10}$$

These closed forms will be used in the proof of the proposition we state in the sequel.

Proposition 4.3. *For the DCSBM with k_0 communities and $\rho_n \geq C\frac{\log n}{n}, n \geq 1$, where C is a sufficiently large constant not depending on n, the estimator defined in (3.2) satisfies*

$$\mathbb{P}_\theta(\widehat{k}_n \geq k_0) = 1$$

for all sufficiently large n.

Proof: We define the profile estimator for the communities based on the observed graph under the model with k communities as

$$\widehat{z}_k = \underset{z\in\{1,2,\ldots,k\}^n}{\arg\max} \sup_{\theta\in\Theta_k} \mathbb{P}_\theta(x,z)\,. \tag{4.11}$$

By Lemma 5.6 we can (and henceforth will) take n sufficiently large so that $x \in \Omega_n$, the set of "good" networks defined in (4.1). In order to show that $\widehat{k}_n(x) \geq k_0$, almost surely when $n \to \infty$, it is sufficient to show that for all $k < k_0$,

$$\log p_{k_0}(x) - (k_0^3 + 3k_0 n)\log(n+1) \;>\; \log p_k(x) - (k^3 + 3kn)\log(n+1)\,, \tag{4.12}$$

almost surely, when $n \to \infty$. But, if we show that

$$\liminf_{n\to\infty} \frac{1}{\rho_n n^2} \log \frac{p_{k_0}(x)}{p_k(x)} > 0\,, \tag{4.13}$$

and due to the fact that

$$\frac{1}{\rho_n n^2}\left(k_0^3 + 3k_0 n - k^3 - 3kn\right)\log(n+1)$$

can be made sufficiently small by assumption on ρ_n, it follows that (4.13) implies (4.12). Observe that for all $x \in \Omega_n$, by Proposition 4.1 and the fact that $p_k(x) \leq \sup_{\theta\in\Theta_k} \mathbb{P}_\theta(x)$, we have that

$$\begin{aligned}
\frac{1}{\rho_n n^2}\log\frac{p_{k_0}(x)}{p_k(x)} \;&=\; \frac{1}{\rho_n n^2}\log\frac{p_{k_0}(x)}{\sup_{\theta\in\Theta_{k_0}}\mathbb{P}_\theta(x)} + \frac{1}{\rho_n n^2}\log\frac{\sup_{\theta\in\Theta_{k_0}}\mathbb{P}_\theta(x)}{\sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x)} \\
&\quad + \frac{1}{\rho_n n^2}\log\frac{\sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x)}{p_k(x)} \\
&\geq\; -\gamma(k_0)\frac{\log n}{\rho_n n} + \frac{1}{\rho_n n^2}\log\frac{\sup_{\theta\in\Theta_{k_0}}\mathbb{P}_\theta(x)}{\sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x)}\,,
\end{aligned}$$

for some constant $\gamma(k_0)$. Then, to show (4.13) it is enough to prove that for $k < k_0$

$$\liminf_{n\to\infty} \frac{1}{\rho_n n^2}\log\frac{\sup_{\theta\in\Theta_{k_0}}\mathbb{P}_\theta(x)}{\sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x)} \;>\; 0 \tag{4.14}$$

as $-\gamma(k_0)\frac{\log n}{\rho_n n}$ is also sufficiently small by hypothesis on ρ_n. First observe that for all $z \in [k_0]^n$ we have that

$$\log \sup_{\theta\in\Theta_{k_0}} \mathbb{P}_\theta(x) \;\geq\; \log \sup_{\theta\in\Theta_{k_0}} \mathbb{P}_\theta(x,z)\,. \tag{4.15}$$

Let

$$\tilde{o}_{ab}(x,z) := \sum_{1\leq i,j\leq n} x_{ij}\mathbb{1}\{z_i = a, z_j = b\}$$

for all pairs a, b and notice is relates to o_{ab} by $\tilde{o}_{ab}(x,z) = o_{ab}(x,z)$ for all $a \neq b$ and $\tilde{o}_{aa}(x,z) = 2o_{aa}(x,z)$ for all a. From (2.4)-(2.6) and the definition of the maximum likelihood estimators we have that

$$\begin{aligned}\log \sup_{\theta\in\Theta_{k_0}} \mathbb{P}_\theta(x,z) &= L(x) + n\sum_{a=1}^{k_0} \widehat{\pi}_a(z)\log\widehat{\pi}_a(z)\\ &\quad + \frac{1}{2}\sum_{1\le a,b\le k_0} \tilde{o}_{ab}(x,z)\log\widehat{\lambda}_{ab}(x,z) + \sum_{1\le a\le k_0} d_a^t(x,z)\log\frac{n_a(z)}{d_a^t(x,z)}\end{aligned} \tag{4.16}$$

with

$$\begin{aligned} L(x) &= -\log c(x) - \sum_{i,j} x_{ij} + \sum_i d_i(x)\log d_i(x)\,;\\ \widehat{\pi}_a(z) &= \frac{n_a(z)}{n}, \qquad 1\le a\le k_0;\\ \widehat{\lambda}_{ab}(x,z) &= \frac{\tilde{o}_{ab}(x,z)}{n_a(z)n_b(z)}, \qquad 1\le a,b\le k_0;\\ d_a^t(x,z) &= \sum_{1\le i,j\le n} x_{ij}\mathbb{1}\{z_i=a\} = \sum_b \tilde{o}_{ab}(x,z), \qquad 1\le a\le k_0\,.\end{aligned}$$

For the denominator in (4.14) we use that

$$\begin{aligned}\log \sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x) &\le \log k^n \sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x,\widehat{z}_k)\\ &\le n\log k + \log\sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x,\widehat{z}_k)\,,\end{aligned} \tag{4.17}$$

with $\widehat{z}_k$ defined by (4.11). Analogously as in (4.16) we have that

$$\begin{aligned}\log\sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x,\widehat{z}_k) = L(x) &+ n\sum_{1\le a\le k}\widehat{\pi}_a(\widehat{z}_k)\log\widehat{\pi}_a(\widehat{z}_k)\\ &+ \frac{1}{2}\sum_{1\le a,b\le k}\tilde{o}_{ab}(\widehat{z}_k)\log\widehat{\lambda}_{ab}(x,\widehat{z}_k)\\ &+ \sum_{1\le a\le k} d_a^t(x,\widehat{z}_k)\log\frac{n_a(\widehat{z}_k)}{d_a^t(x,\widehat{z}_k)}\,.\end{aligned} \tag{4.18}$$

Then, the logarithm in (4.14) can be lower bounded by the difference of (4.15) and (4.17), and using the expressions in (4.16) and (4.18) we obtain that

$$\begin{aligned}\log\frac{\sup_{\theta\in\Theta_{k_0}}\mathbb{P}_\theta(x)}{\sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x)} &\ge n\sum_{1\le a\le k_0}\widehat{\pi}_a(z)\log\widehat{\pi}_a(z) + \frac{1}{2}\sum_{1\le a,b\le k_0}\tilde{o}_{ab}(x,z)\log\widehat{\lambda}_{ab}(x,z)\\ &\quad + \sum_{1\le a\le k_0} d_a^t(x,z)\log\frac{n_a(z)}{d_a^t(x,z)} - n\log k\\ &\quad - n\sum_{1\le a\le k}\widehat{\pi}_a(\widehat{z}_k)\log\widehat{\pi}_a(\widehat{z}_k) - \frac{1}{2}\sum_{1\le a,b\le k}\tilde{o}_{ab}(x,\widehat{z}_k)\log\widehat{\lambda}_{ab}(x,\widehat{z}_k)\\ &\quad - \sum_{1\le a\le k} d_a^t(x,\widehat{z}_k)\log\frac{n_a(\widehat{z}_k)}{d_a^t(x,\widehat{z}_k)}\,.\end{aligned} \tag{4.19}$$

We will now rearrange the six terms of the right-hand side. First, let

$$\mathcal{A}(n) := n\sum_{1\le a\le k_0}\widehat{\pi}_a(z)\log\widehat{\pi}_a(z) - n\log k - n\sum_{1\le a\le k}\widehat{\pi}_a(\widehat{z}_k)\log\widehat{\pi}_a(\widehat{z}_k). \tag{4.20}$$

Second, we can write

$$\begin{aligned}\sum_{1\le a\le k_0} d_a^t(x,z)\log\frac{n_a(z)}{d_a^t(x,z)} &= \frac{1}{2}\sum_{1\le a\le k_0} d_a^t(x,z)\log\frac{n_a(z)}{d_a^t(x,z)}\\ &\quad+\frac{1}{2}\sum_{1\le b\le k_0} d_b^t(x,z)\log\frac{n_b(z)}{d_b^t(x,z)}\\ &= \frac{1}{2}\sum_{1\le a,b\le k_0}\tilde{o}_{ab}(x,z)\log\frac{n_a(z)n_b(z)}{d_a^t(x,z)d_b^t(x,z)}\end{aligned} \tag{4.21}$$

and therefore

$$\begin{aligned}\frac{1}{2}\sum_{1\le a,b\le k_0}\tilde{o}_{ab}(x,z)\log\widehat{\lambda}_{ab}(x,z)+\sum_{1\le a\le k_0} d_a^t(x,z)\log\frac{n_a(z)}{d_a^t(x,z)}\\ =\frac{1}{2}\sum_{1\le a,b\le k_0}\tilde{o}_{ab}(x,z)\log\frac{\tilde{o}_{ab}(x,z)}{d_a^t(x,z)d_b^t(x,z)}.\end{aligned} \tag{4.22}$$

Using (4.20), (4.22) and the counterpart of (4.22) under the k-th order model with $\widehat{z}_k$ instead of z, the right-hand side of (4.19) now reads

$$\begin{aligned}\log\frac{\sup_{\theta\in\Theta_{k_0}}\mathbb{P}_\theta(x)}{\sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x)} \;\ge\; \mathcal{A}(n)+\frac{1}{2}\sum_{1\le a,b\le k_0}\tilde{o}_{ab}(x,z)\log\frac{\tilde{o}_{ab}(x,z)}{d_a^t(x,z)d_b^t(x,z)}\\ -\frac{1}{2}\sum_{1\le a,b\le k}\tilde{o}_{ab}(x,\widehat{z}_k)\log\frac{\tilde{o}_{ab}(x,\widehat{z}_k)}{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}\,.\end{aligned}$$

Now, dividing both sides of (4.19) by $\rho_n n^2$, and summing on the right-hand side the following term (which equals 0)

$$\frac{1}{\rho_n n^2}\left(\frac{1}{2}\sum_{1\le a,b\le k_0}\tilde{o}_{ab}(x,z)\log\rho_n n^2-\frac{1}{2}\sum_{1\le a,b\le k}\tilde{o}_{ab}(x,\widehat{z}_k)\log\rho_n n^2\right)$$

we finally obtain that

$$\begin{aligned}\frac{1}{\rho_n n^2}\log\frac{\sup_{\theta\in\Theta_{k_0}}\mathbb{P}_\theta(x)}{\sup_{\theta\in\Theta_k}\mathbb{P}_\theta(x)} \;\ge\; \frac{1}{2}\Bigg(\sum_{1\le a,b\le k_0}\frac{\tilde{o}_{ab}(x,z)}{\rho_n n^2}\log\frac{\rho_n n^2\tilde{o}_{ab}(x,z)}{d_a^t(x,z)d_b^t(x,z)}\\ -\sum_{1\le a,b\le k}\frac{\tilde{o}_{ab}(x,\widehat{z}_k)}{\rho_n n^2}\log\frac{\rho_n n^2\tilde{o}_{ab}(x,\widehat{z}_k)}{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}\Bigg)+\frac{\mathcal{A}(n)}{\rho_n n^2}.\end{aligned} \tag{4.23}$$

Since $\frac{\mathcal{A}(n)}{\rho_n n^2}$ converges almost surely to 0, proving that (4.23) is bounded from below by a positive constant, eventually almost surely as $n\to\infty$, is equivalent to proving that

$$\begin{aligned}\liminf_{n\to\infty}\sum_{1\le a,b\le k_0}\frac{d_a^t(x,z)d_b^t(x,z)}{\rho_n^2 n^4}\varphi\Big(\frac{\rho_n n^2\tilde{o}_{ab}(x,z)}{d_a^t(x,z)d_b^t(x,z)}\Big)\\ -\sum_{1\le a,b\le k}\frac{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}{\rho_n^2 n^4}\varphi\Big(\frac{\rho_n n^2\tilde{o}_{ab}(x,\widehat{z}_k)}{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}\Big)\;>\;0\,,\end{aligned} \tag{4.24}$$

with $\varphi(u) = u \log u$. By Lemma 5.7 we have that

$$\begin{aligned} \frac{\rho_n n^2 \tilde{o}_{ab}(x,z)}{d_a^t(x,z) d_b^t(x,z)} &= \frac{\tilde{o}_{ab}(x,z)/\rho_n n^2}{d_a^t(x,z) d_b^t(x,z)/\rho_n^2 n^4} \\ &\to \frac{[\mathrm{diag}(\pi)\tilde{\lambda}\mathrm{diag}(\pi)^T]_{ab}}{[\mathrm{diag}(\pi)\tilde{\lambda}\mathrm{diag}(\pi)^T \mathbf{1}_k]_a [\mathrm{diag}(\pi)\tilde{\lambda}\mathrm{diag}(\pi)^T \mathbf{1}_k]_b} \\ &= \frac{\pi_a \pi_b \tilde{\lambda}_{ab}}{\pi_a [\tilde{\lambda}\pi]_b \pi_b [\tilde{\lambda}\pi]_a} \\ &= \frac{\tilde{\lambda}_{ab}}{[\tilde{\lambda}\pi]_a [\tilde{\lambda}\pi]_b} \end{aligned} \tag{4.25}$$

where we recall that $\tilde{\lambda}$ is the matrix such that $\lambda = \rho_n \tilde{\lambda}$. Then we have that

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{2} \sum_{1\le a,b\le k_0} \frac{d_a^t(x,z) d_b^t(x,z)}{\rho_n^2 n^4} \, \varphi\Big(\frac{\rho_n n^2 \tilde{o}_{ab}(x,z)}{d_a^t(x,z) d_b^t(x,z)} \Big) \\ = \frac{1}{2} \sum_{1\le a,b\le k_0} \pi_a \pi_b [\tilde{\lambda}\pi]_a [\tilde{\lambda}\pi]_b \, \varphi\Big(\frac{\tilde{\lambda}_{ab}}{[\tilde{\lambda}\pi]_a [\tilde{\lambda}\pi]_b} \Big) \, . \end{aligned} \tag{4.26}$$

On the other hand, by Lemma 5.8 we have that

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{2} \sum_{1\le a,b\le k} \frac{d_a^t(x,\widehat{z}_k) d_b^t(x,\widehat{z}_k)}{\rho_n^2 n^4} \, \varphi\Big(\frac{\rho_n n^2 \tilde{o}_{ab}(x,\widehat{z}_k)}{d_a^t(x,\widehat{z}_k) d_b^t(x,\widehat{z}_k)} \Big) \\ \le \frac{1}{2} \sum_{1\le a,b\le k} \pi_a^* \pi_b^* [\lambda^*\pi^*]_a [\lambda^*\pi^*]_b \, \varphi\Big(\frac{\lambda_{ab}^*}{[\lambda^*\pi^*]_a [\lambda^*\pi^*]_b} \Big) \end{aligned} \tag{4.27}$$

for some $k \times k$ positive matrix λ^* and k dimensional vector π^* defined by (5.37). Finally, by Lemma 5.9 we have that the difference of (4.26) and (4.27) is lower bounded by

$$\begin{aligned} \frac{1}{2} \Bigg(\sum_{1\le a,b\le k_0} \pi_a \pi_b [\tilde{\lambda}\pi]_a [\tilde{\lambda}\pi]_b \, \varphi\Big(\frac{\tilde{\lambda}_{ab}}{[\tilde{\lambda}\pi]_a [\tilde{\lambda}\pi]_b} \Big) \\ - \sum_{1\le a,b\le k} \pi_a^* \pi_b^* [\lambda^*\pi^*]_a [\lambda^*\pi^*]_b \, \varphi\Big(\frac{\lambda_{ab}^*}{[\lambda^*\pi^*]_a [\lambda^*\pi^*]_b} \Big) \Bigg) \; > \; 0 \end{aligned}$$

unless $\tilde{\lambda}$ has two proportional columns, which contradicts the hypothesis of identifiability of k_0. This concludes the proof of Proposition 4.3. □

Acknowledgments

This work was produced as part of the activities of the *Research, Innovation and Dissemination Center for Neuromathematics* (grant FAPESP 2013/07699-0). It was also supported by FAPESP project (grant 2017/10555-0) *"Stochastic Modeling of Interacting Systems"* and CNPq Universal project (grant 432310/2018-5) *"Statistics, stochastic processes and discrete structures"*. FL is partially supported by a CNPq's research fellowship, grant 311763/2020-0. During the realization of this work, CV was supported by a CAPES Ph.D. fellowship.

5. Appendix

5.1. *Basic results.* We state below Lemmas 5.1 and 5.2 for completeness.

Lemma 5.1. *For integers $m = m_1 + \cdots + m_J$ we have that*

$$\frac{\prod_{j=1}^{J} \left(\frac{m_j}{m}\right)^{m_j}}{\prod_{j=1}^{J} \Gamma\left(m_j + \frac{1}{2}\right)} \leq \frac{1}{\Gamma\left(m + \frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)^{J-1}}. \tag{5.1}$$

Proof: For an integer m, we have that

$$\Gamma\left(m + \frac{1}{2}\right) = (m+1)(m+2)\ldots(2m)\frac{\sqrt{\pi}}{2^m}.$$

Thus, for integers m_j, $j = 1, \ldots, J$, such that $m = m_1 + \cdots + m_J$ we write

$$\frac{\prod_{j=1}^{J} \Gamma\left(m_j + \frac{1}{2}\right)}{\Gamma\left(m + \frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)^{J-1}} = \frac{\prod_{j=1}^{J} (m_j+1)(m_j+2)\cdots(2m_j)}{(m+1)(m+2)\cdots(2m)}. \tag{5.2}$$

Define, for $y \geq 0$ and an integer $r \geq 1$

$$g_r(y) = \prod_{i=1}^{r} \left(y + \frac{i}{r}\right).$$

As it is shown in the Lemma included in the Appendix of Davisson et al. (1981), for integers $m = m_1 + \cdots + m_J$ and $y \geq 0$ we have that

$$g_m(y) \leq \prod_{j=1}^{J} g_{m_j}(y). \tag{5.3}$$

Using this result for $y = 1$ and $r = m$ we have that

$$\frac{\prod_{i=1}^{m} (m+i)}{m^m} \leq \prod_{j=1}^{J} \prod_{i=1}^{m_j} \frac{(m_j+i)}{m_j} = \frac{\prod_{j=1}^{J} (m_j+1)(m_j+2)\cdots(2m_j)}{\prod_{j=1}^{J} m_j^{m_j}}. \tag{5.4}$$

Rearranging (5.4) and combining with (5.2) we conclude that

$$\prod_{j=1}^{J} \left(\frac{m_j}{m}\right)^{m_j} \leq \frac{\prod_{j=1}^{J} (m_j+1)(m_j+2)\cdots(2m_j)}{(m+1)(m+2)\cdots(2m)} = \frac{\prod_{j=1}^{J} \Gamma\left(m_j + \frac{1}{2}\right)}{\Gamma\left(m + \frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)^{J-1}}.$$

□

Lemma 5.2. *For integers $m = m_1 + \cdots + m_J$, with $J \geq 1$ and $m \geq \max(J, 3)$, we have that*

$$\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(m + \frac{J}{2}\right)}{\Gamma\left(\frac{J}{2}\right)\Gamma\left(m + \frac{1}{2}\right)} \leq m^J.$$

Proof: Stirlings' formula for the Γ function states that for all $y \geq 0$ we have

$$y^{y-\frac{1}{2}}e^{-y}\sqrt{2\pi} \leq \Gamma(y) \leq y^{y-\frac{1}{2}}e^{-y}\sqrt{2\pi}e^{\frac{1}{12y}}.$$

Then

$$\begin{aligned}
\log\left(\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(m+\frac{J}{2}\right)}{\Gamma\left(\frac{J}{2}\right)\Gamma\left(m+\frac{1}{2}\right)}\right) \;\leq\; & \left(m+\frac{J-1}{2}\right)\log\left(m+\frac{J}{2}\right)-\left(m+\frac{J}{2}\right) \\
& + \frac{1}{12\left(m+\frac{J}{2}\right)} - m\log\left(m+\frac{1}{2}\right)+\left(m+\frac{1}{2}\right)+\log\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{J}{2})} \\
\leq\; & \left(m+\frac{J-1}{2}\right)\log\left(m\left(1+\frac{J}{2m}\right)\right)+\frac{1}{12m} \\
& - m\log\left(m\left(1+\frac{1}{2m}\right)\right)-\frac{J-1}{2}+\log\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{J}{2})} \\
\leq\; & \left(\frac{J-1}{2}\right)\log m+\left(m+\frac{J-1}{2}\right)\log\left(1+\frac{J}{2m}\right) \\
& - m\log\left(1+\frac{1}{2m}\right)+\frac{1}{12m}-\frac{J-1}{2}+\log\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{J}{2})}\,.
\end{aligned} \tag{5.5}$$

Using that $1-\frac{1}{y}\leq\log y\leq y-1$, for $y>0$ we obtain that

$$\begin{aligned}
\log\left(\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(m+\frac{J}{2}\right)}{\Gamma\left(\frac{J}{2}\right)\Gamma\left(m+\frac{1}{2}\right)}\right) \leq & \left(\frac{J-1}{2}\right)\log m+\left(m+\frac{J-1}{2}\right)\left(\frac{J}{2m}\right)+\frac{1}{12m}-\frac{J-1}{2} \\
& +\log\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{J}{2})}\leq\left(\frac{J-1}{2}\right)\log m+\frac{J(J-1)}{4m}+\frac{1}{12m}+\log\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{J}{2})}\,.
\end{aligned}$$

Observe that for $J\geq 1$ and $m\geq\max(J,3)$ we have that

$$\log\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{J}{2})}\;\leq\;\log(2)$$

and

$$\frac{J(J-1)}{4m}\;\leq\;\frac{J-1}{4}\;\leq\;\frac{J-1}{4}\log m\,.$$

Then

$$\log\left(\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(m+\frac{J}{2}\right)}{\Gamma\left(\frac{J}{2}\right)\Gamma\left(m+\frac{1}{2}\right)}\right)\;\leq\;\left(\frac{J-1}{2}\right)\log m+\frac{J-1}{4}\log m+1\leq\;J\log m\,.$$

□

5.2. *Proof of Proposition 4.1.* We recall that Θ_k denotes the space of hiperparameters $\theta=(\pi,\lambda,w)$. The parameter w depends on z through the counters $\{n_a(z)\colon a\in[k]\}$, then to make this dependence explicit, we write

$$\Theta_k=\Pi_k\times\Lambda_k\times\mathcal{W}(z)$$

where

$$\Pi_k:=\{(\pi_1,\ldots,\pi_k)\in(0,1]^k\colon\;\sum_i\pi_i=1\}$$

is the standard k-dimensional simplex,

$$\Lambda_k:=\{\lambda\in(\mathbb{R}^+)^{k\times k}\colon\lambda_{ab}=\lambda_{ba}\text{ for all }a,b\in[k]\}$$

is the set of $k\times k$ symmetric matrices with positive entries and

$$\mathcal{W}(z)=\mathcal{W}_1(z)\times\mathcal{W}_2(z)\times\cdots\times\mathcal{W}_k(z)$$

with

$$\mathcal{W}_a(z):=\{w\in(\mathbb{R}^+)^{n_a(z)}:\sum_i w_i=n_a(z)\}\tag{5.6}$$

which is the set of possible w's on community a. By the definition of the model we have that the *a priori* distribution over Θ_k is given by

$$\nu_k(\theta) = \nu_k^{(1)}(\pi)\nu_k^{(2)}(\lambda)\nu^{(3)}(w|z) \tag{5.7}$$

where

$$\nu_k^{(1)}(\pi) = \frac{\Gamma(k/2)}{\Gamma(1/2)^k} \prod_{1\le a\le k} {\pi_a}^{-1/2},$$

$$\nu_k^{(2)}(\lambda) = \frac{1}{\Gamma(1/2)^{\frac{k(k+1)}{2}}} \prod_{1\le a\le b\le k} \lambda_{ab}^{-1/2} e^{-\lambda_{ab}}$$

and

$$\nu^{(3)}(w|z) = \prod_{1\le a\le k} \frac{\Gamma\left(\frac{n_a(z)}{2}\right)}{n_a(z)\Gamma\left(\frac{1}{2}\right)^{n_a(z)}} \prod_{i:\, z_i=a} \left(\frac{w_i}{n_a(z)}\right)^{-1/2}.$$

We can now decompose the marginal likelihood as

$$\begin{aligned}
p_k(x) &= \sum_{z\in[k]^n} \int_{\Theta^k} \mathbb{P}_\theta(x,z)\nu_k(\theta)d\theta \\
&= \sum_{z\in[k]^n} \int_{\mathcal{W}(z)} \int_{\Lambda^k} \mathbb{P}_{\lambda,w}(x|z)\nu_k^{(2)}(\lambda)\nu^{(3)}(w|z)d\lambda dw \int_{\Delta^k} \mathbb{P}_\pi(z)\nu_k^{(1)}(\pi)d\pi \\
&=: \sum_{z\in[k]^n} p_k(x|z)p_k(z)
\end{aligned}$$

in which the (conditional) likelihoods were given in (2.7) and (2.5). Then we have that

$$\begin{aligned}
&\int_{\Lambda^k} \mathbb{P}_{\lambda,w}(x|z)\nu_k^{(2)}(\lambda)d\lambda \\
&\quad = \int_{\Lambda^k} \frac{1}{c(x)} \left[\prod_{1\le i\le n} {w_i}^{d_i(x)}\right] \left[\prod_{1\le a\le b\le k} \lambda_{ab}^{o_{ab}(x,z)} e^{-n_{ab}(z)\lambda_{ab}}\right] \nu_k^{(2)}(\lambda)d\lambda \\
&\quad = \frac{1}{c(x)\Gamma(\frac{1}{2})^{\frac{k(k+1)}{2}}} \left[\prod_{1\le i\le n} {w_i}^{d_i(x)}\right] \int_{\Lambda^k} \prod_{1\le a\le b\le k} \lambda_{ab}^{o_{ab}(x,z)-\frac{1}{2}} e^{-(n_{ab}(z)+1)\lambda_{ab}} d\lambda \\
&\quad = \frac{1}{c(x)} \underbrace{\frac{1}{\Gamma(\frac{1}{2})^{\frac{k(k+1)}{2}}} \left[\prod_{1\le a\le b\le k} \frac{\Gamma\left(o_{ab}(x,z)+\frac{1}{2}\right)}{[n_{ab}(z)+1]^{o_{ab}(x,z)+\frac{1}{2}}}\right]}_{A(x,z)} \left[\prod_{1\le i\le n} {w_i}^{d_i(x)}\right].
\end{aligned} \tag{5.8}$$

Therefore

$$
\begin{aligned}
p_k(x|z) &= \frac{A(x,z)}{c(x)} \int_{\mathcal{W}} \prod_{1\le i\le n} {w_i}^{d_i(x)} \nu^{(3)}(w|z)dw \\
&= \frac{A(x,z)}{c(x)} \prod_{1\le a\le k} \frac{\Gamma\left(\frac{n_a(z)}{2}\right)}{n_a(z)\Gamma\left(\frac{1}{2}\right)^{n_a(z)}} \int_{\mathcal{W}_a(z)} \prod_{i\colon z_i=a} {w_i}^{d_i(x)} \Big(\frac{w_i}{n_a(z)}\Big)^{-1/2} dw_a \\
&= \frac{A(x,z)}{c(x)} \prod_{1\le a\le k} \frac{\Gamma\left(\frac{n_a(z)}{2}\right) n_a(z)^{d_a^t(x,z)-1}}{\Gamma\left(\frac{1}{2}\right)^{n_a(z)}} \int_{\mathcal{W}_a(z)} \prod_{i\colon z_i=a} \Big(\frac{w_i}{n_a(z)}\Big)^{d_i(x)-1/2} dw_a \\
&= \frac{A(x,z)}{c(x)} \prod_{1\le a\le k} \frac{\Gamma\left(\frac{n_a(z)}{2}\right) n_a(z)^{d_a^t(x,z)}}{\Gamma\left(\frac{1}{2}\right)^{n_a(z)}} \int_{\mathcal{Y}_a(z)} \prod_{i\colon z_i=a} {y_i}^{d_i(x)-1/2} dy \\
&= \frac{A(x,z)}{c(x)} \underbrace{\prod_{1\le a\le k} \frac{n_a(z)^{d_a^t(x,z)}\Gamma\left(\frac{n_a(z)}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^{n_a(z)}} \frac{\prod_{i\colon z_i=a}\Gamma(d_i(x)+\frac{1}{2})}{\Gamma(d_a^t(x,z)+\frac{n_a(z)}{2})}}_{B(x,z)} .
\end{aligned}
\tag{5.9}
$$

We also have that

$$
\begin{aligned}
p_k(z) &= \int_{\Pi^k} \prod_{a=1}^{k} {\pi_a}^{n_a(z)} \nu_k^{(1)}(\pi)d\pi \\
&= \frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k} \int_{\Pi^k} \prod_{a=1}^{k} {\pi_a}^{n_a(z)-\frac{1}{2}} d\pi \\
&= \underbrace{\frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k} \frac{\prod_{a=1}^{k}\Gamma\left(n_a(z)+\frac{1}{2}\right)}{\Gamma\left(n+\frac{k}{2}\right)}}_{C(z)} .
\end{aligned}
\tag{5.10}
$$

On the other hand, by the definition of the maximum likelihood estimators we have that

$$
\begin{aligned}
\sup_{\lambda,w} \mathbb{P}_{\lambda,w}(x|z) = \frac{1}{c(x)} & \underbrace{\Big[\prod_{1\le a\le b\le k} \left(\frac{o_{ab}(x,z)}{n_{ab}(z)}\right)^{o_{ab}(x,z)} e^{-o_{ab}(x,z)}\Big]}_{\widehat{A}(x,z)} \\
& \times \underbrace{\Big[\prod_{a\in[k], i\in[n]: z_i=a} \left(\frac{n_a(z)d_i(x)}{d_a^t(x,z)}\right)^{d_i(x)}\Big]}_{\widehat{B}(x,z)}
\end{aligned}
\tag{5.11}
$$

and

$$
\sup_{\pi} \mathbb{P}_{\pi}(z) = \underbrace{\prod_{1\le a\le k_0} \left(\frac{n_a(z)}{n}\right)^{n_a(z)}}_{\widehat{C}(z)} . \tag{5.12}
$$

Now, we observe that by canceling the normalizing constant $c(x)$ we obtain that

$$
\frac{\sup_\theta \mathbb{P}_\theta(x)}{p_k(x)} \le \frac{\sum_z \widehat{A}(x,z)\widehat{B}(x,z)\widehat{C}(z)}{\sum_z A(x,z)B(x,z)C(z)} . \tag{5.13}
$$

Now, if we are able to find bounds D_1, D_2 and D_3, uniform on x and z, such that

$$\frac{\widehat{A}(x,z)}{A(x,z)} \;\le\; D_1\,, \quad \frac{\widehat{B}(x,z)}{B(x,z)} \;\le\; D_2 \quad \text{and} \quad \frac{\widehat{C}(x,z)}{C(x,z)} \;\le\; D_3$$

then we automatically get

$$\frac{\sup_\theta \mathbb{P}_\theta(x)}{p_k(x)} \;\le\; D_1\, D_2\, D_3\,. \tag{5.14}$$

These bounds follow by Lemmas 5.3, 5.4 and 5.5 proved below. Using these lemmas we obtain that

$$\begin{aligned}
\log \frac{\sup_\theta \mathbb{P}_\theta(x)}{p_k(x)} \;&\le\; \log\Big[(n+1)^{k(k+1)}(n^2\log n)^n n^k\Big] \\
&\le\; k(k+1)\log(n+1) + n(2\log n + \log\log n) + k\log n \\
&\le\; k(k+2)\log(n+1) + 3n\log n
\end{aligned}$$

concluding the proof of Proposition 4.1.

Lemma 5.3. *For $(x,z) \in \Omega_n \times [k]^n$ we have that*

$$\frac{\widehat{A}(x,z)}{A(x,z)} \;\le\; (n+1)^{k(k+1)}\,. \tag{5.15}$$

Proof: Fix any $(x,z) \in \Omega_n \times [k]^n$. We simplify the notation by writing $o_{ab} = o_{ab}(x,z)$ and similarly for $n_a(z)$ and $n_{ab}(z)$. By (5.8) an (5.11) we obtain that

$$\begin{aligned}
\frac{\widehat{A}(x,z)}{A(x,z)} \;&=\; \frac{\prod_{1\le a\le b\le k}\left(\frac{o_{ab}}{n_{ab}}\right)^{o_{ab}} e^{-o_{ab}}}{\frac{1}{\Gamma(\frac{1}{2})^{\frac{k(k+1)}{2}}}\prod_{1\le a\le b\le k}\frac{\Gamma(o_{ab}+\frac{1}{2})}{[n_{ab}+1]^{o_{ab}+\frac{1}{2}}}} \\
&= \Gamma\left(\frac{1}{2}\right)^{\frac{k(k+1)}{2}} \prod_{1\le a\le b\le k} \frac{\left(\frac{o_{ab}}{n_{ab}}\right)^{o_{ab}} e^{-o_{ab}}(n_{ab}+1)^{o_{ab}+\frac{1}{2}}}{\Gamma\left(o_{ab}+\frac{1}{2}\right)}.
\end{aligned} \tag{5.16}$$

Letting

$$N := \sum_{1\le a\le b\le k} o_{ab}$$

we rewrite

$$\frac{\widehat{A}(x,z)}{A(x,z)} = \Gamma\left(\frac{1}{2}\right)^{\frac{k(k+1)}{2}} \prod_{1\le a\le b\le k} \frac{\left(\frac{o_{ab}}{N}\right)^{o_{ab}}}{\Gamma\left(o_{ab}+\frac{1}{2}\right)} \left(\frac{N}{n_{ab}}\right)^{o_{ab}} (n_{ab}+1)^{o_{ab}+\frac{1}{2}}\, e^{-o_{ab}}. \tag{5.17}$$

We now use Lemma 5.1 to get

$$\prod_{1\le a\le b\le k} \frac{\left(\frac{o_{ab}}{N}\right)^{o_{ab}}}{\Gamma\left(o_{ab}+\frac{1}{2}\right)} \;\le\; \frac{1}{\Gamma\left(N+\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)^{\frac{k(k+1)}{2}-1}}\,. \tag{5.18}$$

On the other hand

$$\begin{aligned}
\left(\frac{N}{n_{ab}}\right)^{o_{ab}} (n_{ab}+1)^{o_{ab}+\frac{1}{2}}\, e^{-o_{ab}} \;&=\; N^{o_{ab}}\left(1+\frac{1}{n_{ab}}\right)^{o_{ab}} (n_{ab}+1)^{\frac{1}{2}}\, e^{-o_{ab}} \\
&\le\; N^{o_{ab}}\, e^{\frac{o_{ab}}{n_{ab}}}\,(n_{ab}+1)^{\frac{1}{2}}\, e^{-o_{ab}}.
\end{aligned} \tag{5.19}$$

Putting (5.18) and (5.19) in (5.17), we get

$$\begin{aligned}\frac{\widehat{A}(x,z)}{A(x,z)} &\leq \frac{\Gamma\left(\frac{1}{2}\right)^{\frac{k(k+1)}{2}}}{\Gamma\left(N+\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)^{\frac{k(k+1)}{2}-1}} \prod_{1\leq a\leq b\leq k} N^{o_{ab}} e^{\frac{o_{ab}}{n_{ab}}}\left(n_{ab}+1\right)^{\frac{1}{2}} e^{-o_{ab}} \\ &= \frac{\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(N+\frac{1}{2}\right)} e^{-N} N^{N} e^{\sum_{1\leq a\leq b\leq k}\frac{o_{ab}}{n_{ab}}} \prod_{1\leq a\leq b\leq k}\left(n_{ab}+1\right)^{\frac{1}{2}}. \end{aligned} \tag{5.20}$$

For any real number $r>0$, $\Gamma(r)\geq r^{r-1/2}e^{-r}\sqrt{2\pi}$, so that

$$\frac{e^{-N}N^{N}}{\Gamma\left(N+\frac{1}{2}\right)} \leq \left(1+\frac{1}{2N}\right)^{-N} e^{\frac{1}{2}}\frac{1}{\sqrt{2\pi}} \leq \frac{1}{\sqrt{2\pi}}. \tag{5.21}$$

Moreover, as $x\in\Omega_n$ we have that

$$\sum_{1\leq a\leq b\leq k}\frac{o_{ab}}{n_{ab}} \leq \sum_{1\leq a\leq b\leq k}\frac{n_{ab}\log n}{n_{ab}} = \frac{k(k+1)}{2}\log n \tag{5.22}$$

and

$$\prod_{1\leq a\leq b\leq k}\left(n_{ab}+1\right)^{\frac{1}{2}} \leq \prod_{1\leq a\leq b\leq k}\left(n^2+1\right)^{\frac{1}{2}} = \left(n^2+1\right)^{\frac{k(k+1)}{4}} \leq (n+1)^{\frac{k(k+1)}{2}}. \tag{5.23}$$

Plugging (5.21), (5.22) and (5.23) into (5.20) proves the lemma. □

Lemma 5.4. *For $(x,z)\in\Omega_n\times[k]^n$ we have*

$$\frac{\widehat{B}(x,z)}{B(x,z)} \leq (n^2\log n)^n. \tag{5.24}$$

Proof: For $(x,z)\in\Omega_n\times[k]^n$ we have that

$$\begin{aligned}\frac{\widehat{B}(x,z)}{B(x,z)} &= \prod_{1\leq a\leq k}\frac{\prod_{i\in[n]:z_i=a}\left(\frac{n_a(z)d_i(x)}{d_a^t(x,z)}\right)^{d_i(x)}}{\frac{n_a(z)^{d_a^t(x,z)}\Gamma\left(\frac{n_a(z)}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^{n_a(z)}}\frac{\prod_{i:\, z_i=a}\Gamma(d_i(x)+\frac{1}{2})}{\Gamma(d_a^t(x,z)+\frac{n_a(z)}{2})}} \\ &= \prod_{1\leq a\leq k}\frac{\Gamma\left(\frac{1}{2}\right)^{n_a(z)}\Gamma(d_a^t(x,z)+\frac{n_a(z)}{2})}{\Gamma\left(\frac{n_a(z)}{2}\right)}\prod_{i:z_i=a}\frac{\left(\frac{d_i(x)}{d_a^t(x,z)}\right)^{d_i(x)}}{\Gamma(d_i(x)+\frac{1}{2})}\end{aligned}$$

But for all $a\in[k]$ we have by Lemma 5.1 that

$$\prod_{i:z_i=a}\frac{\left(\frac{d_i(x)}{d_a^t(x,z)}\right)^{d_i(x)}}{\Gamma(d_i(x)+\frac{1}{2})} \leq \frac{1}{\Gamma\left(d_a^t(x,z)+\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)^{n_a(z)-1}}, \tag{5.25}$$

where we used the equality $d_a^t(x,z)=\sum_{i:z_i=a}d_i(x)$. Putting all the previous bounds together we obtain that

$$\frac{\widehat{B}(x,z)}{B(x,z)} \leq \prod_{1\leq a\leq k}\frac{\Gamma\left(\frac{1}{2}\right)\Gamma(d_a^t(x,z)+\frac{n_a(z)}{2})}{\Gamma\left(\frac{n_a(z)}{2}\right)\Gamma\left(d_a^t(x,z)+\frac{1}{2}\right)}.$$

Finally, by Lemma 5.2 we conclude, as $d_a^t(x,z)\leq n^2\log n$ for all $a\in[k]$ and $\sum_a n_a(z)=n$ that

$$\frac{\widehat{B}(x,z)}{B(x,z)} \leq \prod_{1\leq a\leq k} d_a^t(x,z)^{n_a(z)} \leq (n^2\log n)^n.$$

□

Lemma 5.5. *For any $z \in [k]^n$ we have that*

$$\frac{\widehat{C}(z)}{C(z)} \;\leq\; n^k\,. \tag{5.26}$$

Proof: By definition we have that

$$\frac{\widehat{C}(z)}{C(z)} = \frac{\prod_{a=1}^k \left(\frac{n_a(z)}{n}\right)^{n_a(z)}}{\frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k}\frac{\prod_{a=1}^k \Gamma\left(n_a(z)+\frac{1}{2}\right)}{\Gamma\left(n+\frac{k}{2}\right)}}\,.$$

By Lemma 5.1 we obtain that

$$\prod_{a=1}^k \frac{\left(\frac{n_a(z)}{n}\right)^{n_a(z)}}{\Gamma\left(n_a(z)+\frac{1}{2}\right)} \;\leq\; \frac{1}{\Gamma\left(n+\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)^{k-1}}\,. \tag{5.27}$$

This last inequality and Lemma 5.2 imply that

$$\frac{\widehat{C}(z)}{C(z)} \leq \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(n+\frac{k}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(n+\frac{1}{2}\right)} \;\leq\; n^k\,.$$

□

5.3. *Proof of other auxiliary lemmas.*

Lemma 5.6. *Consider the DCSBM with k_0 communities. The set Ω_n defined in (4.1) satisfies*

$$\mathbb{P}_\theta(\Omega_n) = 1$$

for sufficiently large n.

Proof: We have that

$$\begin{aligned}
\mathbb{P}_\theta(\Omega_n^c) \;&=\; \sum_{z\in[k_0]^n} \mathbb{P}_\theta(\Omega_n^c, z)\\
&=\; \sum_{z\in[k_0]^n} \mathbb{P}_\pi(z)\mathbb{P}_{\lambda,w}(\Omega_n^c|z)\\
&\leq\; \sum_{z\in[k_0]^n} \mathbb{P}_\pi(z) \sum_{i,j=1}^n \sum_{x:x_{ij}<\log n} \mathbb{P}_{\lambda,w}(x|z).
\end{aligned} \tag{5.28}$$

Notice that in the last display, $\sum_{x:x_{ij}<\log n} \mathbb{P}_{\lambda,w}(x|z)$ is, in "random variable notation", $\mathbb{P}_{\lambda,w}(X_{ij} \geq \log n|Z=z)$. Now, conditionally on $Z=z$, the X_{ij}'s have a Poisson distribution with parameter $\lambda_{z_iz_j}$. From Boucheron et al. (2013, Section 2.2), if $Y \sim \text{Poisson}(\mu)$, then for $r>0$

$$\begin{aligned}
\text{Prob}(Y-\mu\geq r) \;&\leq\; e^{-\mu\left(1+\frac{r}{\mu}\right)\log\left(1+\frac{r}{\mu}\right)+r}\\
&\leq\; e^{-r\left(\log\left(1+\frac{r}{\mu}\right)-1\right)}\,.
\end{aligned}$$

Thus, for n such that $\log n > \lambda_{z_iz_j}$

$$\begin{aligned}
\mathbb{P}_{\lambda,w}(X_{ij}\geq \log n|Z=z) \;&\leq\; e^{-(\log n-\lambda_{z_iz_j})\left(\log\left(\frac{\log n}{\lambda_{z_iz_j}}\right)-1\right)}\\
&\leq\; e^{-(\log n-\lambda_{\max})\left(\log\left(\frac{\log n}{\lambda_{\max}}\right)-1\right)}
\end{aligned} \tag{5.29}$$

where in the last line we used the notation $\lambda_{\max} = \max_{a,b} \lambda_{ab}$. Then, by (5.28) and (5.29) we have that

$$\begin{aligned}\mathbb{P}_\theta(\Omega_n^c) \;&\leq\; n^2 e^{-(\log n - \lambda_{\max})\left(\log\left(\frac{\log n}{\lambda_{\max}}\right)-1\right)} \sum_{z\in[k_0]^n} \mathbb{P}_\pi(z)\\ &=\; n^{3+\log\lambda_{\max}-\log(\log n)}\,(\log n)^{\lambda_{\max}}\,(e\lambda_{\max})^{-\lambda_{\max}} \end{aligned} \tag{5.30}$$

which is summable in n and therefore the lemma is proved by application of the Borel Cantelli Lemma. □

To state the next auxiliary result, we need some notation. Define for all $z \in \{1,\dots,k\}^n$ and $z^0 \in \{1,\dots,k_0\}^n$ the $k \times k_0$ matrix $Q_n(z,z^0)$ given by

$$[Q_n(z,z^0)]_{aa'} \;=\; \frac{1}{n}\sum_{i=1}^n w_i\, \mathbb{1}\{z_i = a, z_i^0 = a'\}. \tag{5.31}$$

Observe that the counters $n_{a'}(z^0)$, for $a' \in [k_0]$, can be written as

$$n_{a'}(z^0) = \sum_{i=1}^n w_i\, \mathbb{1}\{z_i^0 = a'\} = \sum_{i=1}^n w_i \sum_{a=1}^k \mathbb{1}\{z_i = a, z_i^0 = a'\}\,. \tag{5.32}$$

Then $n_{a'}(z^0) = n(Q_n^T(z,z^0)\mathbf{1}_k)_{a'}$, with $\mathbf{1}_k$ a column vector of dimension k with all entries equal to 1. Moreover, the matrix $Q_n(z,z^0)$ satisfies

$$\|Q_n(z,z^0)\|_1 \;=\; \sum_{a=1}^k \sum_{a'=1}^{k_0} [Q_n(z,z^0)]_{a,a'} = 1, \tag{5.33}$$

for all (z,z^0) and

$$\begin{aligned}\mathbb{E}_{\lambda,w}(\tilde{o}_{ab}(X,z)\,|\,Z=z^0) \;&=\; n^2 Q_n(z,z^0)\lambda Q_n(z,z^0)^T &(5.34)\\ &=\; \rho_n n^2 Q_n(z,z^0)\tilde{\lambda} Q_n(z,z^0)^T &(5.35)\end{aligned}$$

in which $\mathbb{E}_{\lambda,w}$ denotes the expectation with respect to the measure $\mathbb{P}_{\lambda,w}$.

We are now ready to prove the concentration bounds used in the proof of Proposition 4.3.

Lemma 5.7. *Consider the DCSBM with k_0 communities. For any $\epsilon > 0$ and $a,b \in [k]$ we have that*

$$\mathbb{P}_\theta\Big(\sup_{\bar{z}\in[k]^n}\Big|\frac{\tilde{o}_{ab}(X,\bar{z})}{\rho_n n^2} - [Q_n(\bar{z},Z)\tilde{\lambda}Q_n(\bar{z},Z)^T]_{ab}\Big| > \epsilon\Big) \;\leq\; \exp\Big(-\frac{\rho_n n^2\epsilon^2}{\tilde{\lambda}_{\max}+\epsilon} + n\log k\Big)$$

and

$$\mathbb{P}_\theta\Big(\sup_{\bar{z}\in[k]^n}\Big|\frac{d_a^t(X,\bar{z})}{\rho_n n^2} - [Q_n(\bar{z},Z)\tilde{\lambda}Q_n(\bar{z},Z)^T\mathbf{1}_k]_a\Big| \;>\; \epsilon\Big) \;\leq\; \exp\Big(-\frac{\rho_n n^2\epsilon^2}{\tilde{\lambda}_{\max}+\epsilon} + n\log k\Big)\,,$$

with $\tilde{\lambda}_{\max} = \max_{a,b}\tilde{\lambda}_{ab}$.

Proof: For any fixed $z \in [k_0]^n$ and $\bar{z} \in [k]^n$ we have that

$$\begin{aligned}&\tilde{o}_{ab}(X,\bar{z}) - n^2[Q_n(\bar{z},z)\lambda Q_n(\bar{z},z)^T]_{ab}\\ &\quad= \tilde{o}_{ab}(X,\bar{z}) - \rho_n n^2[Q_n(\bar{z},z)\tilde{\lambda}Q_n(\bar{z},z)^T]_{ab}\\ &\quad= \sum_{1\leq i,j\leq n}\;\sum_{1\leq a',b'\leq k_0}(X_{ij} - \rho_n w_i w_j \tilde{\lambda}_{a'b'})\mathbb{1}\{\bar{z}_i = a, z_i = a'\}\mathbb{1}\{\bar{z}_j = b, z_j = b'\}\,.\end{aligned}$$

Observe that given $Z = z$, $\tilde{o}_{ab}(X,\bar{z})$ corresponds to the sum of $n_a(\bar{z})n_b(\bar{z})$ independent Poisson random variables, given by $X_{ij}\mathbb{1}\{\bar{z}_i = a, \bar{z}_i = b\}$, with expected value given by $\rho_n w_i w_j \tilde{\lambda}_{z_i z_j}$. Then the sum is also Poisson distributed with a parameter that is the sum of the corresponding parameters. Using one more time Boucheron et al. (2013, Section 2.2), we have for $Y \sim \text{Poisson}(\mu)$ and $t > 0$

$$\text{Prob}(Y - \mu \geq r) \leq e^{-\mu\left(1+\frac{r}{\mu}\right)\log\left(1+\frac{r}{\mu}\right)+r}$$
$$\text{Prob}(Y - \mu \leq -r) \leq e^{-\mu\left(1-\frac{r}{\mu}\right)\log\left(1-\frac{r}{\mu}\right)-r}$$

which, after some algebra, yields

$$\text{Prob}(|Y - \mu| > r) \;\leq\; 2e^{-\frac{r^2}{2(\mu+r)}}\,.$$

Therefore, for any $\delta > 0$

$$\mathbb{P}_\theta\Big(\,\Big|\tilde{o}_{ab}(X, \bar{z}) - \rho_n n^2[Q_n(\bar{z}, z)\tilde{\lambda}Q_n(\bar{z}, z)^T]_{ab}\Big| \;>\; \delta\,|Z\; = z\Big)$$
$$\leq\; 2\exp\Big(-\frac{\delta^2}{2(\rho_n n^2[Q_n(\bar{z}, z)\tilde{\lambda}Q_n(\bar{z}, z)^T]_{ab} + \delta)}\Big)\,.$$

Since, for any $\bar{z}$ and z, we have that

$$\rho_n n^2[Q_n(\bar{z}, z)\tilde{\lambda}Q_n(\bar{z}, z)^T]_{ab} \;\leq\; \rho_n n^2\tilde{\lambda}_{\max}$$

with $\tilde{\lambda}_{\max} = \max_{a,b}\tilde{\lambda}_{ab}$, it follows that, for any $z, \bar{z}$ and $\epsilon > 0$

$$\mathbb{P}_\theta\Big(\,\Big|\frac{\tilde{o}_{ab}(X, \bar{z})}{\rho_n n^2} - [Q_n(\bar{z}, z)\tilde{\lambda}Q_n(\bar{z}, z)^T]_{ab}\Big| \;>\; \epsilon\,|\,Z = z\Big)$$
$$\leq\; \exp\Big(-\frac{\rho_n^2 n^4\epsilon^2}{\rho_n n^2\tilde{\lambda}_{\max} + \epsilon\rho_n n^2}\Big)$$
$$\leq\; \exp\Big(-\frac{\rho_n n^2\epsilon^2}{\tilde{\lambda}_{\max} + \epsilon}\Big)\,.$$

Now, using a union bound over all $\bar{z} \in [k]^n$ and integrating over z we obtain that

$$\mathbb{P}_\theta\Big(\sup_{\bar{z}\in[k]^n}\Big|\frac{\tilde{o}_{ab}(X, \bar{z})}{\rho_n n^2} - [Q_n(\bar{z}, Z)\tilde{\lambda}Q_n(\bar{z}, Z)^T]_{ab}\Big| \;>\; \epsilon\Big)$$
$$\leq\; \exp\Big(-\frac{\rho_n n^2\epsilon^2}{\tilde{\lambda}_{\max} + \epsilon} + n\log k\Big)$$

and this proves the first inequality of the lemma. Now, given $Z = z$,

$$d_a^t(X, \bar{z}) \;=\; \sum_{b\in[k]}\tilde{o}_{ab}(X, \bar{z})$$

is also a sum of independent random variables with Poisson distribution and

$$\mathbb{E}_{\lambda,w}(d_a^t(X, \bar{z})\,|\,Z = z) \;=\; [\rho_n n^2 Q_n(\bar{z}, z)\tilde{\lambda}Q_n(\bar{z}, z)^T\mathbf{1}_k]_a,$$

thus we also obtain that

$$\mathbb{P}_\theta\Big(\sup_{\bar{z}\in[k]^n}\Big|\frac{d_a^t(X, \bar{z})}{\rho_n n^2} - [Q_n(\bar{z}, Z)\tilde{\lambda}Q_n(\bar{z}, Z)^T\mathbf{1}_k]_a\Big| > \epsilon\Big) \leq \exp\Big(-\frac{\rho_n n^2\epsilon^2}{\tilde{\lambda}_{\max} + \epsilon} + n\log k\Big)\,.$$

This concludes the proof of Lemma 5.7 □

In the sequel, we state and prove the lemmas cited in the proof of Proposition 4.3.

Lemma 5.8. *For $k < k_0$ there exists a $k \times k$ positive matrix λ^* and k dimensional vector π^* such that*

$$\limsup_{n\to\infty} \frac{1}{2} \sum_{1\le a,b\le k} \frac{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}{\rho_n^2 n^4}\,\varphi\Big(\frac{\rho_n n^2 \tilde{o}_{ab}(x,\widehat{z}_k)}{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}\Big) \le \frac{1}{2}\sum_{1\le a,b\le k} \pi_a^*\pi_b^*[\lambda^*\pi^*]_a[\lambda^*\pi^*]_b\,\varphi\Big(\frac{\lambda_{ab}^*}{[\lambda^*\pi^*]_a[\lambda^*\pi^*]_b}\Big)\,. \tag{5.36}$$

Moreover, (π^*, λ^*) are given by

$$\begin{aligned} \pi_a^* &= [R^*\mathbf{1}_{k_0}]_a, \qquad a\in\{1,\ldots,k\} \\ \lambda_{ab}^* &= \frac{[R^*\lambda R^{*^T}]_{ab}}{[R^*\mathbf{1}_{k_0}\mathbf{1}_{k_0}^T R^{*^T}]_{ab}}\,, \qquad a,b\in\{1,\ldots,k\}\,, \end{aligned} \tag{5.37}$$

for a $k \times k_0$ real matrix R^* satisfying $\|R^*\|_1 = 1$ and having one and only one non-zero entry on each column.

Proof: Observe that

$$\sum_{1\le a,b\le k} \frac{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}{\rho_n^2 n^4}\varphi\Big(\frac{\rho_n n^2 \tilde{o}_{ab}(x,\widehat{z}_k)}{\rho_n\, d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}\Big) = \sum_{1\le a,b\le k} \frac{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}{\rho_n^2 n^4}\varphi\left(\frac{\tilde{o}_{ab}(x,\widehat{z}_k)/\rho_n n^2}{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)/\rho_n^2 n^4}\right), \tag{5.38}$$

where $\varphi(u) = u\log u$. Then by Lemma 5.7, taking $\epsilon_n = \rho_n = \frac{\log n}{n}$ we have that

$$\Big|\frac{\tilde{o}_{ab}(x,\widehat{z}_k)}{\rho_n n^2} - [Q_n(\widehat{z}_k,z)\tilde{\lambda}Q_n^T(\widehat{z}_k,z)]_{ab}\Big| \le \epsilon_n$$

and similarly

$$\Big|\frac{d_a^t(x,\widehat{z}_k)}{\rho_n n^2} - [Q_n(\widehat{z}_k,z)\tilde{\lambda}Q_n(\widehat{z}_k,z)^T\mathbf{1}_k]_a\Big| \le \epsilon_n$$

eventually almost surely as $n \to \infty$. Then as φ is continuous, substituting $\tilde{o}_{ab}(x,\widehat{z}_k)/\rho_n n^2$ by $[Q_n(\widehat{z}_k,z)\tilde{\lambda}Q_n^T(\widehat{z}_k,z)]_{ab}$ and substituting $d_a^t(x,\widehat{z}_k)/\rho_n n^2$ by $[Q_n(\widehat{z}_k,z)\tilde{\lambda}Q_n(\widehat{z}_k,z)^T\mathbf{1}_k]_a$ in the right-hand side of (5.38) we obtain that

$$\sum_{1\le a,b\le k} \frac{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}{\rho_n^2 n^4}\,\varphi\Big(\frac{\rho_n n^2 \tilde{o}_{ab}(x,\widehat{z}_k)}{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}\Big) \le \sup_{\substack{Q_n\colon \|Q_n\|_1=1\\ Q_n^T\mathbf{1}_k=n(z)/n}} \sum_{1\le a,b\le k} [Q_n\tilde{\lambda}Q_n^T\mathbf{1}_k]_a[Q_n\tilde{\lambda}Q_n^T\mathbf{1}_k]_b\varphi\left(\frac{[Q_n\tilde{\lambda}Q_n^T]_{ab}}{[Q_n\tilde{\lambda}Q_n^T\mathbf{1}_k]_a[Q_n\tilde{\lambda}Q_n^T\mathbf{1}_k]_b}\right) + \eta_n. \tag{5.39}$$

for some sequence $\eta_n \to 0$ as $n \to \infty$. Then taking $\limsup$ on both sides, we must have that

$$\limsup_{n\to\infty} \sum_{1\le a,b\le k} \frac{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}{\rho_n^2 n^4}\,\varphi\Big(\frac{\rho_n n^2 \tilde{o}_{ab}(x,\widehat{z}_k)}{d_a^t(x,\widehat{z}_k)d_b^t(x,\widehat{z}_k)}\Big) \le \sup_{\substack{R\colon \|R\|_1=1\\ R^T\mathbf{1}_k=\pi}} \frac{1}{2}\sum_{1\le a,b\le k} [R\tilde{\lambda}R^T\mathbf{1}_k]_a[R\tilde{\lambda}R^T\mathbf{1}_k]_b\,\varphi\Big(\frac{[R\tilde{\lambda}R^T]_{ab}}{[R\tilde{\lambda}R^T\mathbf{1}_k]_a[R\tilde{\lambda}R^T\mathbf{1}_k]_b}\Big) \tag{5.40}$$

almost surely. Then, the supremum in the right-hand side of (5.40) is a maximum of a convex function over a convex polyhedron defined by $\{R\colon \|R\|_1 = 1, R^T\mathbf{1}_k = \pi\}$. Then, the maximum must be attained at one of the vertices of the polyhedron; that is, on those matrixes R such that one and only one entry by column is greater than zero, given that $\pi_a > 0$ for all $a \in \{1, \dots, k_0\}$. We denote by R^* one of these maximums (if there is more than one) and let

$$\begin{aligned} \pi^*_a &= [R^* \mathbf{1}_{k_0}]_a \qquad a \in \{1, \dots, k\} \\ \lambda^*_{ab} &= \frac{[R^* \tilde{\lambda} {R^*}^T]_{ab}}{[R^* \mathbf{1}_{k_0} \mathbf{1}_{k_0}^T {R^*}^T]_{ab}} \,, \qquad a, b \in \{1, \dots, k\} \,. \end{aligned} \tag{5.41}$$

Then

$$\begin{aligned} \sup_{\substack{R:\, \|R\|_1 = 1 \\ R^T \mathbf{1}_k = \pi}} \; & \frac{1}{2} \sum_{1 \le a, b \le k} [R \tilde{\lambda} R^T \mathbf{1}_k]_a [R \tilde{\lambda} R^T \mathbf{1}_k]_b \, \varphi\Big(\frac{[R \tilde{\lambda} R^T]_{ab}}{[R \tilde{\lambda} R^T \mathbf{1}_k]_a [R \tilde{\lambda} R^T \mathbf{1}_k]_b} \Big) \\ &= \frac{1}{2} \sum_{1 \le a, b \le k} \pi^*_a \pi^*_b [\lambda^* \pi^*]_a [\lambda^* \pi^*]_b \, \varphi\Big(\frac{\lambda^*_{ab}}{[\lambda^* \pi^*]_a [\lambda^* \pi^*]_b} \Big) \,. \end{aligned} \tag{5.42}$$

This concludes the proof of Lemma 5.8. □

Lemma 5.9. *Assume $\tilde{\lambda}$ has no two proportional columns. Then for all $k < k_0$ and (π^*, λ^*) as in Lemma 5.8 we have that*

$$\begin{aligned} & \sum_{1 \le a, b \le k_0} \pi_a \pi_b [\tilde{\lambda} \pi]_a [\tilde{\lambda} \pi]_b \, \varphi\Big(\frac{\tilde{\lambda}_{ab}}{[\tilde{\lambda} \pi]_a [\tilde{\lambda} \pi]_b} \Big) \\ & \qquad - \sum_{1 \le a, b \le k} \pi^*_a \pi^*_b [\lambda^* \pi^*]_a [\lambda^* \pi^*]_b \, \varphi\Big(\frac{\lambda^*_{ab}}{[\lambda^* \pi^*]_a [\lambda^* \pi^*]_b} \Big) \; > \; 0 \,. \end{aligned} \tag{5.43}$$

Proof: First consider the case $k = k_0 - 1$. As R^* has one and only one non-zero entry in each column, we have that there is a surjective function $h\colon [k_0] \to [k]$ connecting each community in $[k_0]$ (columns of R^*) with is corresponding community in $[k]$ (line with non-zero entry). Then for $k = k_0 - 1$, there are $k - 1$ communities in $\{1, \dots, k_0\}$ that are mapped into $k - 1$ communities in $\{1, \dots, k\}$ and two communities in $\{1, \dots, k_0\}$ that are mapped into a single community in $\{1, \dots, k\}$. Without loss of generality assume that the communities $k_0 - 1$ and k_0 satisfy $h(k_0 - 1) = h(k_0) = k = k_0 - 1$. Moreover, as ${R^*}^T \mathbf{1}_k = \pi$ we must have that the non-zero entries are given by

$$\begin{aligned} R^*_{aa} &= \pi_a \,, \qquad 1 \le a \le k_0 - 1 \\ R^*_{(k_0 - 1) k_0} &= \pi_{k_0} \,. \end{aligned} \tag{5.44}$$

Then the parameters π^* and λ^* defined in (5.41) are given by

$$\begin{aligned} \pi^*_a &= \pi_a \,, \qquad 1 \le a \le k_0 - 1 \\ \pi^*_{k_0 - 1} &= \pi_{k_0 - 1} + \pi_{k_0} \end{aligned}$$

and

$$\begin{aligned} \lambda^*_{ab} &= \tilde{\lambda}_{ab} \,, \qquad 1 \le a, b \le k_0 - 2 \\ \lambda^*_{a(k_0-1)} &= \frac{\pi_{k_0-1} \tilde{\lambda}_{a(k_0-1)} + \pi_{k_0} \tilde{\lambda}_{a k_0}}{\pi_{k_0-1} + \pi_{k_0}} \,, \qquad 1 \le l \le k_0 - 2, \\ \lambda^*_{(k_0-1)(k_0-1)} &= \frac{\pi^2_{k_0-1} \tilde{\lambda}_{(k_0-1)(k_0-1)} + 2 \pi_{k_0-1} \pi_{k_0} \tilde{\lambda}_{(k_0-1) k_0} + \pi^2_{k_0} \tilde{\lambda}_{k_0 k_0}}{\pi^2_{k_0-1} + 2 \pi_{k_0-1} \pi_{k_0} + \pi^2_{k_0}} \,. \end{aligned}$$

Observe that for all $1 \le a \le k_0 - 1$ we have that $[\lambda^* \pi^*]_a = [\tilde{\lambda} \pi]_a$ then for all $1 \le a, b \le k_0 - 2$

$$\pi^*_a \pi^*_b [\lambda^* \pi^*]_a [\lambda^* \pi^*]_b \, \varphi\Big(\frac{\lambda^*_{ab}}{[\lambda^* \pi^*]_a [\lambda^* \pi^*]_b} \Big) \; = \; \pi_a \pi_b [\tilde{\lambda} \pi]_a [\tilde{\lambda} \pi]_b \, \varphi\Big(\frac{\tilde{\lambda}_{ab}}{[\tilde{\lambda} \pi]_a [\tilde{\lambda} \pi]_b} \Big) \,.$$

On the other hand we have that

$$[\lambda^*\pi^*]_{k_0-1} = \frac{\pi_{k_0-1}[\tilde{\lambda}\pi]_{k_0-1} + \pi_{k_0}[\tilde{\lambda}\pi]_{k_0}}{\pi_{k_0-1} + \pi_{k_0}}\,.$$

Then for $1 \le a \le k_0 - 2$ it follows, by the log-sum inequality, that

$$\begin{aligned}
&\pi^*_a\pi^*_{k_0-1}[\lambda^*\pi^*]_a[\lambda^*\pi^*]_{k_0-1}\varphi\Big(\frac{\lambda^*_{a(k_0-1)}}{[\lambda^*\pi^*]_a[\lambda^*\pi^*]_{k_0-1}}\Big)\\
&= \pi_a[\tilde{\lambda}\pi]_a(\pi_{k_0-1}[\tilde{\lambda}\pi]_{k_0-1} + \pi_{k_0}[\tilde{\lambda}\pi]_{k_0})\varphi\left(\frac{\pi_{k_0-1}\tilde{\lambda}_{a(k_0-1)} + \pi_{k_0}\tilde{\lambda}_{ak_0}}{[\tilde{\lambda}\pi]_a(\pi_{k_0-1}[\tilde{\lambda}\pi]_{k_0-1} + \pi_{k_0}[\tilde{\lambda}\pi]_{k_0})}\right)\\
&= \pi_a(\pi_{k_0-1}\tilde{\lambda}_{a(k_0-1)} + \pi_{k_0}\tilde{\lambda}_{ak_0})\log\left(\frac{\pi_a(\pi_{k_0-1}\tilde{\lambda}_{a(k_0-1)} + \pi_{k_0}\tilde{\lambda}_{ak_0})}{\pi_a[\tilde{\lambda}\pi]_a(\pi_{k_0-1}[\tilde{\lambda}\pi]_{k_0-1} + \pi_{k_0}[\tilde{\lambda}\pi]_{k_0})}\right)\\
&\le \pi_a\pi_{k_0-1}\tilde{\lambda}_{a(k_0-1)}\log\Big(\frac{\tilde{\lambda}_{a(k_0-1)}}{[\tilde{\lambda}\pi]_a[\tilde{\lambda}\pi]_{k_0-1}}\Big) + \pi_a\pi_{k_0}\tilde{\lambda}_{ak_0}\log\Big(\frac{\tilde{\lambda}_{ak_0}}{[\tilde{\lambda}\pi]_a[\tilde{\lambda}\pi]_{k_0}}\Big)\\
&= \pi_a\pi_{k_0-1}[\tilde{\lambda}\pi]_a[\tilde{\lambda}\pi]_{k_0-1}\varphi\Big(\frac{\tilde{\lambda}_{a(k_0-1)}}{[\tilde{\lambda}\pi]_a[\tilde{\lambda}\pi]_{k_0-1}}\Big) + \pi_a\pi_{k_0}[\tilde{\lambda}\pi]_a[\tilde{\lambda}\pi]_{k_0}\varphi\Big(\frac{\tilde{\lambda}_{ak_0}}{[\tilde{\lambda}\pi]_a[\tilde{\lambda}\pi]_{k_0}}\Big)\,.
\end{aligned} \tag{5.45}$$

Moreover, we have that the inequality must be strict unless

$$\frac{\tilde{\lambda}_{a(k_0-1)}}{[\tilde{\lambda}\pi]_a[\tilde{\lambda}\pi]_{k_0-1}} = \frac{\tilde{\lambda}_{ak_0}}{[\tilde{\lambda}\pi]_a[\tilde{\lambda}\pi]_{k_0}} \qquad \text{for all } a \le k_0 - 2\,. \tag{5.46}$$

On the other hand, for $a = k_0 - 1$ and $b = k_0 - 1$, also by the log-sum inequality we have that

$$\begin{aligned}
&\pi^*_{k_0-1}\pi^*_{k_0-1}[\lambda^*\pi^*]_{k_0-1}[\lambda^*\pi^*]_{k_0-1}\varphi\Big(\frac{\lambda^*_{(k_0-1)(k_0-1)}}{[\lambda^*\pi^*]_{k_0-1}[\lambda^*\pi^*]_{k_0-1}}\Big)\\
&= (\pi^2_{k_0-1}\tilde{\lambda}_{(k_0-1)(k_0-1)} + 2\pi_{k_0-1}\pi_{k_0}\tilde{\lambda}_{(k_0-1)k_0} + \pi^2_{k_0}\tilde{\lambda}_{k_0k_0})\\
&\qquad\qquad \times \log\Big(\frac{\pi^2_{k_0-1}\tilde{\lambda}_{(k_0-1)(k_0-1)} + 2\pi_{k_0-1}\pi_{k_0}\tilde{\lambda}_{(k_0-1)k_0} + \pi^2_{k_0}\tilde{\lambda}_{k_0k_0}}{\pi^2_{k_0-1}[\tilde{\lambda}\pi]^2_{k_0-1} + 2\pi_{k_0-1}\pi_{k_0}[\tilde{\lambda}\pi]_{k_0-1}[\tilde{\lambda}\pi]_{k_0} + \pi^2_{k_0}[\tilde{\lambda}\pi]^2_{k_0}}\Big)\\
&\le \pi^2_{k_0-1}\tilde{\lambda}_{(k_0-1)(k_0-1)}\log\Big(\frac{\tilde{\lambda}_{(k_0-1)(k_0-1)}}{[\tilde{\lambda}\pi]^2_{k_0-1}}\Big)\\
&\qquad + 2\pi_{k_0-1}\pi_{k_0}\tilde{\lambda}_{(k_0-1)k_0}\log\Big(\frac{\tilde{\lambda}_{(k_0-1)k_0}}{[\tilde{\lambda}\pi]_{k_0-1}[\tilde{\lambda}\pi]_{k_0}}\Big) + \pi^2_{k_0}\tilde{\lambda}_{k_0k_0}\log\Big(\frac{\tilde{\lambda}_{k_0k_0}}{[\tilde{\lambda}\pi]^2_{k_0}}\Big)\\
&= \pi^2_{k_0-1}[\tilde{\lambda}\pi]^2_{k_0-1}\varphi\Big(\frac{\tilde{\lambda}_{(k_0-1)(k_0-1)}}{[\tilde{\lambda}\pi]^2_{k_0-1}}\Big)\\
&\qquad + 2\pi_{k_0-1}\pi_{k_0}[\tilde{\lambda}\pi]_{k_0-1}[\tilde{\lambda}\pi]_{k_0}\varphi\Big(\frac{\tilde{\lambda}_{(k_0-1)k_0}}{[\tilde{\lambda}\pi]_{k_0-1}[\tilde{\lambda}\pi]_{k_0}}\Big) + \pi^2_{k_0}[\tilde{\lambda}\pi]^2_{k_0}\varphi\Big(\frac{\tilde{\lambda}_{k_0k_0}}{[\tilde{\lambda}\pi]^2_{k_0}}\Big)\,,
\end{aligned} \tag{5.47}$$

with equality if and only if

$$\frac{\tilde{\lambda}_{(k_0-1)(k_0-1)}}{[\tilde{\lambda}\pi]^2_{k_0-1}} = \frac{\tilde{\lambda}_{(k_0-1)k_0}}{[\tilde{\lambda}\pi]_{k_0-1}[\tilde{\lambda}\pi]_{k_0}} = \frac{\tilde{\lambda}_{k_0k_0}}{[\tilde{\lambda}\pi]^2_{k_0}}\,. \tag{5.48}$$

From (5.46) and (5.48) we obtain that the inequality (5.43) must be strict unless

$$\tilde{\lambda}_{a(k_0-1)} = \frac{[\tilde{\lambda}\pi]_{k_0-1}}{[\tilde{\lambda}\pi]_{k_0}}\tilde{\lambda}_{ak_0} \qquad \text{for all } a \le k_0 \tag{5.49}$$

which is a contradiction with the hypothesis for the identifiability of k_0. □

References

Abbe, E. Community detection and stochastic block models: recent developments. *J. Mach. Learn. Res.*, **18**, Paper No. 177, 86 (2017). MR3827065.

Amini, A. A., Chen, A., Bickel, P. J., and Levina, E. Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.*, **41** (4), 2097–2122 (2013). MR3127859.

Bickel, P. J. and Chen, A. A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA*, **106** (50), 21068–21073 (2009). DOI: 10.1073/pnas.0907096106.

Biernacki, C., Celeux, G., and Govaert, G. Exact and Monte Carlo calculations of integrated likelihoods for the latent class model. *J. Statist. Plann. Inference*, **140** (11), 2991–3002 (2010). MR2659830.

Boucheron, S., Lugosi, G., and Massart, P. *Concentration inequalities. A nonasymptotic theory of independence.* Oxford University Press, Oxford (2013). ISBN 978-0-19-953525-5. MR3185193.

Celisse, A., Daudin, J.-J., and Pierre, L. Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.*, **6**, 1847–1899 (2012). MR2988467.

Cerqueira, A. and Leonardi, F. Estimation of the number of communities in the stochastic block model. *IEEE Trans. Inform. Theory*, **66** (10), 6403–6412 (2020). MR4173547.

Chen, K. and Lei, J. Network cross-validation for determining the number of communities in network data. *J. Amer. Statist. Assoc.*, **113** (521), 241–251 (2018). MR3803461.

Daudin, J.-J., Picard, F., and Robin, S. A mixture model for random graphs. *Stat. Comput.*, **18** (2), 173–183 (2008). MR2390817.

Davisson, L. D., McEliece, R. J., Pursley, M. B., and Wallace, M. S. Efficient universal noiseless source codes. *IEEE Trans. Inform. Theory*, **27** (3), 269–279 (1981). MR619112.

Decelle, A., Krzakala, F., Moore, C., and Zdeborová, L. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, **84**, 066106 (2011). DOI: 10.1103/PhysRevE.84.066106.

Holland, P. W., Laskey, K. B., and Leinhardt, S. Stochastic blockmodels: first steps. *Social Networks*, **5** (2), 109–137 (1983). MR718088.

Hu, J., Qin, H., Yan, T., and Zhao, Y. Corrected Bayesian information criterion for stochastic block models. *J. Amer. Statist. Assoc.*, **115** (532), 1771–1783 (2020). MR4189756.

Jin, J. Fast community detection by SCORE. *Ann. Statist.*, **43** (1), 57–89 (2015). MR3285600.

Karrer, B. and Newman, M. E. J. Stochastic blockmodels and community structure in networks. *Phys. Rev. E (3)*, **83** (1), 016107, 10 (2011). MR2788206.

Latouche, P., Birmelé, E., and Ambroise, C. Variational Bayesian inference and complexity control for stochastic block models. *Stat. Model.*, **12** (1), 93–115 (2012). MR2953099.

Le, C. M. and Levina, E. Estimating the number of communities by spectral methods. *Electron. J. Stat.*, **16** (1), 3315–3342 (2022). MR4422967.

Lei, J. A goodness-of-fit test for stochastic block models. *Ann. Statist.*, **44** (1), 401–424 (2016). MR3449773.

Lei, J. and Rinaldo, A. Consistency of spectral clustering in stochastic block models. *Ann. Statist.*, **43** (1), 215–237 (2015). MR3285605.

Ma, S., Su, L., and Zhang, Y. Determining the number of communities in degree-corrected stochastic block models. *J. Mach. Learn. Res.*, **22**, Paper No. 69, 63 (2021). MR4253762.

Newman, M. E. J. and Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E*, **69**, 026113 (2004). DOI: 10.1103/PhysRevE.69.026113.

Qin, T. and Rohe, K. Regularized Spectral Clustering under the Degree-Corrected Stochastic Blockmodel. In Burges, C. et al., editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc. (2013). ISBN 9781632660244.

Rohe, K., Chatterjee, S., and Yu, B. Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.*, **39** (4), 1878–1915 (2011). MR2893856.

Sarkar, P. and Bickel, P. J. Role of normalization in spectral clustering for stochastic blockmodels. *Ann. Statist.*, **43** (3), 962–990 (2015). MR3346694.

van der Pas, S. L. and van der Vaart, A. W. Bayesian community detection. *Bayesian Anal.*, **13** (3), 767–796 (2018). MR3807866.

Wang, Y. X. R. and Bickel, P. J. Likelihood-based model selection for stochastic block models. *Ann. Statist.*, **45** (2), 500–528 (2017). MR3650391.

Yan, X., Shalizi, C., et al. Model selection for degree-corrected block models. *J. Stat. Mech. Theory Exp.*, **5**, P05007 (2014). DOI: 10.1088/1742-5468/2014/05/P05007.

Zhao, Y., Levina, E., and Zhu, J. Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.*, **40** (4), 2266–2292 (2012). MR3059083.